دانشکده علوم ریاضی

رساله دکتری کنترل و بهینه سازی

حل رده ای از مسائل ماشین بردار پشتیبان با استفاده از روش های بهینه سازی دینامیکی

نگارنده: امیر فیضی

استاد راهنما

دکتر علیرضا ناظمی

استاد مشاور

دکتر محمدرضا ربیعی

بهمن ۱۳۹۹

شماره: ۳-۲/۲-۴۰۵
تاریخ: ۱۴۰۰/۲/۷
ویرایش:

باسمه تعالی

مدیریت تحصیلات تکمیلی

پیوست شماره ۲

دانشکده : ریاضی

گروه : آمار

رساله دکتری: **آقای امیر فیضی**

تحت عنوان: **حل رده ای از مسائل ماشین بردار پشتیبان با استفاده از روش های بهینه سازی دینامیکی**

در تاریخ **۱۳۹۹/۱۱/۱۹** توسط کمیته تخصصی زیر جهت اخذ مدرک **رساله دکتری** ارزیابی گردید و با درجه ..عالی..... مورد پذیرش قرار گرفت.

اساتید راهنما	امضاء	اساتید مشاور	امضاء
نام و نام خانوادگی: **دکتر علیرضا ناظمی**		نام و نام خانوادگی: **دکتر محمدرضا ربیعی**	
نام و نام خانوادگی: -------	-------	نام و نام خانوادگی: -------	-------

اساتید داور	امضاء	نماینده تحصیلات تکمیلی	امضاء
نام و نام خانوادگی: **دکتر جعفر فتحعلی**		نام و نام خانوادگی: **دکتر سیدحیدر جعفری**	
نام و نام خانوادگی: **دکتر مهرداد غزنوی**			
نام و نام خانوادگی: **دکتر مهدی روزبه**			

وزارت علوم، تحقیقات و فناوری
دانشکده ریاضی

تقدیم به

پیشگاه مقدس حضرت فاطمة الزّهرا
سلام الله علیها

همسر مهربانم

و دختران عزیزم رقیه و فاطمه

سپاس گزاری...

من لم یشکر المخلوق لم یشکر الخالق

حمد و ستایش خداوند سبحان را، که مرا توفیق عنایت فرمود تا در راه کسب علم و دانش گام بردارم. هم‌چنین درود و سلام خداوند به پیشگاه مقدس ائمه اطهار(علیهم السلام) خاصه ولی نعمتمان آقا علی بن موسی‌الرضا (علیه السلام)، که در تمامی عمر، به‌ویژه دوران تحصیل در مجاورت آن امام همام، از الطاف کریمانه ایشان برخوردار بوده‌ام.

و به پاس احترام به حرمت دانش بر خود لازم می‌دانم که از زحمات و راهنمایی‌های استاد گرانقدر و ارجمندم

جناب آقای دکتر علیرضا ناظمی

که نظارت این رساله را بر عهده داشته و در تمام مراحل از یاری و مساعدت ایشان برخوردار بوده‌ام، تشکر و قدرانی نمایم و موفقیت ایشان را از خدای متعال خواهانم.

از

جناب آقای دکتر محمدرضا ربیعی

که در امر مشاوره این رساله مساعدت نمودند و در این امر نهایت مراقبت، توجه و دقت خود را مبذول فرموده اند کمال تشکر و امتنان را دارم و برای ایشان از خداوند سلامت و سعادت ابدی را خواهانم.

هم‌چنین از اساتید فرهیخته جناب آقای دکتر جعفر فتحعلی، جناب آقای دکتر مهدی روزبه و جناب آقای دکتر مهرداد غزنوی که زحمت داوری این رساله را به عهده گرفتند با تمام وجود تشکر و قدردانی می‌نمایم.

امیر فیضی
بهمن ۱۳۹۹

تعهد نامه

اینجانب **امیر فیضی** دانشجوی دکتری رشته **ریاضی کاربردی علوم ریاضی** دانشگاه شاهرود، نویسنده پایان‌نامه با عنوان **حل رده ای از مسائل ماشین بردار پشتیبان با استفاده از روش های بهینه سازی دینامیکی** ، تحت راهنمایی **علیرضا ناظمی** متعهد می‌شوم:

- تحقیقات در این پایان‌نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است.
- در استفاده از نتایج پژوهش‌های دیگر پژوهش‌گران، به مرجع مورد استفاده استناد شده است.
- مطالب این پایان‌نامه، تا کنون توسط خود، یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ‌جا ارایه نشده است.
- حقوق معنوی این اثر، به دانشگاه صنعتی شاهرود تعلق دارد، و مقالات مستخرج با نام ‘‘ دانشگاه صنعتی شاهرود’’ یا “ Shahrood University of Technology ” به چاپ خواهد رسید.
- حقوق معنوی تمام افرادی که در به‌دست آوردن نتایج اصلی پایان‌نامه تاثیرگذار بوده‌اند، در مقالات مستخرج از پایان‌نامه رعایت می‌گردد.
- در تمام مراحل انجام این پایان‌نامه، در مواردی که از موجود زنده (یا بافت‌های آنها) استفاده شده است، ضوابط و اصول اخلاقی رعایت شده است.
- در تمام مراحل انجام این پایان‌نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته (یا استفاده شده است)، اصل رازداری و اصول اخلاق انسانی رعایت شده است.

امیر فیضی
بهمن ۱۳۹۹

مالکیت نتایج و حق نشر

- تمام حقوق معنوی این اثر و محصولات آن (مقالات مستخرج، کتاب، برنامه‌های رایانه‌ای، نرم‌افزارها و تجهیزات ساخته شده) متعلق به دانشگاه صنعتی شاهرود می‌باشد. این مطلب باید به نحو مقتضی، در تولیدات علمی مربوطه ذکر شود.
- استفاده از اطلاعات و نتایج موجود در این پایان‌نامه بدون ذکر منبع مجاز نمی‌باشد.

چکیده

ماشین های بردار پشتیبان ابزارهای قدرتمندی برای طبقه بندی داده ها و رگرسیون هستند. در سال های اخیر، الگوریتم های سریع بسیاری برای ماشین‌های بردار پشتیبان توسعه یافته اند. از ماشین بردار پشتیبان در بسیاری از کابردهای نظامی و مهندسی، مسائل زمان واقعی، مانند طبقه بندی در محیط الکترومغناطیسی پیچیده، تشخیص در پزشکی و ... استفاده می‌شود. با توجه به این‌که روش های همگرایی عددی نیاز به محاسبات بیشتری داشته و نمی تواند نیاز ما در مسائل زمان واقعی را برآورده سازد، یک روند میسر و امیدوار کننده برای آموزش ماشین های بردار پشتیبان در زمان های واقعی بکارگیری شبکه های عصبی بازگشتی بر پایه اجرای دوره ای است.

در این رساله ما با معرفی شبکه عصبی به حل مسائلی از بردار پشتیبان مانند مساله رگرسیون بردار پشتیبان و نیز مساله رگرسیون بردار پشتیان تصادفی با قیدهای احتمالی توسط شبکه عصبی می‌پردازیم. ثابت می‌کنیم نقطه تعادل شبکه عصبی با جواب بهینه مساله بردار پشتیبان معادل هستند. هم‌چنین پایداری و همگرایی شبکه عصبی ارائه شده اثبات خواهد شد. در نهایت با ارائه مثال‌هایی نتایج نظری را تایید می‌کنیم.

کلمات کلیدی: بردار پشتیبان، شبکه عصبی، بهینه‌سازی، رگرسیون، متغیر تصادفی

لیست مقالات مستخرج از پایان‌نامه

۱. مقاله اول:

Feizi, Amir and Nazemi, Alireza. An application of a practical neural network model for solving support vector regression problems. *Intelligent Data Analysis*, 21:1443–1461, 2017.

۲. مقاله دوم:

Feizi, Amir, Nazemi, Alireza, and Mohammad Reza, Rabeie. Solving the stochastic support vector regression with probabilistic constraints by a high performance neural network model. *Engineering with Computers*, 2020.

فهرست مطالب

فهرست تصاویر

فهرست جداول

پیشگفتار

ماشین بردار پشتیبان یکی از روش‌های یادگیری با نظارت است که از آن برای طبقه‌بندی و رگرسیون استفاده می‌کنند. این روش از جملهٔ روش‌های نسبتاً جدیدی است که در سال‌های اخیر کارایی خوبی نسبت به روش‌های قدیمی‌تر برای طبقه‌بندی نشان داده است. مبنای کاری دسته‌بندی کننده‌ی SVM دسته‌بندی خطی داده‌ها است و در تقسیم خطی داده‌ها سعی می‌کنیم خطی را انتخاب کنیم که حاشیه اطمینان بیشتری داشته باشد. حل معادله پیدا کردن خط بهینه برای داده‌ها به وسیله روش‌های مسائل درجه دوم که روش‌های شناخته شده‌ای در حل مسائل محدودیت‌دار هستند صورت می‌گیرد. برای به‌دست آوردن بردار پشتیبان از قضیه دوگانی لاگرانژ و شرایط کاروش کان تاکر استفاده می‌کنیم.

فصل اول این رساله به معرفی بردار پشتیبان با استفاده از مسائل کلاس بندی اختصاص دارد.

در فصل دوم این رساله شبکه عصبی بیولوژیکی و مصنوعی را توضیح می‌دهیم.

فصل سوم این رساله به معرفی یک شبکه عصبی بازگشتی برای حل مسائل رگرسیون می‌پردازد. در این فصل ثابت می‌شود که نقطه تعادل شبکه عصبی با جواب بهینه مساله رگرسیون بردار پشتیبان معادل است. هم‌چنین همگرایی و پایداری شبکه عصبی معرفی شده، اثبات می‌گردد.

با توجه به این که در اغلب مسائل، کمیت‌های استفاده شده به دلیل انواع خطاهای ممکن دقیق نیستند، در فصل آخر با اصلاح شبکه عصبی استفاده شده در فصل سوم، مساله رگرسیون بردار پشتیبان با قیدهای احتمالی را به‌وسیله شبکه عصبی حل می‌کنیم.

فصل ۱

ابرصفحه بهینه برای جداسازی نمونه‌ها و معرفی بردار پشتیبان

ماشین بردار پشتیبان یکی از روش‌های یادگیری با نظارت است که از آن برای طبقه‌بندی و رگرسیون استفاده می‌شود. به عبارت دیگر، با دریافت داده های آموزشی برچسب خورده (آموزش نظارت شده)، الگوریتم یک ابرصفحه جدا کننده بهینه را خروجی می دهد که نمونه های جدید را طبقه‌بندی می کند. در فضای دو بعدی این ابرصفحه یک خط تقسیم کننده‌ی صفحه به دو بخش است که در آن هر کلاس در یک طرف خط قرار می گیرد. در این فصل به معرفی بردار پشتیبان با استفاده از مساله کلاس‌بندی می‌پردازیم. عمدتاً مطالب این فصل از مرجع [۱] گرفته شده است.

۱.۱ شناسایی الگو و مفاهیم مقدماتی

اگر تعدادی سیب و گلابی در اختیار شما قرار دهند و از شما بخواهند که تشخیص دهید هر یک به کدام دسته از دو نوع میوه تعلق دارند، چه می‌کنید؟ این مسأله ظاهراً زیاد پیچیده نیست، هر کسی می‌تواند بر اساس مشاهده به‌راحتی آن‌ها را از هم جدا کند. به علاوه، خیلی سخت نیست که تکه‌های فاسد، کثیف و بقیه اشیائی که نه سیب هستند و نه گلابی را تشخیص داد.

اگر چه این مساله تشخیص سیب‌ها و گلابی‌ها از یکدیگر به نظر مشکل نیست، اما ماشینی کردن کردن فرایند کم و بیش پیچیده است. مثلاً چه چیزی باید مبنای تصمیم‌گیری باشد که بگوییم یک شئ «سیب» است یا «گلابی»؟ آیا وزن آن شئ است، اندازه یا رنگ آن؟ شاید شکل آن، بو یا طعم آن باشد؟ احتمالاً ترکیبی از همه این ویژگی‌هاست. لحظه‌ای فکر کنید ما اندازه‌هایی از رنگ و بوی یک شئ را داریم، چگونه یک سیب که روی آن کمی گِلی است رده بندی می‌شود؟ آیا آن یک سیب است یا یک شئ گِلی؟

این همان مساله رده‌بندی است: تخصیص یک شئ جدید به یک مجموعه از کلاس‌ها (در این‌جا سیب‌ها یا گلابی‌ها) که از پیش معلوم هستند. رده‌بندی کننده‌ای که باید عمل جداسازی را انجام دهد (یا به هر شئ ورودی یک برچسب خروجی تخصیص دهد)، برپایه مجموعه‌ای از مثال‌ها عمل می‌کند. در این‌جا واژه‌ی شئ کاملاً عمومی است، می‌تواند از «سیب» و «گلابی» تا «ارقام دست‌نویس»، از «سیگنال‌های گفتاری» تا «صدای موزیک» متغیر باشد [۴۸].

۱.۱.۱ نظریه یادگیری به‌وسیله سرپرست

آیا ماشین می‌تواند فقط دستوراتی را که ما به آن می‌دهیم انجام دهد؟ در موارد بسیاری ما از ماشین می‌خواهیم که وظایفی را انجام دهد که نمی‌توان آن‌ها را با الگوریتم شرح داد. این مساله هنگامی که مدل ریاضی مشخص و قابل دسترسی وجود ندارد یا موقعی که محاسبه راه حل نشدنی یا خیلی گران دارد، پیش می‌آید. حل کردن این مساله با استفاده از روش‌های برنامه‌نویسی کلاسیک غیرممکن است. به عنوان مثال شناسایی اعداد در دست‌نوشته‌های مختلف، را در نظر بگیرید (این موضوع یکی از مساله‌های کلاسیک در زمینه‌ی آموزش ماشین است). ما برای حل این مساله، نیاز داریم مساله را از یک دیدگاه دیگر بررسی کنیم. احتمالاً می‌توانیم ماشین را آموزش دهیم، به همان شیوه‌ای که کودکان را آموزش می‌دهیم، با این تفاوت که به آن‌ها خلاصه مفاهیم و تعاریف و مسائل تئوری را آموزش نمی‌دهیم، ولی توجه آن‌ها را به تابع‌هایی از ورودی و خروجی جلب می‌کنیم. چگونه بچه‌ها خواندن حروف الفبا را یاد می‌گیرند، معلم به آن‌ها تعریف دقیق و کاربرد حروف را به‌طور مجزا نمی‌گوید اما به آن‌ها مثال‌هایی را نشان می‌دهد که این حروف در آن‌ها به‌کار رفته‌اند. شاگرد با دقت و تمرین کردن زیاد این مثال‌ها یک استنباط کلی از مشخصات آن حرف پیدا می‌کند، در پایان اگر شاگردان به نوع حروف فکر کنند، می‌توانند لغات را در متن بخوانند و تشخیص دهند. از هدف‌های آشکار و اصلی آموزش، دسته‌بندی کردن صحیح مشاهدات آینده است. ما در این‌جا به یک سری روش‌های دسته‌بندی با تاثیرهای روشنی از اثرشان روی داده نیاز داریم. چگونه آموزش به عنوان مثال، در مساله ریاضی فرموله می‌شود؟

۲.۱.۱ مدل‌سازی مساله آموزش

مساله آموزش یک مدل کلی را شرح می‌دهد که شامل سه قسمت اصلی است. این سه قسمت عبارتند از: تولید کننده داده، یک سرپرست که یک برچسب y را به داده مربوط می‌کند و ماشین آموزش که خروجی y را تحت عنوان پاسخ سرپرست برمی‌گرداند.

هدف تقلید از سرپرست می‌باشد، تلاش برای ساختن یک عملکرد، که بهترین پیشگویی را از خروجی‌های سرپرست به‌دست آورد، که برپایه‌ی داده‌های تولید شده است. این موضوع با این‌که تلاش کنیم تا عملکرد سرپرست یا پاسخ سرپرست را مشخص کنیم فرق دارد.

تولید کننده محیطی که سرپرست ماشین آموزش باید در آن فعالیت کند را مشخص می‌کند. در واقع بردارهای $x \in \mathbb{R}^m$ را تولید می‌کند که مستقل از هم هستند و توزیع آن‌ها را با توجه به یک توزیع احتمال غیرمشخص $P(x)$، مشخص می‌کند. سرپرست مقدار واقعی را با توجه به توزیع احتمالی شرطی منصوب می‌کند. در این موضوع ما با مساله دسته‌بندی دوتایی سروکار داریم. ماشین آموزش با مجموعه‌ای ممکن از نگاشت‌های $x \to F(x,a)$ تعریف می‌شود، به‌طوری که a عضوی از فضای پارامتری A است که $A \equiv (w,b) \in \mathbb{R}^{m+1}$. به‌عنوان یک ماشین آموزش که با ابرصفحه جهت‌دار $\{x :< w,x > +b = 0\}$ تعریف می‌شود، موقعیت یک نقطه در $\mathbb{R}^m$ را مشخص می‌کند، که در آن $< w,x >$ ضرب داخلی دو بردار w و x است. این نتایج در ماشین آموزش به‌صورت زیر است:

$$f(x,(w,b)) = sign(< w,x > +b) : (w,b) \in \mathbb{R}^{m+1}.$$

تابع $f : \mathbb{R}^m \to \{-1,1\}$ که x را به $\{-1,1\}$ نگاشت می‌کند، **تابع تصمیم** نام دارد.

تعریف ۱.۱.۱. انتخاب مشخص از پارامتر a که بر اساس مشاهده می‌باشد، **مجموعه‌ی تمرینی** نام دارد.

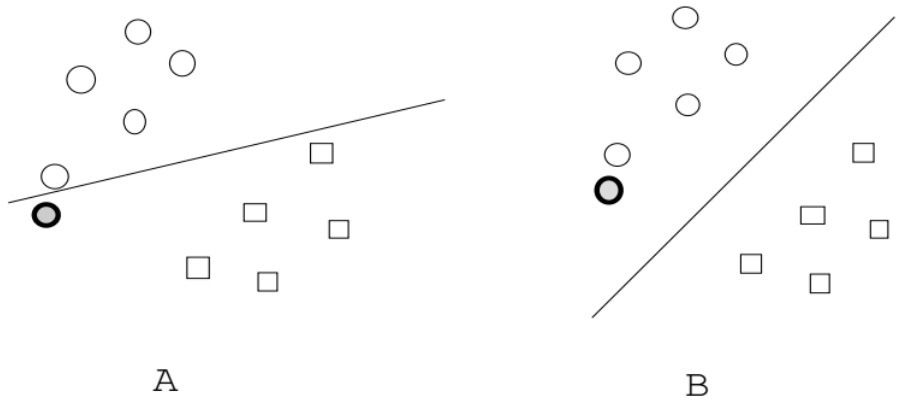
۳.۱.۱ رده‌بندی کننده خطی

به عنوان اولین قدم در ساخت ماشین بردار پشتیبان، دسته‌بندی کردن با ابرصفحه‌های خطی ساده را مورد مطالعه قرار می‌دهیم. رده‌ای از ابرصفحه‌های $< w,x > +b = 0$، $x \in \mathbb{R}^n$ که نظیر توابع تصمیم زیر هستند را در نظر بگیرید:

$$f(x) = sign(< w,x > +b).$$

برای شروع رده‌بندی مثال‌های جداشدنی (نقاط تمرینی جداشدنی) را در نظر می‌گیریم. راه‌های زیادی برای جداکردن دو کلاس از هم وجود دارد. ایده یادگیری از مثال‌ها یعنی استفاده از اثر اعضای سازنده کلاس‌ها با امتحان کردن نقاط تمرینی که متعلق به آن کلاس هستند، تعریف می‌شود.

ابرصفحه‌ها باید به‌گونه‌ای انتخاب شوند که تغییرات خطی در اطلاعات، باعث تغییر پیش‌بینی‌ها نشوند. اگر فاصله بین جداکننده و نقاط تمرینی خیلی کوچک شوند، حتی مثال‌هایی که برای امتحان کردن نقاط بسیار نزدیک به نقاط تمرین زده‌ایم نیز به‌درستی دسته‌بندی نمی‌شوند. شکل (۱.۱) این موضوع را توضیح می‌دهد، دایره‌ها و مربع‌های سفید نقاط تمرینی هستند و دایره خاکستری یک نمونه برای تست نمودن ابرصفحه جداکننده است. رویکرد وپنیک[1] و



شکل ۱.۱: کدام جداسازی بهتر است؟ حاشیه تقریباً صفر در A و یا حاشیه بزرگ در B؟

لرنر در سال ۱۹۳۶ و وپنیک در سال ۱۹۶۴ ریشه در این مشاهده دارد که قدرت عمومی سازی بستگی به فاصله ابرصفحه تا نقاط تمرینی دارد. آن‌ها یک الگوریتم یادگیری برای مسائل جداشدنی به نام تعمیم تصویر ارائه دادند، به‌طوری که ابرصفحه‌ای را می‌سازد که بیشترین فاصله را تا اعضای کلاس‌ها دارد،

$$\max_{w,b} \quad \min\{\|x - x_i\| : x \in R^n, < w, x > +b = ۰, \quad i = ۱, ..., n\}.$$

در این بخش ما یک راه مؤثر برای ساخت این ابرصفحه ارائه می‌دهیم.

تعریف ۲.۱.۱. (جداشوندگی) مجموعه تمرینی $\{x_i, y_i\}_{i=۱}^n$ که $x_i \in \mathbb{R}^m$ و $y_i \in \{-۱, ۱\}$، **جدا شونده** توسط ابرصفحه $< w, x > +b = ۰$ نامیده می‌شود اگر بردار w و مقدار اسکالر b وجود داشته باشند به‌طوری که روابط زیر برقرار باشند:

$$< w, x_i > +b > ۰, \quad y_i = ۱ \text{ برای برچسب } , \tag{۱.۱}$$

$$< w, x_i > +b < ۰, \quad y_i = -۱ \text{ برای برچسب }. \tag{۲.۱}$$

ابرصفحه‌ای که توسط w و b تعریف می شود **ابرصفحه جداکننده** نامیده می‌شود.

می‌توان نامعادله‌های (۱.۱) و (۲.۱) را به صورت معادل زیر نوشت:

$$y_i(< w, x_i > +b) > ۰, \quad y_i \in \{-۱, ۱\}. \tag{۳.۱}$$

[1] Vapnik

تعریف ۱.۱.۳. (حاشیه) ابرصفحه جداسازنده H را به‌صورت زیر در نظر بگیرید:

$$H :< w, x > +b = ۰.$$

حاشیه $\gamma_i(w,b)$ از یک نقطه تمرینی x_i به‌صورت فاصله بین H و x_i تعریف می‌شود:

$$\gamma_i(w,b) = y_i(< w, x_i > +b).$$

حاشیه $\gamma_s(w,b)$ از یک مجموعه بردار $S = \{x_۱, ..., x_n\}$ به‌صورت حداقل فاصله از H تا بردارهای S تعریف می‌شود:

$$\gamma_s(w,b) = \min_{x_i \in S} \gamma_i(w,b).$$

اکنون بردار w^* و اسکالر b^* که حاشیه مجموعه‌ی تمرینی γ_i را تحت شرایط نامعادله‌های (۱.۱) و (۲.۱) ماکزیمم می‌کند را در نظر بگیرید.
زوج (w^*, b^*) ابرصفحه‌ای را تعیین می‌کند که مثال‌های مثبت را از مثال‌های منفی با حاشیه ماکسیمال جدا می‌کند.این ابرصفحه را **جداکننده بهینه** می‌نامیم.

تعریف ۱.۱.۴. (جدا شوندگی) ابرصفحه جداکننده‌ی بهینه برای مجموعه‌ی تمرینی χ به‌صورت زیر تعریف می‌شود:

$$(w^*, b^*) = arg \max_{w,b} \gamma(w,b).$$

قضیه ۱.۱.۱. ابر صفحه جدا کننده بهینه یکتاست.

برهان. به مرجع [۵۰] مراجعه شود. □

با استفاده از تعریف داریم:

$$\begin{aligned}(w^*, b^*) &= arg \max_{w,b} \gamma_\chi(w,b) \\ &= arg \max_{w,b} \min_{x_i \in \chi} \gamma_i(w,b) \\ &= arg \max_{w,b} \min_{x_i \in \chi} y_i(< w, x_i > +b).\end{aligned}$$

فرض می‌کنیم w^* نقطه ماکزیمم تابع پیوسته زیر باشد:

$$\gamma(w) = \min_{x_i \in \chi} y_i(< w, x_i >),$$

که روی مجموعه $\|w\| \leq ۱$ تعریف شده است. در این‌صورت داریم:
۱) پیوستگی γ در ناحیه‌ی بسته $\|w\| \leq ۱$، وجود ماکزیمم را نتیجه می‌دهد.

۲) فرض کنید که ماکزیمم در نقطه‌ای داخلی مانند w' به‌دست می‌آید، که در آن $\|w'\| < ۱$. اکنون بردار $w'' = \frac{w'}{\|w'\|}$ حاشیه‌ی بزرگ‌تری را تعریف می‌کند، چون

$$\gamma(w'') = \frac{\gamma(w')}{\|w'\|} > \gamma(w'),$$

و این با بهینگی w' تناقض دارد. پس $\|w\| = ۱$ و ماکزیمم فقط می‌تواند روی مرز به دست آید.

با داشتن w^* می‌توانیم b^* را این‌گونه تعریف کنیم.

$$b^* = \frac{۱}{۲}(\min_{i \in I_+} < w^*, x_i > - \max_{i \in I_-} < w^*, x_i >),$$

که در آن $I_+ = \{i | y_i = ۱\}$ و $I_- = \{i | y_i = -۱\}$.

بنابراین مسأله‌ای که برای پیدا کردن ابرصفحه بهینه به‌دست می‌آید به‌صورت زیر است:

$$\begin{aligned} &\text{maximize } \gamma_\chi(w, b), \\ &\text{subject to } \gamma_\chi(w, b) > ۰, \\ &\|w\|^۲ = ۱. \end{aligned} \tag{۴.۱}$$

حل مساله (۴.۱) با توجه به این‌که تابع مورد نظر نه خطی است و نه درجه دوم و هم‌چنین قیود مساله غیرخطی هستند، دشوار است.

برای پیدا کردن ابرصفحه جدا کننده بهینه می‌توان مساله زیر را در نظر گرفت:

$$\begin{aligned} &\text{minimize } \frac{۱}{۲}\|w\|^۲, \\ &\text{subject to } < w, x_i > + b \geq +۱, \quad y_i = -۱ \text{ برای برچسب } , \\ &\qquad\qquad < w, x_i > + b \leq -۱, \quad y_i = +۱ \text{ برای برچسب } . \end{aligned} \tag{۵.۱}$$

قضیه ۲.۱.۱. مسائل بهینه‌سازی (۴.۱) و (۵.۱) معادل هستند.

برهان. برای اثبات به مرجع [۳] مراجعه نمایید. □

مساله بهینه‌سازی (۵.۱) را به روش لاگرانژ حل می‌کنیم. سعی می‌کنیم نقطه بهینه را با استفاده از تابع لاگرانژ زیر پیدا کنیم.

$$L_P(w, b, \alpha) = \frac{۱}{۲}\|w\|^۲ - \sum_{i=۱}^{n} \alpha_i [y_i(< w, x_i > + b) - ۱],$$

که در آن $\alpha_i \geq ۰$ ضرایب لاگرانژ هستند، تابع $L_P(w, b, \alpha)$ باید نسبت به متغیرهای اولیه w و b می‌نیمم و نسبت به متغیرهای دوگان (α_iها) ماکزیمم گردد.

در **نقطه زینی** مشتق‌های L_P نسبت به متغیرهای اولیه باید صفر شود، یعنی:

$$\frac{\partial L_P}{\partial w}(w, b, \alpha) = ۰,$$

و

$$\frac{\partial L_P}{\partial b}(w,b,\alpha)=۰.$$

بنابراین

$$w=\sum_{i=۱}^{n}\alpha_i y_i x_i, \tag{۶.۱}$$

و

$$\sum_{i=۱}^{n}\alpha_i y_i=۰. \tag{۷.۱}$$

با جایگزین کردن روابط فوق در L_P به دوگان ولف[1] مساله بهینه‌سازی زیر می‌رسیم:

$$\begin{aligned}\text{maximize}\quad L_D(\alpha)=&\frac{۱}{۲}\sum_{i=۱}^{n}\sum_{j=۱}^{n}\alpha_i\alpha_j y_i y_j<x_i,x_j>-\sum_{i=۱}^{n}\sum_{j=۱}^{n}\alpha_i\alpha_j y_i y_j<x_i,x_j>\\&+\sum_{i=۱}^{n}\alpha_i=\sum_{i=۱}^{n}\alpha_i-\frac{۱}{۲}\sum_{i=۱}^{n}\sum_{j=۱}^{n}\alpha_i\alpha_j y_i y_j<x_i,x_j>.\end{aligned}$$

برای ساختن ابرصفحه بهینه باید ضرایب α_i^* که مساله‌ی فوق را حل می‌کند، بیابیم.
P مخفف ابتدایی و D مخفف دوگان است. L_P و L_D از تابع هدف یکسان ولی با قیود متفاوت می‌باشند. جواب بهینه با می‌نیمم کردن L_P و ماکزیمم کردن L_D به‌دست می‌آید.

لم ۱.۱.۱. فرض کنید بردار w^* ابرصفحه بهینه را تعریف می‌کند. مقدار ماکزیمم تابع $L_D(\alpha^*)$ برابر است با:

$$L_D(\alpha^*)=\frac{۱}{۲}\|w^*\|^۲=\frac{۱}{۲}\sum_{i=۱}^{n}\alpha_i^*,$$

برهان. جواب‌های بهینه w^* و α^* باید در شرط مکمل کان ـ تاکر صدق کنند، یعنی:

$$\alpha_i^*[y_i(<w^*,x_i>+b^*)-۱]=۰,\quad \forall i.$$

با جمع بر روی تمام (α_iها) داریم:

$$\sum_{i=۱}^{n}\alpha_i^* y_i<w^*,x_i>+b^*\sum_{i=۱}^{n}\alpha_i^* y_i-\sum_{i=۱}^{n}\alpha_i^*=۰.$$

با استفاده از شرایط (۶.۱) و (۷.۱) داریم:

$$\sum_{i=۱}^{n}\alpha_i^*=\sum_{i=۱}^{n}\alpha_i^* y_i<w^*,x_i>=\|w^*\|^۲. \tag{۸.۱}$$

[1] Wolf dual

از طرفی از رابطه‌ی (۶.۱) داریم:

$$\|w^*\|^۲ = w^{*^T}.w^* = \sum_{i=۱}^{n}\sum_{j=۱}^{n}\alpha_i^*\alpha_j^* y_i y_j < x_i, x_j >.$$

با جای‌گذاری (۷.۱) و (۸.۱) در دوگان ولف مساله بهینه سازی، مقدار تابع هدف دوگان مساله به‌صورت زیر نوشته می‌شود:

$$L_D(\alpha^*) = \sum_{i=۱}^{n}\alpha_i^* - \frac{۱}{۲}\sum_{i=۱}^{n}\sum_{j=۱}^{n}\alpha_i^*\alpha_j^* y_i y_j < x_i, x_j > = \|w^*\|^۲ - \frac{۱}{۲}\|w^*\|^۲ = \frac{۱}{۲}\|w^*\|^۲,$$

و اثبات تمام است. □

اکنون شرط مکمل کان ـ تاکر $(\alpha_i^*[y_i(< w^*, x_i > +b^*) - ۱] = ۰, \quad \forall i)$ را دقیق‌تر بررسی می‌کنیم. معنی این شرط در مساله ماشین بردار پشتیبان این است که برای یک نقطه تمرینی x_i، یا مؤلفه‌ی لاگرانژ α_i نظیر آن صفر است یا x_i بر روی یکی از ابرصفحه‌های $H_۱$ یا $H_۲$ است:

$$H_۱ : \quad < w, x > +b = +۱,$$

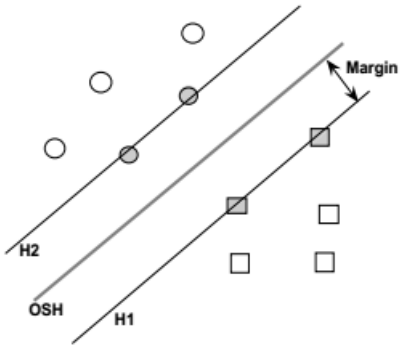
$$H_۲ : \quad < w, x > +b = -۱.$$

ابرصفحه‌های $H_۱$ و $H_۲$، **ابرصفحه‌های حاشیه** نامیده می‌شوند. آن‌ها دارای نقاط تمرینی با حداقل فاصله از ابرصفحه جداکننده بهینه هستند. ابرصفحه‌های حاشیه، مرز حاشیه را تشکیل می‌دهند و بردارهایی که α_i نظیر آن‌ها مثبت است، بر آن‌ها واقع هستند.

تعریف ۵.۱.۱. نقطه تمرینی x_i یک **بردار پشتیبان** نامیده می‌شود اگر و تنها اگر ضریب لاگرانژ نظیر آن مثبت باشد$(\alpha_i > ۰)$.

بردارهای پشتیبان نزدیک‌ترین نقاط تمرینی به ابرصفحه‌های جدا کننده هستند. بقیه نقاط تمرینی دارای $\alpha_i = ۰$ هستند، یعنی یا بر یکی از ابرصفحه‌های حاشیه‌ای قرار دارند و یا به یکی از نیم فضاهای تعریف (۲.۱.۱) تعلق دارند. توجه داشته باشید که شرط مکمل فقط می‌گوید که تمام بردارهای پشتیبان روی صفحه‌های حاشیه هستند، نه این‌که بردارهای پشتیبان تنها نقاط تمرینی روی آن ابرصفحه‌ها هستند. ممکن است موردی باشد که هر دوی α_i و $y_i(< w, x_i > +b)$ برابر صفر شوند. در این حالت x_i روی یکی از ابرصفحه‌های حاشیه‌ای است، بدون این‌که بردار پشتیبان باشد.

فرایند یافتن ابرصفحه بهینه جدا کننده‌ی دو کلاس همان مرحله **یادگیری** است، یعنی با استفاده از نقاط تمرینی ماشین را برای تشخیص الگو آموزش می‌دهیم. لذا از نقاط تمرینی به عنوان **نقاط آموزشی** نیز نام می‌بریم.



شکل ۲.۱: ابر صفحه جدا کننده بهینه با استفاده از بردارهای پشتیبان خاکستری بر روی ابرصفحه‌های حاشیه‌ای به‌دست می‌آید.

۲.۱ ماشین بردار پشتیبان

مساله و مفاهیم شناسایی الگو به عنوان یک عامل مهم در طراحی سیستم‌های اطلاعاتی بسیار مورد توجه قرار گرفته است. تئوری شناسایی مبتنی بر روش‌های تحلیلی قدرتمندی است که کاربردهای وسیعی در سیستم‌های کنترل و مخابرات دارد. میزان علاقه در این رشته (مخصوصاً در دهه‌های ۶۰ و ۷۰) در حال گسترش بوده و می‌باشد، به‌طوری که تبدیل به یک موضوع مشترک تحقیقاتی برای رشته‌های آکادمیک مختلف از قبیل مهندسی، علوم پایه مانند فیزیک، آمار و شیمی، زبان‌شناسی، روان‌شناسی، بیولوژی، تئوری اطلاعات، علم کامپیوتر و ... گشته است.

هر یک از حوزه‌های تحقیقاتی بر ویژگی خاصی از مساله شناسایی الگو از قبیل مدل‌سازی فرایندهای فیزیولوژی و یا مثلاً توسعه تکنیک‌های تحلیلی برای تصمیم‌گیری‌های اتوماتیک تاکید دارند.

مساله شناسایی الگو، عملاً چیزی جز جداسازی داده‌ها یا الگوهای ورودی بین دسته‌های مختلف نیست. در طراحی یک سیستم شناسایی الگو چند کار باید انجام گیرد:
نخست اشیاء مورد شناسایی را باید به الگوها و یا بردارهای ورودی که به سیستم شناسایی اعمال می‌گردند، تبدیل نمود. مؤلفه‌های بردار ورودی از طریق اندازه‌گیری به‌دست می‌آیند. هر کمیت اندازه‌گیری شده، یک ویژگی از شئ مورد نظر را بیان می‌کند. از نظر هندسی هر شئ یا موضوع شناسایی را می‌توان به عنوان یک نقطه در فضای چندبعدی اقلیدسی در نظر گرفت.
کار دوم استخراج مشخصه‌های مهم از روی داده‌های ورودی اندازه‌گیری شده و کاهش ابعاد بردار الگوها می‌باشد، به این کار اصطلاحاً استخراج شاخص می‌گویند. مشخصه‌های یک طبقه آن‌هایی هستند که در تمامی الگوهای متعلق به آن طبقه خاص مشترک می‌باشند. عناصری از بردار مشخصه‌ها که در تمامی طبقه‌ها مشترک هستند حذف می‌گردند.

سومین کار در شناسایی الگو، این است که روندی جهت تصمیم‌گیری بهینه اتوماتیک، برای طبقه بندی الگوها به‌دست آوریم.

به بیان ریاضی مساله شناسایی الگو به شکل ساده، یعنی تخصیص الگوهای ورودی به یکی از طبقاتی که فضای چندبعدی اقلیدسی برای تصمیم‌گیری به تعداد متناهی از آن تقسیم شده است. یکی از ایده‌های جدید در شناسایی و دسته‌بندی الگوها، ماشین بردار پشتیبان یا SVM است. ماشین بردار پشتیبان دارای خواص ارزشمندی است که آن‌را برای شناسایی الگو مناسب می‌سازد، از جمله این‌که SVM در آموزش خود مشکل بهینه‌های محلی را ندارد، دسته‌بندی کننده‌ها را با حداکثر تعمیم بنا می‌کند، ساختار و توپولوژی خود را به‌صورت بهینه تعیین می‌کند و توابع متمایز غیرخطی را به راحتی و با محاسبات کم با استفاده از مفهوم حاصل‌ضرب داخلی در فضای هیلبرت تشکیل می‌دهد.

اولین الگوریتم برای طبقه‌بندی الگوها در سال ۱۹۳۶ توسط فیشر[1] ارائه شد و معیار آن برای بهینه بودن، کم کردن خطای طبقه بندی الگوهای آموزشی بود. بسیاری از الگوریتم‌ها و روش‌هایی که تاکنون نیز برای طبقه‌بندی کننده‌های الگو ارائه شده است از همین استراتژی پیروی می‌کنند. در هیچ‌یک از این روش‌ها خاصیت تعمیم طبقه‌بندی کننده به‌طور مستقیم در تابع هزینه‌ی روش دخالت داده نشده است و طبقه‌بندی کننده طراحی شده نیز دارای خاصیت تعمیم دهنده‌گی کمی می‌باشد. اگر طراحی دسته‌بندی کننده‌ی الگو را به‌عنوان یک مساله بهینه‌سازی در نظر بگیریم بسیاری از این روش‌ها با مشکل بهینه‌های محلی در تابع هزینه مواجه هستند و در دام بهینه‌های محلی گرفتار می‌شوند.

شکل دیگری نیز وجود دارد و آن تعیین ساختار و توپولوژی دسته‌بندی کننده قبل از طراحی است که به‌عنوان مثال تعیین تعداد بهینه‌ی گره‌های لایه‌ی مخفی در شبکه‌های عصبی MLP[2]، تعداد توابع گاوسی در شبکه‌های عصبی RBF[3] و یا تعداد بهینه‌ی حالت‌ها و توابع گاوسی در مدل پنهان مارکف می‌باشد. همه این عوامل باعث می‌شوند که در عمل با روش‌های پیشنهاد شده‌ی قبلی نتوانیم به یک دسته‌بندی کننده‌ی بهینه برسیم.

محقق روسی به‌نام ولادمیر وپنیک[4] در سال ۱۹۶۵ گامی بسیار مهم در طراحی دسته‌بندی کننده‌ها برداشت و نظریه آماری یادگیری را به‌صورت مستحکم‌تری بنا نهاد و ماشین‌های بردار پشتیبان را بر این اساس ارائه داد. ماشین‌های بردار پشتیبان دارای خواص زیر هستند:

۱) طراحی دسته‌بندی کننده با حداکثر تعمیم

۲) رسیدن به بهینه سراسری تابع هزینه

۳) تعیین خوکار ساختار و توپولوژی بهینه برای طبقه‌بندی کننده

۴) مدل کردن توابع متمایز غیرخطی با استفاده از هسته‌های غیرخطی و مفهوم حاصل‌ضرب داخلی در فضاهای هیلبرت.

---

[1] Fisher
[2] Multi Layer Perceptron
[3] Radial Basis Function
[4] Veladimir Vapnik

۱.۲.۱ ابرصفحه بهینه برای نمونه‌های جدایی‌پذیر خطی

نمونه آموزش $\{(x_i, y_i)\}_{i=۱}^{n}$ را که در آن x_i نمونه‌ی ورودی iام و y_i پاسخ مطلوب نظیر (خروجی هدف) است را در نظر می‌گیریم. فرض می‌کنیم نمونه نمایش داده شده با $y_i = ۱$ و نمونه نمایش داده شده با $y_i = -۱$ به طور خطی جدایی‌پذیراند. معادله‌ی صفحه‌ی تصمیم به فرم ابرصفحه‌ای است که خاصیت جداسازی دارد، یعنی به‌صورت:

$$H \ : \ w^T x + b = ۰, \tag{۹.۱}$$

که در آن x بردار ورودی و w بردار وزنی قابل تنظیم و b مقدار ثابت است. در این‌صورت داریم:

$$\begin{cases} w^T x + b \geq ۰, & y_i = ۱ \text{ برای برچسب } , \\ w^T x + b \leq ۰, & y_i = -۱ \text{ برای برچسب } . \end{cases} \tag{۱۰.۱}$$

فرض نمونه‌های جدایی‌پذیر خطی برای بیان ایده‌ی اصلی یک SVM به‌کار می‌رود. برای بردار وزنی w و ثابت b، فضای جدایی‌پذیری بین ابرصفحه در معادله (۹.۱) و نزدیک‌ترین داده، **حاشیه‌ی جدایی پذیری** نامیده می‌شود که با ρ نمایش داده می‌شود. هدف از یافتن SVM یافتن ابرصفحه‌ای است که این حاشیه را ماکزیمم کند.

با هدف درک بهتر، تصویر حالتی از فضای ورودی دو بعدی را بررسی می‌کنیم. داده‌ها به‌طور خطی جدایی‌پذیر هستند و ابرصفحه‌های مختلفی می‌توانند دو کلاس را جداسازی کنند. در واقع برای $x \in \mathbb{R}^۲$ جداسازی توسط صفحه های

$$w_۱ x_۱ + w_۲ x_۲ + b = ۰,$$

انجام می‌شود. چگونه می‌توان بهترین ابرصفحه را یافت؟ از بین همه‌ی ابرصفحه‌های ممکن آن ابرصفحه‌ای را می‌یابیم که حاشیه‌ی جدا کننده‌ی دو کلاس را ماکزیمم کند. این یک روش قابل قبول شهودی است.

اگر حاشیه جداسازی را با ρ نشان دهیم، هدف SVM این است که اندازه ρ ماکزیمم شود. شکل (۳.۱) ساختار هندسی یک ابرصفحه بهینه برای یک فضای ورودی دو بعدی را شرح می‌دهد.

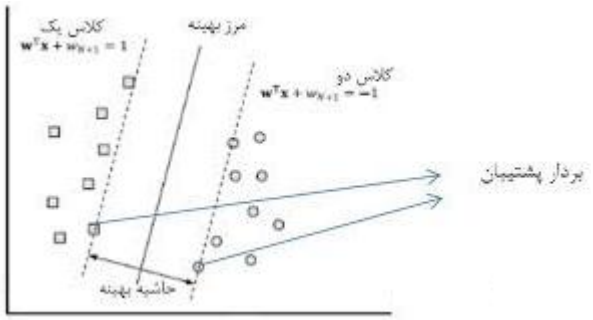
فرض کنی $w_۰$ و $b_۰$ به‌ترتیب مقادیر بهینه از بردار وزنی و مقدار ثابت باشند. ابرصفحه بهینه زیر

$$w_۰^T x + b_۰ = ۰,$$

فضای تصمیم خطی چندبعدی را در فضای ورودی نمایش می‌دهد. تابع زیر

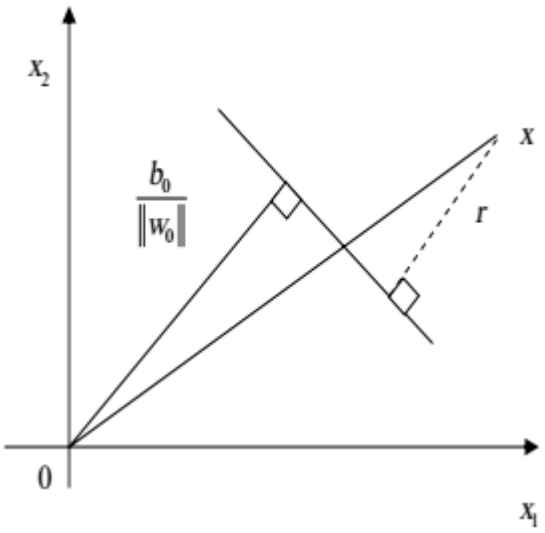
$$g(x) = w_۰^T x + b_۰,$$

یک اندازه‌ی جبری از فاصله‌ی x تا ابرصفحه بهینه را نشان می‌دهد. ساده‌ترین راه برای دیدن این مطلب ارائه x به‌صورت $x = x_\rho + r\frac{w_۰}{\|w_۰\|}$ است.



شکل ۳.۱: ابرصفحه بهینه برای نمونه‌های جدایی‌پذیر خطی

که در آن x_ρ تصویر نرمال x به توی ابرصفحه بهینه و r فاصله جبری مطلوب است. اگر x در طرف مثبت ابرصفحه بهینه باشد، r مثبت و اگر در طرف منفی ابرصفحه بهینه باشد، r منفی است. شکل (۴.۱) را ببینید. با تعریف $g(x_\rho) = 0$ نتیجه می‌شود:



شکل ۴.۱: تصویر هندسی فاصله جبری نمونه‌ها تا ابرصفحه بهینه برای حالت دو بعدی

$$g(x) = g(x_\rho + r\frac{w_\circ}{\|w_\circ\|}),$$

$$\Rightarrow g(x) = w_\circ^T x_\rho + b_\circ + r w_\circ^T \frac{w_\circ}{\|w_\circ\|},$$

$$\Rightarrow g(x) = w_\circ^T x + b_\circ = r\|w_\circ\|,$$

و یا

$$r = \frac{g(x)}{\|w_\circ\|},$$

به ویژه فاصله‌ی مبدا یعنی $(x = 0)$ تا ابرصفحه بهینه، $\frac{b_\circ}{\|w_\circ\|}$ است. اگر $b_\circ > 0$، مبدا در طرف ابرصفحه بهینه و اگر $b_\circ < 0$، مبدا در طرف منفی ابرصفحه بهینه و اگر $b_\circ = 0$، ابرصفحه بهینه از مبدا می‌گذرد.

مساله، یافتن پارامترهای $w_\circ$ و $b_\circ$ برای ابرصفحه بهینه با استفاده از مجموعه‌ی آموزشی $\{(x_i, y_i)\}_{i=1}^n$ است که زوج $(w_\circ, b_\circ)$ در محدودیت‌های زیر صدق کند،

$$\begin{cases} w_\circ^T x + b_\circ \geq \circ, & y_i = 1 \text{ برای برچسب }, \\ w_\circ^T x + b_\circ \leq \circ, & y_i = -1 \text{ برای برچسب }. \end{cases} \quad (11.1)$$

اگر معادلات (۱۰.۱) برقرار باشند، نمونه‌ها به‌طور خطی جدایی پذیراند. بردارهایی که نامعادلات (۱۱.۱) را به تساوی‌هایی به ترتیب برابر ۱ و ۱ـ تبدیل می‌کنند SV نام دارند. این بردارها نقش اساسی در کاربرد ماشین‌های یادگیری دارند. SVها داده‌هایی هستند که در نزدیک‌ترین فاصله تا صفحه تصمیم قرار می‌گیرند و طبقه بندی آن‌ها بسیار مشکل است و نسبت مستقیم با مکان صفحه تصمیم دارند. یک SV، $x^{(s)}$ با $y^{(s)} = \pm 1$ را در نظر بگیرید، داریم:

$$g(x^{(s)}) = w_\circ^T x^{(s)} \pm b_\circ = \pm 1, \quad y^{(s)} = \pm 1. \quad (12.1)$$

از معادله (۱۲.۱)، فاصله جبری SV یعنی فاصله $x^{(s)}$ تا ابرصفحه بهینه برابر

$$r = \frac{g(x^{(s)})}{\|w_\circ\|} = \begin{cases} \frac{1}{\|w_\circ\|}, & y_i = 1 \text{ برای برچسب }, \\ \frac{-1}{\|w_\circ\|}, & y_i = -1 \text{ برای برچسب }, \end{cases} \quad (13.1)$$

است. علامت $+$ نشان می‌دهد که $x^{(s)}$ در طرف مثبت ابرصفحه بهینه و علامت منفی یعنی $x^{(s)}$ در طرف منفی ابرصفحه بهینه است. مقدار بهینه حاشیه از معادله (۱۳.۱) به‌صورت زیر به‌دست می‌آید:

$$\rho = \frac{1}{\|w_\circ\|} - (\frac{-1}{\|w_\circ\|}) = \frac{2}{\|w_\circ\|}.$$

بنابراین با توجه به مطالب ارائه شده‌ی قبل مساله پیدا کردن ابرصفحه بهینه جداساز را می‌توان به شکل زیر نوشت:

$$\begin{aligned} &\text{minimize} \quad \phi(w) = \frac{1}{2}\|w\|^2, \\ &\text{subject to} \quad y_i(< w, x_i > + b) \geq 1, \quad i = 1, ..., n. \end{aligned} \quad (14.1)$$

مساله (۱۴.۱) را با استفاده از ضرایب لاگرانژ حل می‌کنیم. ابتدا تابع لاگرانژ را به‌صورت زیر در نظر می‌گیریم:

$$L(w, b, \alpha) = \frac{1}{2} w^T w - \sum_{i=1}^{n} \alpha_i [y_i (w^T x_i + b) - 1].$$

که در آن متغیرهای نامنفی α_i ضرایب لاگرانژ نامیده می‌شوند. جواب مساله بهینه سازی مقید (۱۴.۱) توسط نقطه زینی تابع لاگرانژ $L(w, b, \alpha)$ که نسبت به w و b می‌نیمم و نسبت به α

ماکزیمم می‌شود، به‌دست می‌آید. از $L(w,b,\alpha)$ نسبت به w و b مشتق گرفته و مساوی صفر قرار می‌دهیم و به شرایط بهینگی زیر می‌رسیم :

$$\frac{\partial L(w,b,\alpha)}{\partial w} = ۰, \tag{۱۵.۱}$$

$$\frac{\partial L(w,b,\alpha)}{\partial b} = ۰, \tag{۱۶.۱}$$

از شرایط بهینگی (۱۵.۱) و (۱۶.۱) داریم:

$$w - \sum_{i=۱}^{n} \alpha_i y_i x_i = ۰ \Rightarrow w = \sum_{i=۱}^{n} \alpha_i y_i x_i, \tag{۱۷.۱}$$

$$\sum_{i=۱}^{n} \alpha_i y_i = ۰. \tag{۱۸.۱}$$

در هر نقطه زینی برای هر ضریب لاگرانژ α_i داریم:

$$\alpha_i[y_i(w^T x_i + b) - ۱] = ۰, \quad i = ۱, ..., n.$$

این خواص از شرایط کان‌ـ تاکر نتیجه می‌شود. حال از روی مساله بهینه سازی مقید (مساله اولیه) مساله دیگری می‌سازیم که مساله دوگان نامیده می‌شود. این مساله همان مقدار بهینه مساله اولیه را دارد با این تفاوت که با ضرایب لاگرانژ جواب بهینه به‌دست می‌آید.

در ادامه قضیه **دوگان** را بیان می‌کنیم:

۱) شرط لازم و کافی برای آن‌که $w_۰$ جواب بهینه مساله اولیه و $\alpha_۰$ جواب بهینه مساله دوگان باشد، این است که $w_۰$ برای مساله اولیه شدنی و $\phi(w_۰) = L(w_۰, b_۰, \alpha_۰) = \min\limits_{w} L(w, b_۰, \alpha_۰)$ باشد.

برای به‌دست آوردن مساله دوگان، تابع لاگرانژ را به‌صورت زیر جمله به جمله باز نویسی می‌کنیم :

$$L(w,b,\alpha) = \frac{۱}{۲} w^T w - \sum_{i=۱}^{n} \alpha_i y_i w^T x_i - b \sum_{i=۱}^{n} \alpha_i y_i + \sum_{i=۱}^{n} \alpha_i. \tag{۱۹.۱}$$

با توجه به شرط بهینگی از معادله (۱۸.۱) جمله سوم در طرف راست معادله (۱۹.۱)، صفر است. بنابراین با توجه به معادله (۱۷.۱) داریم:

$$w^T w = \sum_{i=۱}^{n} \alpha_i y_i w^T x_i = \sum_{i=۱}^{n} \sum_{j=۱}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j.$$

به‌این ترتیب تابع هدف $L(w,b,\alpha) = Q(\alpha)$ به‌صورت زیر فرموله می‌شود:

$$Q(\alpha) = \sum_{i=۱}^{n} \alpha_i - \frac{۱}{۲} \sum_{i=۱}^{n} \sum_{j=۱}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j, \tag{۲۰.۱}$$

که در آن α_iها نامنفی هستند.

اکنون مساله دوگان را به‌صورت زیر بیان می‌کنیم:
نمونه آموزشی $\{(x_i, y_i)\}_{i=۱}^{n}$ داده شده، ضرایب لاگرانژ $\{\alpha_i\}_{i=۱}^{n}$ را به‌گونه‌ای می‌یابیم که تابع هدف

$$Q(\alpha) = \sum_{i=۱}^{n} \alpha_i - \frac{۱}{۲} \sum_{i=۱}^{n} \sum_{j=۱}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j,$$

با قیود زیر ماکزیمم شود:

$$\begin{cases} \sum_{i=۱}^{n} \alpha_i y_i = ۰, \\ \alpha_i \geq ۰, \quad i = ۱, ..., n. \end{cases}$$

۲.۲.۱ ابرصفحه بهینه برای نمونه‌های جدایی ناپذیر

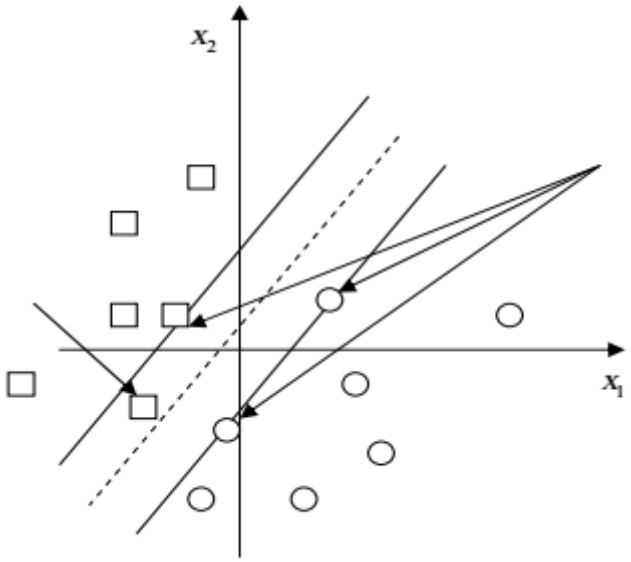
با یک مجموعه داده‌های آموزشی، ساخت ابرصفحه جداکننده بدون در نظر گرفتن خطاهای طبقه‌بندی ممکن نیست. با وجود این، ابرصفحه‌ای که خطای طبقه‌بندی را می‌نیمم کند، می‌یابیم.

در بخش قبل مساله SVM را برای حالت جدایی پذیر بررسی کردیم ولی در عمل تقریباً تمامی مسائل به‌صورت جدایی ناپذیر می‌باشند. برای حالت جدایی ناپذیر یک دسته از متغیرهای ζ_i را به‌عنوان متغیرهای کمبود تعریف می‌کنیم به‌طوری که شرایط زیر برقرار باشد:

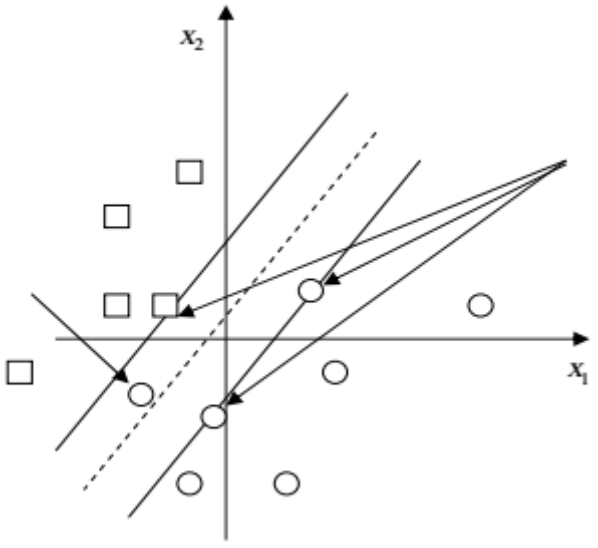
$$y_i(w^T x_i + b) \geq ۱ - \zeta_i, \tag{۲۱.۱}$$

$$\zeta_i \geq ۰, \quad i = ۱, ..., n. \tag{۲۲.۱}$$

در (۲۱.۱) و (۲۲.۱) اگر $۰ \leq \zeta_i \leq ۱$، داده x_i در داخل ناحیه جدا کننده در سمت موافق صفحه تصمیم قرار می‌گیرد(شکل (۵.۱)). و اگر $\zeta_i > ۱$، رده‌بندی اشتباه است و داده در سمت مخالف ابر صفحه جداکننده قرار می‌گیرد(شکل (۶.۱)).



شکل ۵.۱: داده x_i در داخل ناحیه جداکننده در سمت موافق صفحه تصمیم قرار دارد.



شکل ۶.۱: داده x_i در داخل ناحیه جداکننده در سمت مخالف صفحه تصمیم قرار دارد.

هم‌چنین اگر $\zeta_i = ۰$، در این حالت برای نقطه تمرینی x_i خطای طبقه‌بندی وجود ندارد و داده خارج از ناحیه جدا کننده است و به‌درستی طبقه‌بندی می‌شود.
بردارهای پشتیبان نمونه‌های خاصی هستند که در معادله‌های (۲۱.۱) و (۲۲.۱) صدق می‌کنند.
واضح است که هرچه مقادیر متغیرهای ζ_i بیشتر شود از حالت بهینه دورتر خواهیم شد و خطا بیشتر می‌شود. پس مساله بهینه‌سازی مقید را به‌صورت زیر تعریف می‌کنیم:

$$\begin{aligned} \text{minimize} \quad & \frac{۱}{۲}\|w\|^۲ + C\sum_{i=۱}^{n}\zeta_i, \\ \text{subject to} \quad & y_i(< w, x_i > + b) \geq ۱ - \zeta_i, \quad i = ۱, ...n, \\ & \zeta_i \geq ۰, \quad i = ۱, ..., n, \end{aligned}$$

که در آن C پارامتر جریمه است.
تابع لاگرانژ این مساله به‌صورت زیر است:

$$L_P(w, b, \zeta, \alpha, \beta) = \frac{۱}{۲}w^T w + C\sum_{i=۱}^{n}\zeta_i - \sum_{i=۱}^{n}\alpha_i[y_i(< w, x_i > + b) - ۱] - \sum_{i=۱}^{n}\beta_i\zeta_i.$$

شرط لازم و کافی برای بهینگی به‌صورت زیر است:

$$\begin{aligned} \frac{\partial L}{\partial w} = ۰ &\Rightarrow w = \sum_{i=۱}^{n}\alpha_i y_i x_i, \\ \frac{\partial L}{\partial b} = ۰ &\Rightarrow \sum_{i=۱}^{n}\alpha_i y_i = ۰, \\ \frac{\partial L}{\partial \zeta_i} = ۰ &\Rightarrow C - \alpha_i - \beta_i = ۰. \end{aligned}$$

برای این مساله نیز مشابه قبل، مساله دوگان به‌صورت زیر است:

$$\begin{aligned} \text{maximize} \quad & -\frac{۱}{۲}\sum_{i=۱}^{n}\sum_{j=۱}^{n}\alpha_i\alpha_j y_i y_j x_i^T x_j + \sum_{i=۱}^{n}\alpha_i, \\ \text{subject to} \quad & \sum_{i=۱}^{n}\alpha_i y_i = ۰, \\ & ۰ \leq \alpha_i \leq C, \quad i = ۱, ..., n. \end{aligned}$$

همان‌طور که ملاحظه شد، حل مساله SVM در حالت جدایی ناپذیر مشابه حل آن در حالت جدایی پذیر است. با این تفاوت که محدوده تغییرات ضرایب لاگرانژ α_i، فرق می‌کند.
پس از به‌دست آوردن ضرایب لاگرانژ α_i، نمونه‌هایی که ضرایب لاگرانژ آن‌ها در روابط $۰ < \alpha_i \leq C$ صدق می‌کنند، **بردار پشتیبان** می‌باشند. بردارهای پشتیبان با $\alpha_i = C$ در جهت مخالف صفحه تصمیم قرار دارند. مقدار w و شکل تابع جداکننده مشابه حالت جدایی پذیر خواهد بود.

۳.۲.۱ رده‌بندی کننده‌های غیرخطی

در بخش‌های گذشته مشاهده کردیم که چگونه رده‌بندی کننده‌های خطی در داده‌های ورودی به آسانی توسط تکنیک‌های بهینه‌سازی استاندارد محاسبه می‌گردند. با این‌که کار کردن با رده‌بندی کننده‌های خطی آسان است اما آن‌ها محدودیت‌های شدیدی بر کار آموزش می‌گذارند.

در بیشتر مسائل زندگی داده‌ها به‌طور خطی جدا پذیر نیستند. در این موارد روش حل، استفاده از تبدیلات غیرخطی است. به‌عنوان مثال ایجاد یک بردار با بعد بالاتر به‌وسیله ضرب همه بردارها در بردار ویژگی و نگاشتی که داده‌ها را از فضای ورودی به فضایی با بعد بالاتر می‌برد، که یک تابع هسته غیرخطی است. ماشین‌های بردار پشتیبان ذکر شده در بخش‌های قبل، برای دسته‌بندی الگوهای یک مساله دوکلاسه، از مرزهای جداکننده خطی و از یک ابرصفحه استفاده می‌کند و در واقع حاصل‌ضرب داخلی بردار ورودی با هر کدام از بردارهای پشتیبان در فضای n بعدی ورودی محاسبه می‌گردد.

وپنیک با استفاده از مفهوم ضرب داخلی در فضاهای هیلبرت و قضیه هیلبرت‌ـاشمیت نشان داد که ابتدا می‌توان بردار ورودی x را با یک تبدیل غیرخطی به یک فضای با بعد زیاد انتقال داد و در آن فضا حاصلضرب داخلی را انجام داد و ثابت کرد که اگر یک هسته متقارن، شرایط قضیه مرسر[1] را داشته باشد، اعمال این هسته در فضای ورودی با بعد کم می‌تواند به‌عنوان حاصلضرب داخلی در یک فضای هیلبرت با بعد زیاد تلقی شود و محاسبات را به شدت کاهش دهد [۲].

تصمیم نهایی ما به‌عنوان یک ترکیب خطی از صفات داده شده ممکن است بسیار مشکل باشد. ایده کلی به این‌صورت است که اطلاعات ورودی را به فضایی با بعد بالاتر می‌نگاریم و سپس به کمک یک رده‌بندی کننده خطی نقاط تمرینی را جدا می‌کنیم، که این رده بندی کننده خطی یک جداکننده غیرخطی در فضای ورودی را نتیجه می‌دهد. به‌وجود آمدن این تکنیک به دهه ۶۰ برمی‌گردد و برای اولین بار در مقاله‌ای از آیزرمن[2] مطرح شده است.

نحوه نمایش نمونه‌های آموزشی را با نگاشت به یک فضای بی‌نهایت بعدی مانند فضای هیلبرت انتقال می‌دهیم. معمولاً H بعد بسیار بالاتری نسبت به فضای اولیه H دارد (L و H به ترتیب به‌عنوان مخففی از کمترین و بیشترین بعد به‌کار می‌روند). این نگاشت قبل از یادگیری بر هر کدام از مثال‌ها اعمال می‌شود.

ابرصفحه جداکننده بهینه در فضای H ساخته می‌شود. iامین عضو $\Phi(x_i)$ مشخصه iام تحت نگاشت $\Phi : \mathbb{R}^n \to H$ ویژگی فضا خوانده می‌شود. عمل انتخاب کردن بهترین نمایش اطلاعات نیز انتخاب مشخصه نامیده می‌شود.

[1] Mercer
[2] Aizerman

۴.۲.۱ هسته ضرب داخلی

تعریف ۱.۲.۱. تعریف تابع کرنل (هسته ضرب داخلی) :
فرض کنید نگاشت $\Phi : L \to H$ داده‌های فضای ورودی L را به فضای مشخصه H می‌برد. تابع $K : L \times L \to H$ را یک تابع **کرنل** گویند اگر و تنها اگر داشته باشیم :

$$K(x, z) =< \Phi(x), \Phi(z) > .$$

عملکرد تابع کرنل مانند یک ضرب داخلی در H است که می‌تواند مانند یک تابع در فضای اولیه L ارزیابی شود. در بخش قبل ما با معرفی نگاشت Φ قبل از اجرای الگوریتم یادگیری آغاز کردیم. حال برعکس شروع کرده و با انتخاب یک کرنل مانند K آغاز می‌کنیم که خود به‌طور غیرمستقیم یک نگاشت مانند Φ را معرفی خواهد کرد.

فرض کنید x برداری از فضای ورودی با بعد m_0 است و $\{\Phi_j(x)\}_{j=1}^{m_1}$ یک مجموعه از تبدیلات غیرخطی از فضای ورودی به فضای ویژگی است. با وجود این تبدیلات غیرخطی، ما می‌توانیم ابرصفحه‌ای به‌صورت زیر تعریف کنیم که مانند یک ابرصفحه تصمیم عمل می‌کند:

$$\sum_{j=1}^{m_1} w_j \Phi_j(x) + b = 0, \qquad (۲۳.۱)$$

که در آن $\{w_j\}_{j=1}^{m_1}$ مجموعه‌ای از وزن‌های خطی است که فضای ویژگی را به فضای خروجی مرتبط می‌کند و b یک مقدار ثابت است . معادله (۲۳.۱) را به‌صورت خلاصه زیر می‌نویسیم:

$$\sum_{j=0}^{m_1} w_j \Phi_j(x) = 0,$$

که در آن $\Phi_0(x) = 1$ برای هر x و لذا w_0 مقدار ثابت b را نشان می‌دهد. $\Phi_j(x)$ نظیر وزن w_j در فضای تصویر است. $\Phi(x)$ تصویر تولید شده نظیر بردار ورودی x است.

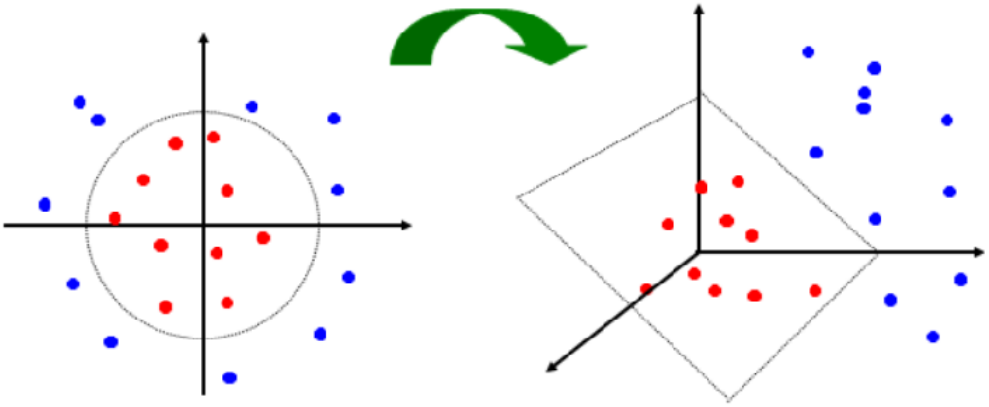
با تعریف $\Phi(x)$ به‌صورت $\Phi(x) = (\Phi_0(x)\Phi_1(x)...\Phi_{m_1}(x))$ که بنا به فرض $\Phi_0(x) = 1$، بردار $\Phi(x)$ تصویر القا شده بردار ورودی x در فضای ویژگی است، همان‌طور که در شکل (۷.۱) نشان داده شده است.

بنابراین می‌توانیم صفحه تصمیمی به شکل

$$w^T \Phi(x) = 0, \qquad (۲۴.۱)$$

تعریف کنیم. با تطبیق فرمول (۱۷.۱) به موقعیت فعلی که اکنون انتظار داریم ترکیبی با قابلیت جدایی پذیری خطی داشته باشد، می‌توان نوشت:

$$w = \sum_{i=1}^{n} \alpha_i y_i \Phi(x_i), \qquad (۲۵.۱)$$



شکل ۷.۱: انتقال مساله از فضای دو بعدی به فضای سه بعدی

که بردار ویژگی $\Phi(x_i)$ متناظر الگوی ورودی x_i در نمونه iاُم است. بنابراین با جای‌گذاری معادله (۲۵.۱) در (۲۴.۱) می‌توانیم صفحه تصمیم محاسبه شده در فضای ویژگی را به‌صورت زیر تعریف کنیم:

$$\sum_{i=1}^{n} \alpha_i y_i \Phi^T(x_i)\Phi(x) = 0. \qquad (۲۶.۱)$$

$\Phi^T(x_i)\Phi(x)$، ضرب داخلی دو بردار در فضای ویژگی توسط بردار ورودی x و الگوی ورودی x_i برای iاُمین نمونه است.

هسته ضرب داخلی $K(x, x_i)$ به‌صورت زیر تعریف می‌شود:

$$K(x, x_i) = \Phi^T(x_i)\Phi(x) = \sum_{i=1}^{n} \Phi_j(x)\Phi_j(x_i), \quad i = 1, ..., n. \qquad (۲۷.۱)$$

از این تعریف مشاهده می‌شود که هسته ضرب داخلی تابعی متقارن است:

$$K(x, x_i) = K(x_i, x), \quad \forall i. \qquad (۲۸.۱)$$

از هسته ضرب داخلی برای ساخت ابرصفحه بهینه در فضای ویژگی بدون در نظر گرفتن شکل ضمنی فضای ویژگی استفاده می‌کنیم. از معادلات (۲۶.۱) و (۲۷.۱) داریم:

$$\sum_{i=1}^{n} \alpha_i y_i K(x, x_i) = 0. \qquad (۲۹.۱)$$

تابع تصمیم در H به‌صورت زیر است:

$$f(x) = sign(\sum_{i=1}^{n} \alpha_i y_i < \Phi(x_i), \Phi(x) > + b).$$

تاکنون دیدیم که در دو مرحله یادگیری و آموزش تابع مشخصه فقط به مقدار ضرب داخلی در فضای ویژگی بستگی دارد. بنابراین آن‌ها می‌توانند با استفاده از توابع کرنل محاسبه شوند. لذا تابع تصمیم فوق را می‌توان به‌صورت زیر نوشت:

$$f(x) = sign(\sum_{i=1}^{n} \alpha_i y_i K(x, x_i) + b).$$

در نتیجه لازم نیست که نگاشت مشخصه را برای حل مساله یادگیری در فضای مشخصه بدانیم.

چه توابعی را می‌توان به عنوان تابع کرنل انتخاب کرد؟ به آسانی قابل مشاهده است که تابع هسته ضرب داخلی K متقارن است:

$$K(x, z) = \Phi(x)\Phi(z) = \Phi(z)\Phi(x) = K(z, x),$$

و در نامساوی کشی‌ـشوارتز هم صدق می‌کند:

$$K(x, z)^2 =< \Phi(x), \Phi(z) >^2 \leq \|\Phi(x)\|^2 \|\Phi(z)\|^2$$

$$=< \Phi(x), \Phi(x) >< \Phi(z), \Phi(z) >= K(x, x)K(z, z).$$

علاوه‌براین باید در شرایط قضیه مرسر نیز صدق کند.

به‌عنوان مثال تابع هسته کرنل می‌تواند به فرم‌های زیر باشد:

هسته چندجمله‌ای	$K(x, y) = (x.y + 1)^p, \quad p = 2, 3, ...$
هسته گاوسی	$K(x, y) = exp(-\frac{\|x - y\|^2}{2\delta^2}),$
هسته تانژانت هیپربولیک	$K(x, y) = tanh(\beta_0 x.y + \beta_1),$

فصل ۲

شبکه عصبی

شبکه‌های عصبی مصنوعی یا شبکه‌های عصبی صناعی یا به زبان ساده‌تر شبکه‌های عصبی، سیستم‌ها و روش‌های محاسباتی نوین برای یادگیری ماشینی، نمایش دانش و در انتها اعمال دانش به دست آمده در جهت بیش‌بینی پاسخ‌های خروجی از سامانه‌های پیچیده هستند. ایده‌ی اصلی این گونه شبکه‌ها تا حدودی الهام‌گرفته از شیوه‌ی کارکرد سیستم عصبی زیستی برای پردازش داده‌ها و اطلاعات به منظور یادگیری و ایجاد دانش می‌باشد. عنصر کلیدی این ایده، ایجاد ساختارهایی جدید برای سامانه‌ی پردازش اطلاعات است.

در این فصل به‌طور خلاصه به معرفی شبکه عصبی و کاربرد آن در بهینه‌سازی می‌پردازیم.

۱.۲ مقدمه‌ای بر شبکه‌های عصبی

۱.۱.۲ شبکه‌های عصبی بیولوژیکی

واحد اصلی دستگاه عصبی، سلولی خاص به‌نام نرون می‌باشد و بدون شک نحوه کار مغز و راز شعور آدمی در آن نهفته است. نرون‌ها با وجود تفاوت‌های زیاد از نظر اندازه و شکل ظاهری، مشخصه‌های مشترکی دارند. همان‌طور که در شکل (۱.۲) نشان داده شده است از تنه سلولی نرون تعدادی شاخک کوتاه خارج می‌شود که دندریت نام دارند. دندریت‌ها و تنه سلولی، سیگنال‌ها را از نرون‌های مجاور دریافت می‌کنند و از طریق یک لوله باریک به نام آکسون به نرون‌های دیگر منتقل می‌گردد.

آکسون در انتهای خود به تعدادی رشته جانبی باریک تقسیم می‌گردد که پایانه‌های آکسونی نام دارد و با دندریت‌های سایر نرون‌ها مرتبط است. ارتباط بین آکسون یک نرون و دندریت نرون دیگر را سیناپس می‌نامند. یک سیناپس مرکب از پایانه قبل از سیناپسی، شکاف سیناپسی و پایانه بعد از سیناپسی می‌باشد. همه سیگنال‌های جمع‌بندی شده، در تنه نرون ترکیب می‌شوند و اگر وسعت سیگنال‌های ترکیب شده به آستانه نرون برسد، مرحله تحریک شدن فعال می‌شود و یک سیگنال خروجی تولید می‌شود. این سیگنال به‌صورت یک پالس منفرد یا بخشی از پالس‌ها در یک میزان خاص به موازات آکسون به پایانه‌های سیناپسی انتقال می‌یابد و موجب ترشح موادی به‌نام عصب‌ـ رسانه می‌شود. عصب‌ـ رسانه در داخل شکاف سیناپسی پخش می‌شود و نرون بعدی را تحریک می‌کند و به این ترتیب یک سیگنال از یک نرون به دیگری انتقال می‌یابد. تعداد بسیار زیادی آکسون از نرون‌های مختلف با دندریت‌های یک نرون به این‌صورت ارتباط برقرار می‌کنند.

نرون‌ها با توجه به عملکردشان به سه دسته تقسیم می‌شوند:

۱) **نرون‌های حسی:** این نرون‌ها پیام‌ها را از گیرنده‌ها به دستگاه عصبی مرکزی منتقل می‌کنند. گیرنده سلول خاصی است که در اندام‌های حسی، عضلات، پوست و ... قرار دارد و تغییرات فیزیکی یا شمیایی را در می‌یابد، این تغییرات را به سیگنال‌های عصبی تبدیل می‌کند.

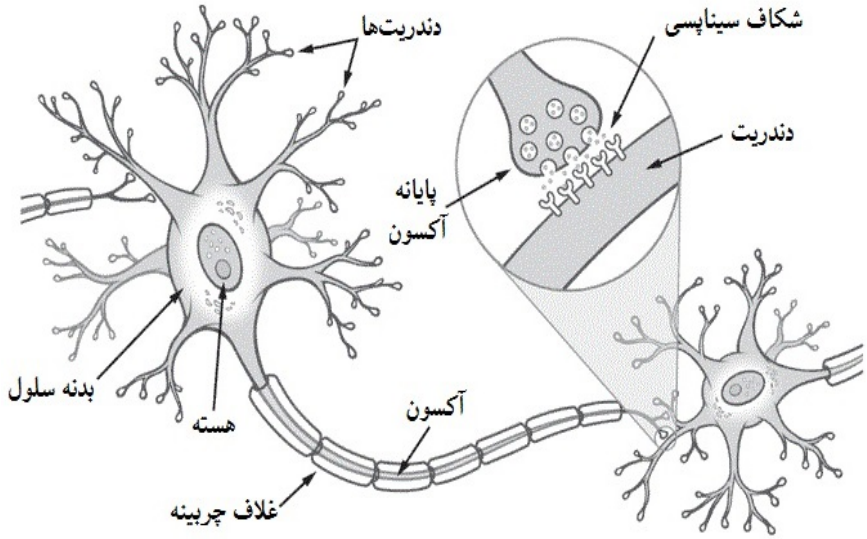
۲) **نرون‌های حرکتی:** این نرون‌ها پیام‌ها را از مغز یا طناب نخاعی به اندام‌های عمل کننده می‌رسانند.

۳) **نرون‌های رابط:** این نرون‌ها پیام‌ها را از نرون‌های حسی می‌گیرند و به یک نرون رابط دیگر انتقال می‌دهند و یا پیام‌ها را به یک نرون حرکتی می‌رسانند. نرون‌های رابط فقط در مغز، چشم و طناب نخاعی یافت می‌شوند.

حال منظور از یادگیری عبارت است از تغییری نسبتاً ساختاری در سیناپس (بعد از تحریک)، که این تغییر ساختاری باعث افزایش کارایی سیناپس می‌شود.

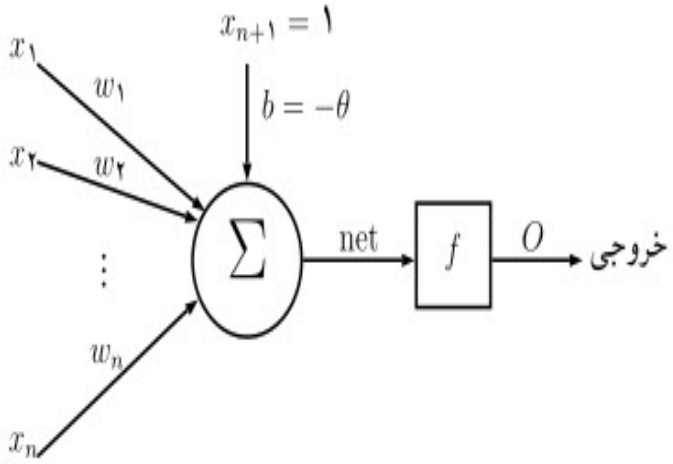
۲.۱.۲ شبکه‌های عصبی مصنوعی

بعد از اشاره‌ای گذرا بر شبکه عصبی بیولوژیکی و نشان دادن گوشه‌ای بسیار کوچک از این جهان مرموز و فوق‌العاده پیچیده به موضوع اصلی یعنی شبکه‌های عصبی مصنوعی می‌پردازیم. شبکه‌های عصبی مصنوعی مدل‌های سخت افزاری یا نرم افزاری الهام گرفته از ساختار و رفتار نرون‌های بیولوژیکی و سیستم‌های عصبی انسان هستند که شامل عناصر پردازشی (به نام نرون‌ها) و ارتباطات میان آنها با ضرایب (وزن‌ها) اختصاص یافته به ارتباطات است، که ساختار شبکه عصبی را تشکیل می دهند. در این رساله ما شبکه‌های عصبی مصنوعی را یک مدل



شکل ۱.۲: ساختار یک نرون.

ریاضی فرض خواهیم کرد و به‌جای استفاده از عبارت شبکه‌های عصبی مصنوعی به‌طور خلاصه عبارت شبکه‌های عصبی را به‌کار خواهیم برد.

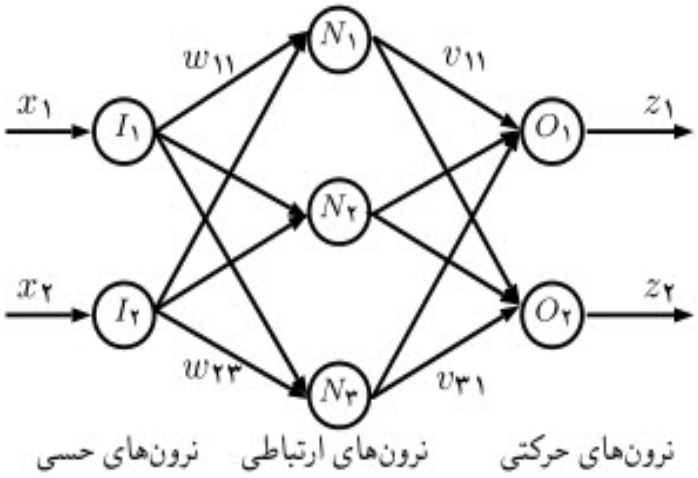


شکل ۲.۲: مدل ریاضی یک نرون.

مدل شبیه‌سازی شده یک نرون، در شکل (۲.۲) نمایش داده شده است. یک نرون در حالت کلی مجموعه‌ای از $n+1$ ورودی x_j $(j = 1,\ldots,n+1)$ دارد. این ورودی‌ها همان سیگنال‌های عصبی می‌باشند که از محیط یا از یک نرون دیگر فرستاده می‌شوند. w_j $(j = 1,\ldots,n+1)$ که روی آکسون‌ها واقع هستند نقش ساختار سیناپسی را ایفا می‌کنند. این‌ها هستند که باید تغییر کنند تا فرآیند یادگیری اتفاق بیفتد حال این‌که این تغییر بر چه قاعده‌ای استوار باشد بستگی به الگوریتم آموزش یا تعلیم دارد. وزن آخرین ورودی یعنی $-\theta$ را با b نشان می‌دهند و آن‌را بایاس می‌نامند. سیگنال‌های عصبی x_1 و x_2 و ... و x_{n+1} به‌ترتیب در w_1، w_2، ... ، w_{n+1} ضرب شده و مجموع این حاصل‌ضرب‌ها که در شکل آن را با net نشان می‌دهیم، وارد نرون می‌شوند. در این مرحله نرون همانند یک تابع روی net عمل می‌کند و

حاصل کار خود (برد تابع) را از طریق آکسون‌های طرف راست به نرون‌های دیگر می‌فرستد. این تابع را تابع فعال‌سازی می‌نامند.

البته خاطر نشان می‌کنیم که اگر نرون به‌عنوان تابع فرض شود این تابع از $\mathbb{R}^n$ به $\mathbb{R}$ تعریف می‌گردد. یعنی سیگنال خروجی از نرون در تمام آکسون های طرف راست یکسان می‌باشد که روی شکل با O نمایش داده‌ایم. همان‌طور که قبلاً اشاره شد نرون‌ها می‌توانند حسی، ارتباطی و یا حرکتی باشند. اطلاعات از طریق نرون‌های حسی دریافت می‌شود و از طریق نرون‌های ارتباطی انتقال می‌یابد و بالاخره نتیجه کار شبکه عصبی و یا واکنش شبکه عصبی به اطلاعات دریافت شده از طریق نرون‌های حرکتی به خارج منتقل می‌گردد(شکل ۳.۲).



شکل ۳.۲: ساختار یک نرون با لایه ورودی ، لایه پنهان و لایه خروجی.

نرون‌های حسی را لایه ورودی، نرون های حرکتی را لایه خروجی و نرون‌های ارتباطی را لایه پنهان نیز می گویند. لایه مخفی می تواند از چندین لایه تشکیل گردد البته در شکل (۳.۲) لایه مخفی تک لایه در نظر گرفته شده است.

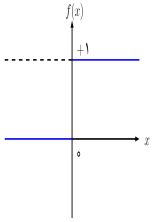
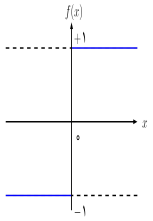
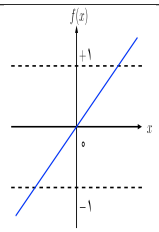
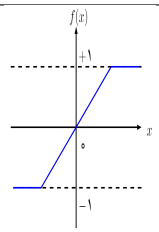
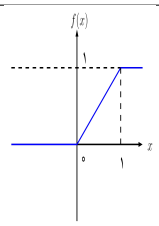
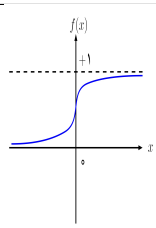
مشخصه‌های اصلی یک شبکه عصبی که در طراحی آن باید مدنظر قرار گیرند عبارت‌اند از

معماری شبکه عصبی: نحوه اتصالات بین نرون‌ها، تعداد آن‌ها و تعداد لایه‌های تشکیل دهنده بخش نرون‌های ارتباطی را معماری شبکه عصبی می‌گوییم.

تابع‌های فعال‌سازی: این‌که چه تابعی روی نرون قرار بگیرد تا بر ورودی های نرون اثر کرده و خروجی نرون را تولید کند تابع فعال‌سازی آن نرون گویند. در جدول (۱.۲) تعدادی از مهم‌ترین توابع فعال‌سازی نشان داده شده است.

الگوریتم آموزش: روشی که وزن‌های روی آکسون‌ها بر اساس آن تغییر می‌کنند (w_{ij}ها و v_{jk}ها در شکل (۳.۲)) تا خروجی شبکه عصبی به حالت مطلوب درآید را الگوریتم آموزش می‌گوییم.

جدول ۱.۲: برخی توابع فعال‌سازی نرون‌ها

ردیف	نام تابع	تعریف تابع	شکل تابع
۱	آستانه‌ای دو مقداره	$f(x)=\begin{cases}۰ & x<۰,\\ ۱ & x\geq ۰.\end{cases}$	
۲	آستانه‌ای دو مقداره متقارن	$f(x)=\begin{cases}-۱ & x<۰,\\ ۱ & x\geq ۰.\end{cases}$	
۳	خطی	$f(x)=x.$	
۴	آستانه‌ای خطی متقارن	$f(x)=\begin{cases}-۱, & x<-۱,\\ x, & -۱\leq x\leq ۱,\\ ۱, & x>۱.\end{cases}$	
۵	آستانه‌ای خطی	$f(x)=\begin{cases}۰, & x<۰,\\ x & ۰,\leq x\leq ۱,\\ ۱, & x>۱.\end{cases}$	
۶	سیگموئید	$f(x)=\frac{۱}{۱+e^{-x}}.$	

با توجه به مطالب قبل می‌توان برای شبکه‌های عصبی مصنوعی خصوصیات عمومی زیر را نسبت داد:

۱. پردازش اطلاعات در عناصر بسیار ساده به نام نرون (سلول عصبی) انجام می‌گیرد.

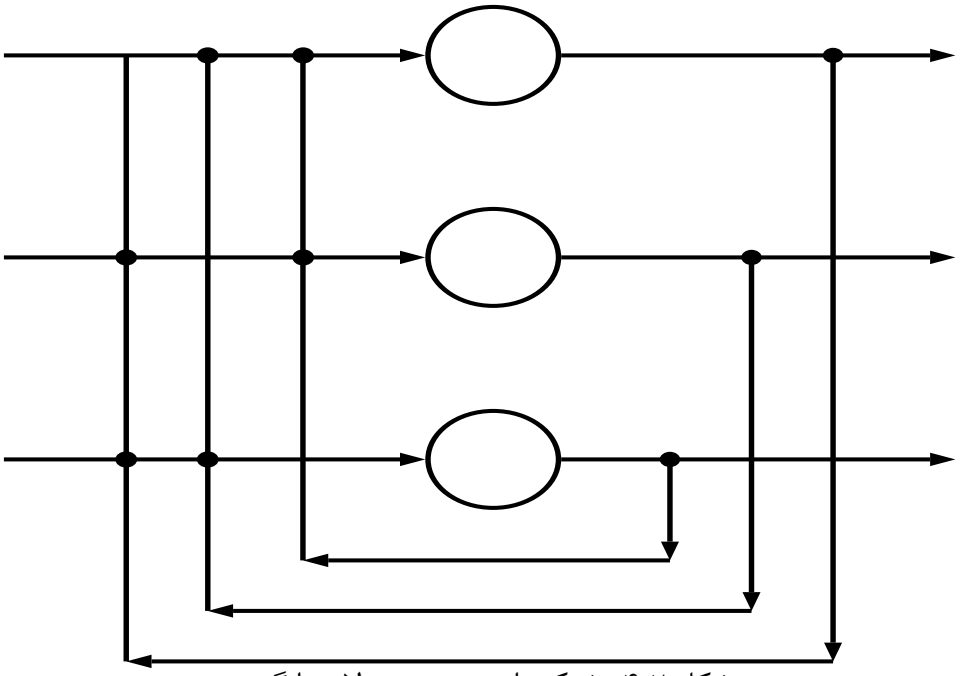
۲. سیگنال‌های عصبی بین سلول‌های عصبی از طریق اتصالات بین آن‌ها مبادله می‌شوند.

۳. به هر خط اتصال (آکسون) یک وزن نسبت داده می‌شود که در شبکه‌های عصبی معمولی این وزن ها در سیگنال‌های منتقل شده ضرب می گردند.

۴. هر نرون دارای یک تابع فعال‌سازی می‌باشد که این تابع در اکثر موارد غیرخطی است.

شبکه‌های عصبی بازگشتی

در شبکه های عصبی بازگشتی، حداقل یک سیگنال برگشتی از یک نرون به همان نرون یا نرون‌های همان لایه و یا لایه قبل وجود دارد. چنین شبکه ای حافظه‌ای را نگه می دارد و حالت بعد نه تنها به سیگنالهای ورودی بلکه به حالات قبل شبکه نیز وابسته است. این شبکه‌های عصبی معمولاً در مدل ریاضی به‌صورت معادلات دیفرانسیل مرتبه اول ظاهر می‌شوند.



شکل ۴.۲: شبکه‌های عصبی دو لایه بازگشتی.

۳.۱.۲ تاریخچه تکامل شبکه عصبی

در سال ۱۹۴۳ یک نرولوژیست به نام وارن مک کلاچ[1] به همراه یک متخصص آمار به نام والتر پیتز[2] اولین مدل ریاضی نرون را ارائه دادند [۲۹]. این نرون ساده یک عنصر محاسبه‌گر بود که عناصری به نام ورودی را در مقادیر ثابتی به نام وزن ضرب می‌کرد و آنها را از یک عملگر خطی عبور می‌داد و حاصل خروجی نرون را تشکیل می‌داد. در این مدل وزن‌ها مقادیر ثابتی

[1] Warren McCulloch
[2] Walter Pitts

بودند. در سال ۱۹۴۹ دونالد هب[1] [۶] کتابی درباره نحوه یادگیری در مغز انسان نوشت. بر این مبنا قانون یادگیری برای نرون را ارائه داد که به قانون هب معروف است و بعدها در تعلیم شبکه‌های عصبی از آن استفاده شد. در سال ۱۹۵۴ فارلی و کلارک[2] [۴] مدلی از نرون‌ها را به‌طور تصادفی به هم وصل کردند و قانون هب را پیاده‌سازی نمودند و نشان دادند که می‌توان دو الگوی ورودی را با این قانون تعلیم از هم تشخیص داد. با پیشرفت علم کامپیوتر اولین مدل از نرون عصبی مصنوعی توسط ناتانیال روچستر[3] محقق شرکت IBM در سال ۱۹۵۸ با استفاده از کامپیوتر شبیه‌سازی شد. روزنبلات[4] [۴۲، ۴۳] در سال ۱۹۵۸ نرون ساده مک‌کلاچ پیتز را اصلاح کرد و به آن قابلیت یادگیری وسازگاری را اضافه کرد و این نرون را پرسپترون نامید. قانون تعلیم پرسپترون اولین قانون رسمی برای شبکه‌های عصبی است.

برنارد ویدرو[5] و مارسین هاف[6] و همکارانشان یک مدل سه سطحی از نرون‌ها را ابداع کردند که آدلاین نام‌گذاری شد و مختصر شده حروف" عنصر خطی تطبیقی" بود [۷، ۵۲–۵۴] و برای آدلاین یک قانون تعلیم بر مبنای مشتق گیری بیان نمودند. این قانون تعلیم به $\alpha - LMS$ معروف است. از متصل کردن چند آدلاین به یکدیگر مادلاین ساخته شد. ویدرو و هاف از شبکه عصبی مادلاین برای حذف اکو از خطوط تلفن استفاده نمودند.

سال ۱۹۶۹ آغاز یک دوره رکود برای شبکه عصبی بود. پاپرت و مینسکی[7] [۳۱] کتابی به نام پرسپترون نوشتند و نشان دادند که اگرچه گیت‌های OR و AND را می‌توان با استفاده از پرسپترون پیاده سازی کرد ولی گیت XOR قابل پیاده سازی نمی‌باشد و لذا نمی‌توان از شبکه عصبی برای پردازش اطلاعات استفاده کرد. الگوریتم پس انتشار پال ورباس[8] [۵۱] در سال ۱۹۷۴ ارائه شد و بعدها توسط رامل هارت[9] [۴۴] به‌طور مستقل کشف گردید. این الگوریتم از زمان پیدایش به‌طور گسترده به‌عنوان یک الگوریتم آموزش در شبکه‌های عصبی پیش‌خورد مورد استفاده قرار گرفته است. سال ۱۹۸۲ تا ۱۹۸۶ را می‌توان تولد دوباره شبکه عصبی دانست. در این سال‌ها دو اتفاق مهم در این زمینه رخ داد. اول ارائه شبکه عصبی بازگشتی و مفهوم تابع انرژی توسط جان هاپفیلد[10] [۲۰] بود که در سال ۱۹۸۲ منتشر شد. دومین اتفاق، حل مسأله فروشنده دوره‌گرد (TSP) توسط شبکه هاپفیلد بود [۲۱]. هاپفیلد نشان داد که از این شبکه بازگشتی می‌توان برای حل مسائل بهینه‌سازی استفاده کرد و افق جدیدی را در شبکه‌های عصبی مصنوعی گشود.

---

[1] Donald Hebb
[2] Farley and Clark
[3] Nathanial Rochester
[4] Frank Rosenblatt
[5] Bernard Widrow
[6] Marcian Hoff
[7] Papert and Minsky
[8] Paul Werbos
[9] Rumelhart
[10] John Hopfield

۴.۱.۲ تاریخچه حل مسائل بهینه‌سازی با استفاده از شبکه‌های عصبی

طیف گسترده‌ای از مسائل مهم در رشته های علوم پایه و مهندسی از جمله کنترل بهینه، طراحی ساختار بهینه، پردازش تصویر، تقریب توابع، تحلیل رگرسیون و... قابل تبدیل به مسائل بهینه‌سازی خطی و غیرخطی هستند. در دهه‌های گذشته الگوریتم‌های بهینه‌سازی عددی فراوانی برای حل مسائل بهینه‌سازی خطی و غیرخطی ارائه شده‌اند. از آن‌جا که زمان مورد نیاز برای حل مسائل بهینه‌سازی و به‌خصوص مسائل بهینه‌سازی غیرخطی بسیار وابسته به بعد و ساختار مسأله است بنابراین الگوریتم‌های عددی کارایی کمی از خود نشان می‌دهند. یک رهیافت امیدوار کننده برای حل مسائل بهینه‌سازی با ابعاد بالا استفاده از شبکه‌های عصبی مصنوعی می‌باشد.

در سال ۱۹۸۶ تانک و هاپفیلد [۴۶] اولین شبکه عصبی را برای حل مسائل برنامه‌ریزی خطی معرفی کردند و نشان دادند که وضعیت شبکه در هر تکرار به‌گونه‌ای تغییر می‌کند که تابع انرژی متناظر با آن به‌طور یکنواخت کاهش می‌یابد تا جایی که به نقطه مینیمم خود می‌رسد و این نقطه مینیمم متناظر با نقطه تعادل شبکه عصبی می‌باشد. آن‌ها این شبکه عصبی را توسط یک مدار الکتریکی پیاده سازی کردند. هم‌چنین از این شبکه برای حل مسأله فروشنده دوره‌گرد با ۳۰ شهر استفاده نمودند. این شبکه دارای نقص‌هایی بود به‌خصوص این‌که نقطه تعادل شبکه در شرایط کاروش کان تاکر[1] صدق نمی‌کرد و لذا جواب مطلوبی از مسأله حاصل نمی‌شد. با این وجود کارهای هاپفیلد انگیزه بسیار خوبی را برای محققین به‌وجود آورد تا در این زمینه فعالیت کنند. کندی و چا[2] [۲۲] با افزودن یک پارامتر جریمه متناهی به شبکه هاپفیلد آن را توسعه دادند و از آن برای حل مسأله برنامه‌ریزی غیرخطی محدب استفاده نمودند. پارامترهای جریمه این شبکه در یافتن نقطه بهینه دقیق ناتوان است به‌خصوص اگر پارامتر جریمه بزرگ باشد این شبکه به‌سختی عمل می‌کند.

برای جلوگیری از بکار بردن پارامتر جریمه یک شبکه عصبی سوئیچ‌ـ‌خازن توسط رودریگز‌ـ‌وازکز[3] و همکاران [۴۱] معرفی شد. هم‌چنین ما و شانبلات[4] [۲۷] یک شبکه عصبی دو فازی ارائه کردند که فاز اول شبکه مشابه شبکه کندی و چا بود و در فاز دوم آن مسیر شبکه به جواب دقیق مسأله همگرا می‌شد بنابراین این روش جواب‌های دقیق‌تری نسبت به شبکه کندی و چا ارائه می‌داد. مشکل این شبکه در این بود که پایداری فاز دوم شبکه بستگی به انتخاب یک مقدار بزرگ از پارامتر جریمه داشت و لذا اگرچه تأثیر پارامتر جریمه در این روش کاهش یافته بود و جواب‌های حاصل دقیق‌تر بودند اما هنوز مستقل از پارامتر نبود.

ژانگ و کنستانتینیدس[5] [۶۸] برمبنای روش لاگرانژ یک شبکه عصبی ارائه کردند که کاملاً مستقل از پارامتر جریمه و قادر به حل مسائل غیرخطی بود. نقطه تعادل این شبکه عصبی در

[1] Karush-Kuhn-Tucker(KKT)
[2] Kennedy and Chua
[3] Rodriguez-Vazquez
[4] Maa and Shanblatt
[5] Zhang and Constantinides

شرایط بهینگی مرتبه اول و دوم صدق می‌کرد و هم‌چنین شبکه حاصل همگرا بود. در سال ۱۹۹۳ بوزردوم و پتیسن[1] [۱۲] شبکه‌ای را برمبنای گرادیان و روش تصویر و مستقل از پارامتر جریمه ابداع کردند که تنها قادر به حل مسائل درجه دوم با متغیرهای کران‌دار بود. این روش در عمل روش کارایی بود اما نمی‌توانست مسائل کلی برنامه‌ریزی خطی و درجه دوم را حل کند.

زیا[2] و همکاران [۵۷، ۶۲–۶۶] چندین مدل را برای حل مسائل بهینه‌سازی غیرخطی با قیود خطی و غیرخطی ارائه کردند. هم‌چنین چندین مدل شبکه عصبی برای حل مسائل بهینه‌سازی غیرخطی توسط عفتی و همکاران [۱۴، ۱۵، ۱۷، ۳۲، ۳۳] ارائه شده است.

[1] Bouzerdoum and Pattison
[2] Xia

فصل ۳

حل مسئله رگرسیون بردار پشتیبان با استفاده شبکه عصبی

۱.۳ رگرسیون چیست؟

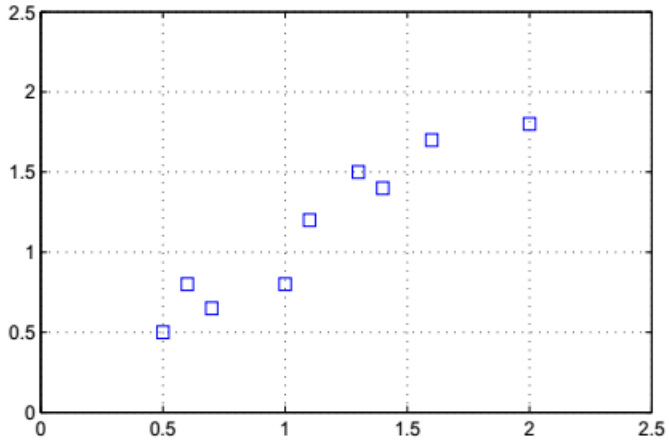
تاریخچه: واژه رگرسیون در فرهنگ لغت به معنی «بازگشت» است و اغلب جهت رساندن مفهوم «بازگشت به یک مقدار متوسط یا میانگین» به‌کار می‌رود. به این معنی که برخی پدیده‌ها به مرور زمان از نظر کمّی به طرف یک مقدار متوسط میل می‌کند.

بیش از ۱۰۰ سال پیش در سال ۱۸۷۷ فرانسیس گالتون[۱] در مقاله‌ای که در همین زمینه منتشر کرد اظهار داشت که متوسط قد پسران دارای پدران قدبلند، کمتر از قد پدرانشان می‌باشد. به نحو مشابه متوسط قد پسران دارای پدران کوتاه قد نیز بیشتر از قد پدرانشان گزارش شده است. به این ترتیب گالتون پدیده بازگشت به طرف میانگین را در داده‌هایش مورد تاکید قرار داد. برای گالتون رگرسیون مفهومی زیست شناختی داشت اما کارهای او توسط کارل پیرسون[۲] برای مفاهیم آماری توسعه داده شد. گرچه گالتون برای تاکید بر پدیده «بازگشت به سمت مقدار متوسط» از تحلیل رگرسیون استفاده کرد، اما به هر حال امروزه واژه تحلیل رگرسیون جهت اشاره به مطالعات مربوط به روابط بین متغیرها به‌کار برده می‌شود.

[1] Francis Galton
[2] Karl Pearson

۱.۱.۳ نمودار پراکندگی

در حقیقت تحلیل رگرسیونی فن و تکنیکی آماری برای بررسی و مدل‌سازی و ارتباط بین متغیرهاست. رگرسیون تقریباً در هر زمینه‌ای از جمله مهندسی، فیزیک، اقتصاد، مدیریت، علوم زیستی، بیولوژی و علوم اجتماعی برای برآورد و پیش‌بینی مورد نیاز است. می‌توان گفت تحلیل رگرسیونی، پرکاربردترین روش در بین تکنیک‌های آماری است. خلاصه‌ای از تحلیل رگرسیونی به‌صورت زیر است:
در ابتدا تحلیل‌گر حدس می‌زند که یک رابطه به شکل یک خط بین دو متغیر وجود دارد و سپس به جمع‌آوری اطلاعات کمّی از دو متغیر می‌پردازد و این داده‌ها را به‌صورت نقاطی در یک نمودار دو بعدی رسم می‌کند (شکل (۱.۳)).



شکل ۱.۳: نمودار پراکندگی داده‌ها

این نمودار که به آن نمودار پراکندگی[1] گفته می‌شود نقش بسیار مهمی در تحلیل‌های رگرسیونی و نمایش ارتباط بین متغیرها ایفا می‌کند.
درصورتی که نمودار نشان دهنده این باشد که داده‌ها تقریباً (نه لزوماً دقیق) در امتداد یک خط مستقیم پراکنده شده‌اند، حدس تحلیل‌گر تایید شده (شکل (۲.۳)) و این ارتباط خطی به‌صورت زیر نمایش داده می‌شود:

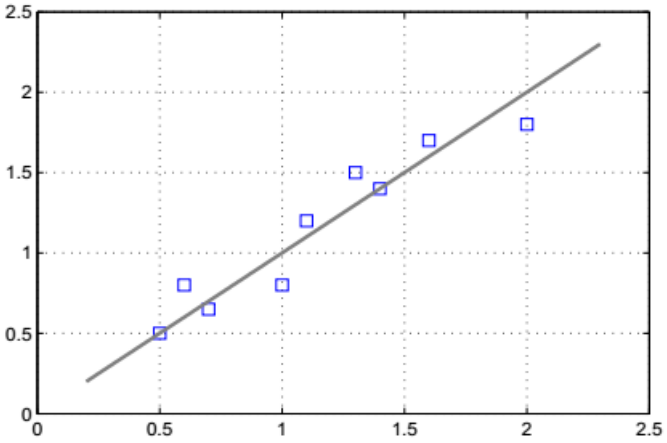
$$y = ax + b,$$

که در آن b عرض از مبدأ و a شیب این خط است.

۲.۱.۳ متغیرها و خطا

بین برخی از نقاط و تصویر آن‌ها بر روی خط رگرسیونی کمی تفاوت به چشم می‌خورد که از آن به عنوان خطای برآورد یاد می‌کنیم. بنابراین معادله اولیه را به صورت زیر اصلاح می‌کنیم:

[1] Scatter Plot



شکل ۲.۳: برآورد خط رگرسیونی

$$y = ax + b + \epsilon,$$

معادله فوق یک مدل رگرسیونی خطی نامیده می‌شود.

۳.۱.۳ روش‌های رگرسیونی

تا این مرحله مدل رگرسیونی معرفی شد و کافی است پارامترهای مجهول a و b برآورد شوند. برآورد پارامترها در مدل‌سازی با استفاده از روش‌های مختلف انجام می‌شود از جمله روش کم‌ترین مربعات خطا. روش کم‌ترین مربعات خطا که یکی از روش‌های مورد استفاده در تحلیل رگرسیونی است، اولین بار به‌صورت واضح و مختصر توسط لژاندر[۱] ریاضی‌دان فرانسوی در سال ۱۸۰۵ و گاوس[۲] ریاضی‌دان آلمانی درسال ۱۸۰۹ معرفی و در مطالعات نجومی به‌کار برده شد.

روش‌های ماشین بردار پشتیبان که در فصل اول آن را توضیح دادیم، در رگرسیون بسیار موفقیت آمیز عمل کرده است. این روش را رگرسیون بردار پشتیبان (SVR) می‌نامند.

۲.۳ استفاده از SVM در رگرسیون

۱.۲.۳ مساله رگرسیون خطی بردار پشتیبان

برای نمونه‌های آموزشی داده شده $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n) \in X \times \mathbb{R}$ که $x_i \in X$ نقطه‌ای در فضای ورودی و y_i متناظر مقدار تابع $f(x_i)$ است. هدف پیدا کردن تابع تخمینی $f(x)$ است

[1] Legendre
[2] Gauss

که برای همه نمونه‌های آموزشی x_i، حداکثر فاصله آن از مقدار هدف y_i، ϵ باشد. در رگرسیون ϵـ بردار پشتیبان ، هر انحراف که کمتر از ϵ است، مجاز می‌باشد.

تابع خطی $f(x) =< w, x > +b$ را در نظر بگیرید که در آن $b \in \mathbb{R}$ و $w \in \mathbb{R}^n$. هدف SVR می‌نیمم سازی نرم اقلیدسی w است به قسمی که انحراف $f(x)$ از مقادیر خروجی y_i کمتر از ϵ باشد. این مساله بهینه سازی محدب به‌صورت زیر بیان می‌شود:

$$\text{minimize } \frac{۱}{۲}\|w\|^۲, \tag{۱.۳}$$

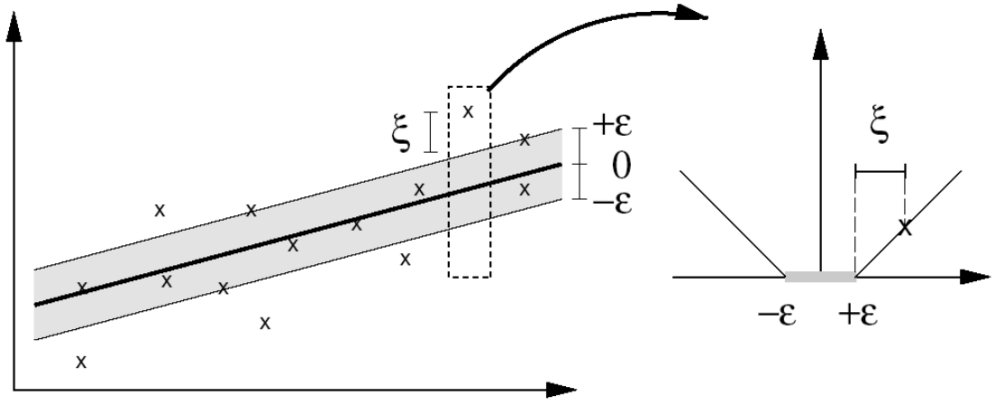
$$\text{subject to } \begin{cases} y_i - < w, x_i > -b \le \epsilon, & i = ۱, ..., n, \\ < w, x_i > +b - y_i \le \epsilon, & i = ۱, ..., n. \end{cases} \tag{۲.۳}$$

در مساله (۱.۳) و (۲.۳) فرض ضمنی بر این است که تابع $f(x)$ که همه اطلاعات ورودی $\{(x_i, y_i)\}_{i=۱}^n$ را تخمین می‌زند، موجود است. برای ساختن یک ماشین بردار پشتیبان که هدف d را تخمین می‌زند، از تابع زیان زیر استفاده می‌کنیم:

$$L_\varepsilon(d, y) = \begin{cases} |d - y| - \varepsilon, & \text{اگر } |d - y| \ge \varepsilon \\ ۰, & \text{در غیر این‌صورت.} \end{cases} \tag{۳.۳}$$

که ϵ پارامتری از پیش تعیین شده و مثبت است. تابع زیان (۳.۳) تابع زیان ϵ ـ حساسیت نامیده می‌شود.

این تعریف نتیجه می‌دهد که تابع زیان تنها زمانی دارای مقدار است که انحراف خروجی y از هدف مطلوب d بزرگ‌تر از پارامتر انحراف ϵ باشد، که به‌طور شهودی در شکل (۳.۳) نشان داده شده است.



شکل ۳.۳: تابع زیان ϵ ـ حساسیت

برای شدنی بودن محدودیت‌های (۲.۳)، متغیرهای کمکی ζ_i و ζ_i' معرفی می‌شوند تا مساله بهینه‌سازی را به‌شکل زیر تشکیل دهیم:

$$\text{minimize } \frac{۱}{۲}\|w\|^۲ + C\sum_{i=۱}^{n}(\zeta_i + \zeta_i'), \tag{۴.۳}$$

$$\text{subject to} \begin{cases} y_i - < w, x_i > -b \leq \epsilon + \zeta_i, & i = ۱, ..., n, \\ < w, x_i > +b - y_i \leq \epsilon + \zeta_i', & i = ۱, ..., n, \\ \zeta_i, \zeta_i' \geq ۰, & i = ۱, ..., n. \end{cases} \quad (۵.۳)$$

که در آن ϵ پارامتر انحراف و C پارامتر جریمه است. برای ساختن دوگان مساله بهینه‌سازی (۴.۳) و (۵.۳) متغیرهای کمکی (α_i, α_i') را معرفی می‌کنیم. α_i متناظر محدودیت‌های شامل ζ_i و α_i' متناظر محدودیت‌های کمکی ζ_i' است. این متغیرها در حقیقت ضرایب لاگرانژ برای مساله اولیه هستند. مشابه روش‌های به‌کار رفته در طراحی رده‌بندی کننده‌های SV، مساله بهینه سازی مقید (۴.۳) و (۵.۳) را با تشکیل تابع لاگرانژین به‌صورت زیر حل می‌کنیم.

$$\begin{aligned} L_P(w, b, \zeta_i, \zeta_i', \alpha_i, \alpha_i', \beta_i, \beta_i') = \quad & \frac{۱}{۲} w^T w + C \sum_{i=۱}^{n} (\zeta_i + \zeta_i') - \sum_{i=۱}^{n} (\beta_i \zeta_i + \beta_i' \zeta_i') - \\ & \sum_{i=۱}^{n} \alpha_i [w^T x_i + b - y_i + \epsilon + \zeta_i] - \\ & \sum_{i=۱}^{n} \alpha_i' [y_i - w^T x_i - b + \epsilon + \zeta_i']. \end{aligned}$$

تابع لاگرانژین $L_P(w, b, \zeta_i, \zeta_i', \alpha_i, \alpha_i', \beta_i, \beta_i')$ باید نسبت به متغیرهای اولیه w، b، ζ_i و ζ_i' می‌نیمم و نسبت به ضرایب لاگرانژ نامنفی α_i، α_i'، β_i و β_i' ماکزیمم شود. بنابراین تابع لاگرانژین در نقطه جواب بهینه مساله (۴.۳) و (۵.۳) یعنی $(w^*, b^*, \zeta_i^*, \zeta_i'^*)$ دارای نقطه زینی است.

در جواب بهینه مشتقات جزئی نسبت به متغیرهای اولیه صفر می‌شود، یعنی :

$$\begin{aligned} \frac{\partial L_P}{\partial w} = ۰ \Rightarrow w - \sum_{i=۱}^{n} (\alpha_i - \alpha_i') x_i = ۰, \quad (۶.۳) \\ \frac{\partial L_P}{\partial b} = ۰ \Rightarrow \sum_{i=۱}^{n} (\alpha_i - \alpha_i') = ۰, \\ \frac{\partial L_P}{\partial \zeta_i} = ۰ \Rightarrow C - \alpha_i - \beta_i = ۰, \qquad i = ۱, ..., n, \\ \frac{\partial L_P}{\partial \zeta_i} = ۰ \Rightarrow C - \alpha_i - \beta_i = ۰. \qquad i = ۱, ..., n. \end{aligned}$$

با استفاده از شرایط KKT دوگان مساله بهینه‌سازی برای رگرسیون خطی با استفاده از SVM به‌صورت زیر است:

$$\text{maximize} \quad -\frac{۱}{۲} \sum_{i=۱}^{n} \sum_{j=۱}^{n} (\alpha_i - \alpha_i')(\alpha_j - \alpha_j') x_i^T x_j + \sum_{i=۱}^{n} y_i (\alpha_i - \alpha_i') - \epsilon \sum_{i=۱}^{n} (\alpha_i + \alpha_i'), \quad (۷.۳)$$

$$\text{subject to} \begin{cases} \sum_{i=۱}^{n} (\alpha_i - \alpha_i') = ۰, \\ \alpha_i \in [۰, C], \ i = ۱, ..., n, \\ \alpha_i' \in [۰, C], \ i = ۱, ..., n. \end{cases} \quad (۸.۳)$$

توجه شود که مساله (۷.۳) و (۸.۳) تنها به‌صورت عباراتی از ضرایب لاگرانژ α و α' بیان شده است.

مرحله‌ی آموزش n جفت ضریب لاگرانژ (α_i, α'_i) را نتیجه می‌دهد. بعد از آموزش تعداد پارامترهای ناصفر α_i یا α'_i مساوی تعداد بردار پشتیبان است. از آن‌جا که حداقل یک مولفه از هر جفت (α_i, α'_i) $(i = ۱, ..., n)$ صفر است لذا حاصل‌ضرب α_i و α'_i همیشه صفر است، یعنی $\alpha_i \alpha'_i = ۰$.

در بهینگی شرایط KKT زیر برقرار هستند:

$$\alpha_i(w^T x_i + b - y_i + \epsilon + \zeta_i = ۰),$$
$$\alpha'_i(-w^T x_i - b + y_i + \epsilon + \zeta_i = ۰),$$
$$\beta_i \zeta_i = (C - \alpha_i)\zeta_i = ۰, \qquad (۹.۳)$$
$$\beta'_i \zeta'_i = (C - \alpha'_i)\zeta'_i = ۰. \qquad (۱۰.۳)$$

شرایط فوق نشان می‌دهد که برای $۰ < \alpha_i < C$، شرط $\zeta_i = ۰$ برقرار است. به‌طور مشابه اگر $۰ < \alpha'_i < C$ آن‌گاه $\zeta'_i = ۰$ و برای $۰ < \alpha_i, \alpha'_i < C$ داریم:

$$\alpha_i(w^T x_i + b - y_i + \epsilon) = ۰,$$
$$\alpha'_i(-w^T x_i - b + y_i + \epsilon) = ۰.$$

شرایط (۹.۳) و (۱۰.۳) در مورد داده‌هایی هستند که دارای انحراف بیشتر از ϵ نسبت به خط رگرسیون هستند. یعنی اگر $\zeta_i > ۰$ آن‌گاه $\alpha_i = C$ و اگر $\zeta'_i > ۰$، آن‌گاه $\alpha'_i = C$. به عبارتی دیگر برای داده‌هایی که بالای محدوده‌ی با خطای مجاز هستند، $\alpha_i = C$، و برای داده‌هایی که پایین محدوده با خطای مجاز هستند، $\alpha'_i = C$، این نقاط را **بردارهای پشتیبان مقید** می‌نامند. هم‌چنین برای همه نقاط تمرینی داخل ناحیه مجاز، یا جایی که $|y - f(x)| < \epsilon$ هر دوی α_i و α'_i مساوی صفر هستند. لذا هیچ‌یک از آن‌ها بردار پشتیبان نیستند و تاثیری در صفحه تصمیم ندارند.

اما پس از حل مساله بهینه سازی (۷.۳) و (۸.۳)با داشتن ضرایب بهینه α_i و α'_i، با استفاده از شرط بهینگی (۶.۳) بردار وزنی بهینه برای ابرصفحه رگرسیون به‌صورت زیر به‌دست می‌آید:

$$w_۰ = \sum_{i=۱}^{n} (\alpha_i - \alpha'_i) x_i = ۰, \qquad (۱۱.۳)$$

و ابرصفحه بهینه به‌صورت زیر به‌دست می‌آید:

$$f(x) = w_۰ x + b = \sum_{i=۱}^{n} (\alpha_i - \alpha'_i) x_i^T x + b.$$

۲.۲.۳ مساله رگرسیون غیرخطی بردار پشتیبان

حل مسائل رگرسیون غیرخطی بیشتر مورد توجه قرار می‌گیرد. یک تعمیم به حالت غیرخطی مشابه روش توسعه رده‌بندی کننده غیرخطی که در (۳.۲.۱) بیان شده است، استفاده از رده‌بندی کننده غیرخطی است، یعنی به‌وسیله یک نگاشت به فضای ویژگی (که معمولاً دارای بعدخیلی بالایی است) داده‌ها را به فضایی با قابلیت رگرسیون خطی می‌بریم و سپس تابع رگرسیون غیرخطی را متناظر معادله رگرسیون خطی در فضای ویژگی به‌دست می‌آوریم.

بنابراین ابر صفحه بهینه رگرسیون بردار پشتیبان در حالت غیرخطی با حل مساله زیر به‌دست می‌آید:

$$\text{minimize } \frac{۱}{۲}\|w\|^۲ + C\sum_{i=۱}^{n}(\zeta_i + \zeta_i'), \tag{۱۲.۳}$$

$$\text{subject to } \begin{cases} y_i - < w, \Phi(x_i) > -b \le \epsilon + \zeta_i, & i = ۱, ..., n, \\ < w, \Phi(x_i) > +b - y_i \le \epsilon + \zeta_i', & i = ۱, ..., n, \\ \zeta_i, \zeta_i' \ge ۰, & i = ۱, ..., n. \end{cases} \tag{۱۳.۳}$$

که در آن ϵ پارامتر انحراف و C پارامتر جریمه است. برای ساختن دوگان مساله بهینه‌سازی (۱۲.۳) و (۱۳.۳) متغیرهای کمکی (α_i, α_i') را معرفی می‌کنیم. α_i متناظر محدودیت‌های شامل ζ_i و α_i' متناظر محدودیت‌های کمکی ζ_i' است. این متغیرها در حقیقت ضرایب لاگرانژ برای مساله اولیه هستند. مشابه روش‌های به‌کار رفته در رگرسیون خطی، مساله بهینه سازی مقید (۱۲.۳) و (۱۳.۳) را با تشکیل لاگرانژین متغیرهای اولیه حل می‌کنیم که تابع لاگرانژین به‌صورت زیر است:

$$\begin{aligned} L_P(w, b, \zeta_i, \zeta_i', \alpha_i, \alpha_i', \beta_i, \beta_i') = \quad & \frac{۱}{۲}w^T w + C\sum_{i=۱}^{n}(\zeta_i + \zeta_i') - \sum_{i=۱}^{n}(\beta_i\zeta_i + \beta_i'\zeta_i') - \\ & \sum_{i=۱}^{n}\alpha_i[w^T\Phi(x_i) + b - y_i + \epsilon + \zeta_i] - \\ & \sum_{i=۱}^{n}\alpha_i'[y_i - w^T\Phi(x_i) - b + \epsilon + \zeta_i']. \end{aligned}$$

تابع لاگرانژین $L_P(w, b, \zeta_i, \zeta_i', \alpha_i, \alpha_i', \beta_i, \beta_i')$ باید نسبت به متغیرهای اولیه w، b، ζ_i و ζ_i' می‌نیمم شود و نسبت به ضرایب لاگرانژ نامنفی α_i، α_i'، β_i و β_i' ماکزیمم گردد. بنابراین تابع لاگرانژین در نقطه جواب بهینه مساله (۱۲.۳) و (۱۳.۳) یعنی $(w^*, b^*, \zeta_i^*, \zeta_i'^*)$ دارای نقطه زینی است.

در جواب بهینه مشتقات جزئی نسبت به متغیرهای اولیه صفر می‌شود، یعنی :

$$\frac{\partial L_P}{\partial w} = ۰ \Rightarrow w - \sum_{i=۱}^{n} (\alpha_i - \alpha_i')\Phi(x_i) = ۰, \qquad (۱۴.۳)$$

$$\frac{\partial L_P}{\partial b} = ۰ \Rightarrow \sum_{i=۱}^{n} (\alpha_i - \alpha_i') = ۰,$$

$$\frac{\partial L_P}{\partial \zeta_i} = ۰ \Rightarrow C - \alpha_i - \beta_i = ۰, \qquad i = ۱, ..., n,$$

$$\frac{\partial L_P}{\partial \zeta_i} = ۰ \Rightarrow C - \alpha_i - \beta_i = ۰. \qquad i = ۱, ..., n.$$

بنابراین مشابه حالت خطی دوگان مساله بهینه‌سازی برای رگرسیون غیرخطی با استفاده از SVR به‌صورت زیر است:

$$\text{maximize} \quad -\frac{۱}{۲}\sum_{i=۱}^{n}\sum_{j=۱}^{n}(\alpha_i - \alpha_i')(\alpha_j - \alpha_j')\Phi(x_i)^T\Phi(x_j) + \sum_{i=۱}^{n} y_i(\alpha_i - \alpha_i') - \epsilon\sum_{i=۱}^{n}(\alpha_i + \alpha_i'), \qquad (۱۵.۳)$$

$$\text{subject to} \begin{cases} \sum_{i=۱}^{n}(\alpha_i - \alpha_i') = ۰, \\ \alpha_i \in [۰, C], \ i = ۱, ..., n, \\ \alpha_i' \in [۰, C], \ i = ۱, ..., n. \end{cases} \qquad (۱۶.۳)$$

از شرط بهینگی (۱۴.۳)، w به‌صورت زیر به‌دست می‌آید:

$$w = \sum_{i=۱}^{n}(\alpha_i - \alpha_i')\Phi(x_i).$$

بنابراین تابع تخمین $f(x) =< w, x > +b$ به‌صورت زیر بیان می‌شود:

$$f(x) = \sum_{i=۱}^{n}(\alpha_i - \alpha_i')\Phi(x_i)^T\Phi(x) + b.$$

می‌توان مساله (۱۵.۳) و (۱۶.۳) را به‌شکل خلاصه زیر نوشت،

$$\text{minimize } \frac{۱}{۲}\theta^T Q\theta + d^T\theta,$$

$$\text{subject to} \begin{cases} A\theta - b \leq ۰, \\ e^T\theta = ۰. \end{cases}$$

که در آن،

$$Q_{۴n\times ۴n} = \begin{cases} \Phi(X_i)^T\Phi(X_j) & ۱ \leq i, j \leq n \\ ۰ & \text{در غیر این‌صورت} \end{cases},$$

$$d = (-y_۱, -y_۲ ..., -y_n, \underbrace{\epsilon, \epsilon, \cdots, \epsilon}_{n}, \underbrace{۰, ۰, \cdots, ۰}_{۲n})^T,$$

$$e = (\underbrace{۱, ۱, \cdots, ۱}_{n}, \underbrace{۰, ۰, \cdots, ۰}_{۳n})^T \in \mathbb{R}^{۴n},$$

$$A = \begin{pmatrix} I_{۲n\times ۲n} & O_{۲n\times ۲n} \\ O_{۲n\times ۲n} & -I_{۲n\times ۲n} \end{pmatrix},$$

$$b = (\underbrace{C, C, \cdots, C}_{n}, \underbrace{۲C, ۲C, \cdots, ۲C}_{n}, \underbrace{C, C, \cdots, C}_{n}, \underbrace{۰, ۰, \cdots, ۰}_{n})^T \in \mathbb{R}^{۴n},$$

$$\theta = (\alpha_۱ - \alpha'_۱, \alpha_۲ - \alpha'_۲, ..., \alpha_n - \alpha'_n, \alpha_۱ + \alpha'_۱, \alpha_۲ + \alpha'_۲ ..., \alpha_n + \alpha'_n, \alpha_۱ - \alpha'_۱, \alpha_۲ - \alpha'_۲, ..., \alpha_n - \alpha'_n,$$
$$\alpha_۱ + \alpha'_۱, \alpha_۲ + \alpha'_۲, ..., \alpha_n + \alpha'_n)^T.$$

در ادامه این فصل با استفاده از شبکه عصبی به حل مساله رگرسیون بردار پشتیبان می‌پردازیم. ثابت می‌شود که نقطه تعادل شبکه عصبی مورد نظر با نقطه بهینه مساله درجه دوم رگرسیون بردار پشتیبان برابر است. وجود داشتن و منحصر به‌فرد بودن نقطه تعادل ثابت شده و با ساخت تابع لیاپانوف مناسب پایداری مجانبی اثبات می‌شود. در ادامه نیز با ارائه مثال‌های عددی کارایی این شبکه بررسی می‌گردد.

۳.۳ معرفی شبکه عصبی

مساله برنامه‌ریزی غیرخطی زیر را در نظر بگیرید. در این رساله فرض بر این است که این مساله دارای جواب یکتا است.

$$\text{minimize } \frac{۱}{۲}\theta^T Q\theta + d^T\theta, \tag{۱۷.۳}$$

$$\text{subject to } \begin{cases} A\theta - b \leq ۰, \\ e^T\theta = ۰. \end{cases} \tag{۱۸.۳}$$

در این بخش می‌خواهیم یک شبکه عصبی با کارایی بالا را معرفی کرده و در مورد ویژگی‌های آن بحث کنیم.

در ادامه توضیح می‌دهیم که سیستم دینامیکی که در این بخش بررسی می‌شود، چگونه شکل می‌گیرد. همانند آنچه در [۳۸] نشان داده شد، مراحل ساخت شبکه عصبی شامل ساخت شبکه عصبی و ایجاد تابع لیاپانوف است. بدین منظور ابتدا شرایط KKT را به صورت زیر برای مساله (۱۷.۳) و (۱۸.۳) در نظر می‌گیریم.

$$\begin{cases} \nu^* \geq ۰, \quad A\theta^* - b \leq ۰, \quad \nu^{*^T}(A\theta^* - b) = ۰, \\ Q\theta^* + d + A^T\nu^* + e\vartheta^* = ۰, \\ e^T\theta^* = ۰. \end{cases} \tag{۱۹.۳}$$

قضیه ۱.۳.۳. [۱۱] $\theta^* \in \mathbb{R}^N$ یک نقطه بهینه (۱۷.۳) و (۱۸.۳) است اگر و فقط اگر نقاطی مانند $\nu^* \in \mathbb{R}^{۲N}$ و $\vartheta^* \in \mathbb{R}$ وجود داشته باشد، به‌طوری که $(\theta^{*T}, \nu^{*T}, \vartheta^{*T})^T$ در شرایط KKT سیستم (۱۹.۳) صدق کند.

که در آن θ^* یک نقطه KKT معادله‌های (۱۷.۳) و (۱۸.۳) و زوج $(\nu^{*T}, \vartheta^{*T})^T$ بردار ضریب لاگرانژ متناظر با θ^* است.

لم ۱.۳.۳. $(\nu + A\theta - b)^+ = \nu$ اگر و فقط اگر $\nu^T(A\theta - b) = ۰$، $A\theta - b \leq ۰$ و $\nu \geq ۰$ که در آن

$$(\nu + A\theta - b)^+ = ([(\nu + A\theta - b)_۱]^+, [(\nu + A\theta - b)_۲]^+, \cdots, [(\nu + A\theta - b)_m]^+)^T,$$
$$[(\nu + A\theta - b)_k]^+ = \max\{(\nu + A\theta - b)_k, ۰\}, \quad k = ۱, ۲, \cdots, m.$$

برهان. اگر $(\nu + A\theta - b)^+ = \nu$، آن‌گاه $\nu = ۰$ یا $\nu > ۰$. در این‌صورت یکی از دو حالت زیر را داریم:

۱. اگر $\nu = ۰$ آن‌گاه $(\nu + A\theta - b)^+ = (A\theta - b)^+ = ۰$ چون $A\theta - b \leq ۰$ ، لذا داریم:
$$\nu^T(A\theta - b) = ۰.$$

۲. اگر $\nu > ۰$ چون $(\nu + A\theta - b)^+ = \nu + A\theta - b = \nu$ بنابراین $A\theta - b = ۰$ لذا داریم:
$$\nu^T(A\theta - b) = ۰.$$

عکس این رابطه نیز به‌راحتی قابل اثبات است.

□

اکنون فرض کنید $\theta(.)$، $\nu(.)$ و $\vartheta(.)$ متغیرهایی وابسته به زمان باشند. هدف ما ساخت شبکه عصبی پیوسته‌ای است که به نقطه KKT مساله (۱۷.۳) و (۱۸.۳) همگرا باشد. شبکه‌ای که در این رساله پیشنهاد می‌شود به شکل زیر است.

$$\frac{dy}{dt} = \tau\Psi(y), \tag{۲۰.۳}$$

$$y(t_۰) = y_۰ = (\theta_۰^T, \nu_۰^T, \vartheta_۰^T)^T, \quad \tau > ۰, \tag{۲۱.۳}$$

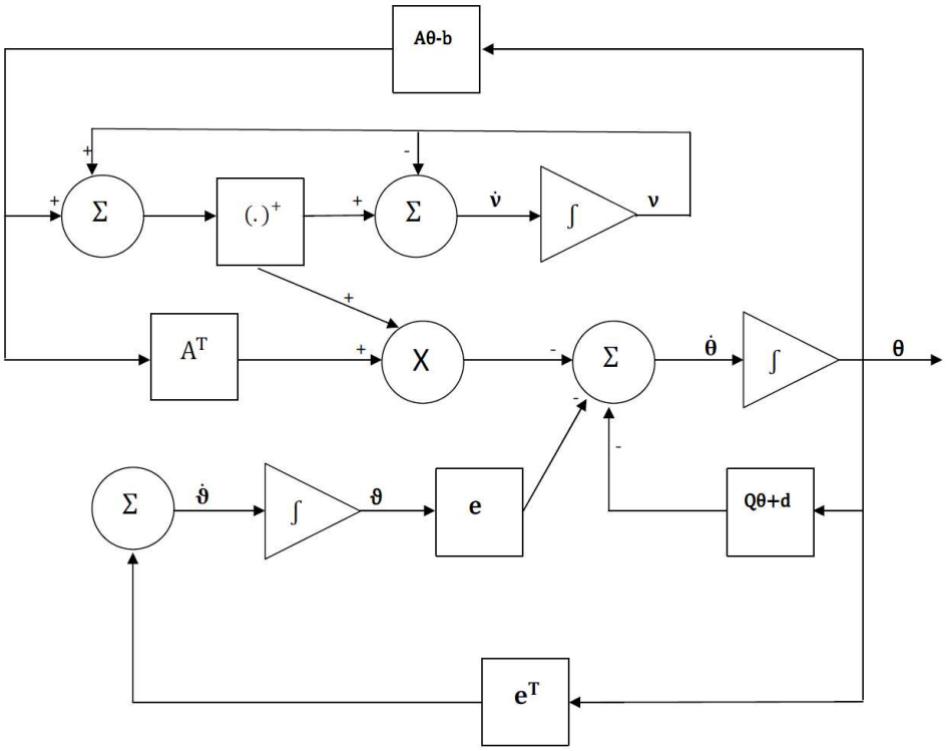
که در آن $y = (\theta^T, \nu^T, \vartheta^T)^T \in \mathbb{R}^{۳N+۱}$، τ یک پارامتر مقیاس و

$$\Psi(y) = \begin{pmatrix} -(Q\theta + d + A^T(\nu + A\theta - b)^+ + e\vartheta) \\ (\nu + A\theta - b)^+ - \nu \\ e^T\theta \end{pmatrix}, \tag{۲۲.۳}$$

است.

لازم به ذکر است که τ نرخ همگرایی شبکه عصبی (۲۰.۳) و (۲۱.۳) است.

تصویر (۴.۳) چگونگی اجرای این شبکه عصبی را توضیح می‌دهد.



شکل ۴.۳: یک نمودار ساده بلوکی شده برای شبکه عصبی (۲۰.۳) و (۲۱.۳)

۱.۳.۳ مقایسه با برخی شبکه‌های عصبی موجود

مساله (۱۷.۳) و (۱۸.۳) یک حالت خاص مساله درجه دوم محدب[1] [۶۰] به شکل زیر است.

$$\text{minimize } \frac{۱}{۲}x^TQx + D^Tx, \tag{۲۳.۳}$$

$$\text{subject to } \begin{cases} Ax - b \leq ۰, \\ Ex - f = ۰. \end{cases} \tag{۲۴.۳}$$

که در آن $Q \in \mathbb{R}^{n\times n}$ یک ماتریس متقارن نیمه معین مثبت، $b \in \mathbb{R}^m$، $E \in \mathbb{R}^{l\times n}$، $f \in \mathbb{R}^l$، $x \in \mathbb{R}^n$ و $rank(A) = m (۰ < m < n)$ است.

یک مدل کلی برای حل مساله (۲۳.۳) و (۲۴.۳) به شکل زیر است:

$$\frac{dy}{dt} = \tau\Psi(y), \tag{۲۵.۳}$$

$$y(t_۰) = y_۰ = (x_۰^T, u_۰^T, v_۰^T)^T \in \mathbb{R}^{n+m+l}, \quad \tau > ۰, \tag{۲۶.۳}$$

که در آن $\psi(y)$ به شکل زیر است:

$$\Psi(y) = \begin{pmatrix} -(Qx + D + A^T(u + Ax - b)^+ + E^Tv) \\ (u + Ax - b)^+ - u \\ Ex - f \end{pmatrix}.$$

[1] degenerate convex quadratic programming (DCQP)

واضح است که مساله (۲۰.۳) و (۲۱.۳) یک حالت خاص مساله (۲۵.۳) و (۲۶.۳) است.
کارایی شبکه عصبی ارائه شده برای حل مسائل DCQP را با مقایسه‌ی این شبکه با برخی شبکه‌های موجود نشان می‌دهیم.
مساله‌ی زیر را در نظر بگیرید:

$$\text{minimum} \quad \frac{۱}{۲}x^TQx + D^Tx, \tag{۲۷.۳}$$

$$\text{subject to} \quad Ax - b \leq ۰, \quad x \in \Omega, \tag{۲۸.۳}$$

که در آن $\Omega = \{x \in \mathbb{R}^n,\ l_i \leq x_i \leq h_i,\ i = ۱,۲,...,n\}$ مقادیر l_i و h_i می‌توانند بی کران بوده و $Ax - b \in \mathbb{R}^m$ است. در [۱۹] شبکه عصبی تصویر فریز[1] برای حل مساله (۲۷.۳) و (۲۸.۳) به شکل زیر بیان شده است:

$$\frac{dx}{dt} = -(x - P_\Omega(x - (Qx + D) - A^Tu)), \tag{۲۹.۳}$$

$$\frac{du}{dt} = -(u - (u + Ax - b)^+), \tag{۳۰.۳}$$

که در آن $\Omega \subset \mathbb{R}^n$ یک مجموعه‌ی محدب بسته و $P_\Omega : \mathbb{R}^n \to \Omega$ یک عملگر تصویر است.
همان‌طور که در [۵۸] نشان داده شده است، همگرایی مجانبی مدل دینامیکی (۲۹.۳) و (۳۰.۳) با یک نگاشت یکنوای متقارن تضمین شده نیست. بنابراین سیستم توصیف شده توسط فریز نمی‌تواند مساله (۲۷.۳) و (۲۸.۳) را در حالتی که Q یک ماتریس نیمه معین مثبت است، مثلاً مساله برنامه ریزی خطی **LP** را حل کند. به عنوان مثال می‌توانید مثال LP را در [۳۷] ببینید.
در حالتی که ماتریس Q فقط معین مثبت باشد، از مدل شبکه لاگرانژی زیر که در [۳۵] برای حل مساله (۲۳.۳) و (۲۴.۳) ارائه شده، استفاده می‌کنیم.

$$\frac{dx}{dt} = -(Qx + D + \frac{۱}{۲}A^Tu^۲ + E^Tv), \tag{۳۱.۳}$$

$$\frac{du}{dt} = diag(u_۱, u_۲, ..., u_m)(Ax - b), \tag{۳۲.۳}$$

$$\frac{dv}{dt} = Ex - f, \tag{۳۳.۳}$$

با یک نقطه‌ی آغازین $(x_۰^T, u_۰^T, v_۰^T)^T$ به‌طوری‌که $u_k(t_۰) > ۰ \quad (k = ۱, ..., m)$.
هرچند همگرایی سراسری شبکه عصبی (۳۱.۳)ـ(۳۳.۳) اثبات شده است ولی این مدل قادر به حل مسائل LP نیست. برای درک بهتر این موضوع می‌توانید مثال LP در [۳۴] را ببینید.
یک مدل شبکه عصبی گرادیانی به نام مدل کندی و چا[2] برای حل مساله (۲۳.۳) و (۲۴.۳) در [۲۲] به‌شکل زیر ارائه شده است:

$$\frac{dx}{dt} = -(Qx + D + ς[A^T(Ax - b)^+ + E^T(Ex - f)]),$$

[1] Friesz
[2] Kennedy and Chau's

که در آن ς یک پارامتر جریمه است.
این مدل هنگامی که پارامتر جریمه آن متناهی است قادر به پیدا کردن جواب دقیق مساله نیست و زمانی که پارامتر جریمه آن بسیار زیاد شود، اجرای آن دشوار است [۱۶]. این مدل به‌ازای هر مقدار متناهی برای پارامتر جریمه به تقریبی از جواب مساله (۲۳.۳) و (۲۴.۳) همگرا است. هم‌چنین می‌توان نشان داد که این مساله همگرای سراسری به یک جواب دقیق مسائل $DCQP$ نیست. به‌عنوان مثال، مساله LP را در [۲۶] ببینید، در حالی‌که این مساله به‌صورت دقیق در [۳۷] حل شده است.

با استفاده از یک تابع مکمل غیرخطی و روش گرادیان در [۳۶] یک شبکه عصبی برای حل مسائل اکیداً محدب درجه دوم به‌شکل زیر ارائه شده است.

$$\frac{dy(t)}{dt} = -k\nabla E(y(t)), \quad k > 0,$$
$$y(0) = y_0,$$

که در آن

$$E(y) = \frac{1}{2}\|\eta(y)\|^2,$$
$$\eta(y) = \begin{bmatrix} Qx + D + E^T v + A^T u \\ f - Ex \\ \phi^{\epsilon}_{FB}(u, b - Ax) \end{bmatrix} = 0,$$
$$\phi^{\epsilon}_{FB}(a, b) = (a + b) - \sqrt{a^2 + b^2 - \epsilon}, \quad \epsilon \to 0_+.$$

این مدل دارای پایداری لیاپانوف و پایدار مجانبی است، ولی دارای نرون های بیشتری با پیچیدگی محاسباتی بالا است و به شرایط همگرایی قوی‌تر نیاز دارد. علاوه براین، این مدل نمی‌تواند مسائل برنامه ریزی خطی (LP) را حل کند، زیرا درمسائل برنامه‌ریزی درجه دوم (QP) فرض براین است که تابع هدف اکیداً محدب است درحالی‌که در مسائل (LP) تابع هدف فقط محدب است.

زیا[1] در [۵۵] شبکه عصبی‌ای برای حل مساله زیر ارائه داد:

$$\text{minimize} \quad \frac{1}{2}x^T Qx + D^T x, \tag{۳۴.۳}$$
$$\text{subject to} \quad x \in \Omega_0, \tag{۳۵.۳}$$
$$Dx = b, \tag{۳۶.۳}$$

به‌طوری که :

$$\frac{dx}{dt} = (I + A_0)[P_0(x - A_0 x + D^T y - a) - x] - D^T(Dx - b),$$
$$\frac{dy}{dt} = -DP_0(x - A_0 x + D^T y - a) + b,$$

[1] Xia

که در آن $A_{\circ_{m\times m}}$ ماتریس متقارن نیمه معین مثبت، D یک ماتریس $n \times m$ و $\Omega_\circ \subset \mathbb{R}^m$ یک مجموعه‌ی محدب بسته است. این مساله ضرایب آنالوگ زیادی دارد و برای مسائل بهینه سازی با ابعاد بالا به‌صرفه نیست. برای حل این مشکل تااو و همکاران[1] در [۴۷] مدل اصلاح شده‌ی زیر را بیان کردند:

$$\frac{dx}{dt} = [P_\circ(x - A_\circ x + D^T y - a) - x] - D^T(Dx - b),$$
$$\frac{dy}{dt} = -DP_\circ(x - A_\circ x + D^T y - a) + b.$$

درباره‌ی این شبکه نیز می‌توان گفت هنوز همگرایی آن در زمان متناهی ثابت نشده است. اگر مدل ارائه شده در [۵۸] را برای حل مساله (۳۴.۳) ـ (۳۶.۳) استفاده کنیم، ابعاد شبکه بزرگ‌تر می‌شود، علاوه‌براین همگرایی این شبکه در زمان متناهی نیز ثابت نشده است. به‌منظور رسیدن به همگرایی در زمان متناهی و هم‌چنین همگرایی نمایی، شیا و همکاران[2] در [۵۶] شبکه عصبی‌ای را برای حل مسائل درجه دوم اکیداً محدب نشان دادند. گرچه شبکه مورد نظر دارای همگرایی نمایی در زمان متناهی به جواب یکتای مساله درجه دوم اکیداً محدب با قیدهای خطی است ولی این شبکه برای حل مساله (۲۳.۳) و (۲۴.۳) درحالی‌که ماتریس $Q \in \mathbb{R}^{n\times n}$ فقط متقارن و نیمه معین مثبت است، کارایی ندارد. لذا این شبکه نمی‌تواند مسائل LP را حل کند.

اخیراً زیو و بی اِن[3] در [۶۰]، بر اساس نتایج به‌دست آمده در [۵۶] برای مسائل درجه دوم اکیداً محدب، یک شبکه عصبی تصویر برای مساله (۳۷.۳)ـ (۳۹.۳) ارائه کردند،

$$\text{minimize} \quad \tfrac{1}{2}x^T Q x + c^T x, \tag{۳۷.۳}$$
$$\text{subject to} \quad b^1 \le Bx \le b^2, \tag{۳۸.۳}$$
$$d^\circ \le h^\circ \le h^\circ, \tag{۳۹.۳}$$

که در آن $Q \in \mathbb{R}^{n\times n}$ یک ماتریس متقارن نیمه معین مثبت، $B \in \mathbb{R}^{m\times n}$، $b^1, b^2 \in \mathbb{R}^m$، $d^\circ, h^\circ \in \mathbb{R}^n$ و $c \in \mathbb{R}^n$ است. در [۶۰] یک شبکه عصبی تصویر به‌گونه‌ای برای حل مساله (۳۷.۳)ـ (۳۹.۳) توسعه یافته که همگرای کامل در زمان متناهی باشد. با این حال، مدل شبکه عصبی پیشنهادی دارای متغیرها و نرون‌های بیشتری است، که باعث می‌شود اجرای شبکه دشوارتر شود.

برای کاهش پیچیدگی شبکه ارائه شده در [۶۰]، زیو در [۵۹] شبکه عصبی جدیدی برای حل مساله (۳۷.۳)ـ (۳۹.۳) ارائه کرده و ادعا دارد که دامنه آن گسترده‌تر از مقاله قبلی است. ولی با دقت متوجه می‌شویم که ساختار عمگر تصویر برای حل مساله (۳۷.۳)ـ (۳۹.۳) پیچیده بوده و لذا شبکه عصبی پیشنهادی براساس آن نیز پیچیدگی محاسباتی بالایی دارد.

[1] Tao et al
[2] Xia et al
[3] Xue and Bian

فورتی و همکاران در [۱۸] همگرایی مسیر دسته‌ای از شبکه‌های عصبی برای حل مسائل برنامه‌ریزی خطی و مسائل برنامه‌ریزی درجه دوم محدب را بررسی کردند. نتیجه اصلی این بررسی این بود که تمامی شبکه‌های مطالعه شده در این بررسی در زمان متناهی به یک مجموعه‌ی تک عضوی از زیرمجموعه‌های نقاط بحرانی قیود، همگرا هستند. با این حال، مسائل درجه دوم بررسی شده در [۱۸] فقط با محدودیت‌های آفین هستند و فضای شدنی آن‌ها یک چند وجهی محدب بسته، بدون نقاط داخلی است.

۲.۳.۳ تحلیل پایداری و همگرایی شبکه عصبی

در این بخش پایداری و همگرایی شبکه عصبی (۲۰.۳) و (۲۱.۳) را بررسی می کنیم.

قضیه ۲.۳.۳. فرض کنید $y^* = (\theta^{*T}, \nu^{*T}, \vartheta^{*T})^T$ نقطه تعادل شبکه عصبی (۲۰.۳) و (۲۱.۳) باشد، آن‌گاه θ^* یک نقطه KKT مساله (۱۷.۳) و (۱۸.۳) است. از طرف دیگر اگر $\theta^* \in \mathbb{R}^n$ یک نقطه بهینه مساله (۱۷.۳) و (۱۸.۳) باشد، آن‌گاه $\nu^* \in R^{۲N}$ و $\vartheta^* \in R$ وجود دارد به‌طوری که $y^* = (\theta^{*T}, \nu^{*T}, \vartheta^{*T})^T$ نقطه تعادل شبکه (۲۰.۳) و (۲۱.۳) است.

برهان. فرض کنید y^* نقطه تعادل شبکه (۲۰.۳) و (۲۱.۳) باشد. در این‌صورت $\frac{d\theta^*}{dt} = ۰$، $\frac{d\nu^*}{dt} = ۰$ و $\frac{d\vartheta^*}{dt} = ۰$.
با توجه به این روابط به‌راحتی نتیجه گرفته می‌شود که

$$Q\theta^* + d + A^T(\nu^* + A\theta^* - b)^+ + e\vartheta^* = ۰, \tag{۴۰.۳}$$

$$(\nu^* + A\theta^* - b)^+ - \nu^* = ۰, \tag{۴۱.۳}$$

$$e^T\theta^* = ۰. \tag{۴۲.۳}$$

بنابر لم (۱.۳.۳) $(\nu^* + A\theta^* - b)^+ = \nu^*$، اگر و فقط اگر

$$\nu^* \geq ۰, \quad A\theta^* - b \leq ۰, \quad \nu^{*T}(A\theta^* - b) = ۰.$$

با جای‌گذاری (۴۰.۳) در (۴۱.۳) داریم:

$$Q\theta^* + d + A^T\nu^* + e\vartheta^* = ۰. \tag{۴۳.۳}$$

با توجه به روابط (۴۳.۳)ـ(۴۲.۳) مشخص است که $y^* = (\theta^{*T}, \nu^{*T}, \vartheta^{*T})^T$ در شرایط KKT ارائه شده در (۱۹.۳) صدق می‌کند.
عکس این قضیه به‌وضوح مشخص است.

□

لم ۲.۳.۳. ماتریس ژاکوبین ($\nabla\Psi(y)$) برای نگاشت Ψ تعریف شده در (۲۲.۳) یک ماتریس نیمه معین منفی است.

برهان. بدون از دست دادن کلیت، فرض می‌کنیم عدد $\circ < p < ۲N$ وجود دارد به‌طوری که

$$(\nu + A\theta - b)^+ = ((\nu + A\theta - b)_۱, (\nu + A\theta - b)_۲, \cdots, (\nu + A\theta - b)_p, \underbrace{\circ, \circ, \cdots, \circ}_{۲N-p})^T.$$

با یک محاسبه ساده، مشخص است که

$$\nabla\Psi(y) = \begin{pmatrix} -(Q + (A^p)^T A^p) & -(A^p)^T & -e \\ A^p & W_{۲N\times ۲N} & O_{۲N\times ۱} \\ e^T & O_{۱\times ۲N} & \circ \end{pmatrix},$$

که در آن O، ماتریس صفر،

$$A^p = \begin{pmatrix} U_{p\times N} \\ O_{(۲N-p)\times N} \end{pmatrix} = \begin{pmatrix} A_{۱.} \\ A_{۲.} \\ \cdots \\ \cdots \\ A_{p.} \\ O_{۱\times N} \\ O_{۱\times N} \\ \cdots \\ \cdots \\ O_{۱\times N} \end{pmatrix},$$

و

$$W_{۲N\times ۲N} = \begin{pmatrix} O_{p\times p} & O_{p\times(۲N-p)} \\ O_{(۲N-p)\times p} & -I_{(۲N-p)\times(۲N-p)} \end{pmatrix}.$$

در [۴۰] می‌توان دید که $(A^p)^T A^p$ یک ماتریس نیمه معین مثبت است. ماتریس Q نیز فرض می‌شود نیمه معین مثبت باشد. علاوه براین به‌وضوح مشخص است ماتریس $W_{۲N\times ۲N}$ یک ماتریس نیمه معین منفی است.

با استفاده از مشاهدات بیان شده، می‌توان فهمید که ماتریس ژاکوبین $(\nabla\Psi(y))$، یک ماتریس نیمه معین منفی است.

اگر $p = ۲N$ یعنی $(\nu + A\theta - b)^+ = ((\nu + A\theta - b)_۱, (\nu + A\theta - b)_۲, \cdots, (\nu + A\theta - b)_{۲N})^T$، آن‌گاه

$$\nabla\Psi(y) = \begin{pmatrix} -(Q + (A^T)A) & -A^T & -e \\ A & O_{۲N\times ۲N} & O_{۲N\times ۱} \\ e^T & O_{۱\times ۲N} & \circ \end{pmatrix}.$$

شبیه قسمت قبلی به سادگی ثابت می‌شود که $\nabla\Psi(y)$ یک ماتریس نیمه معین منفی است.

در نهایت اگر $p=۰$، یعنی $(\nu+A\theta-b)^+=(\underbrace{۰,۰,\cdots,۰}_{۲N})^T$، داریم:

$$\nabla\Psi(y)=\begin{pmatrix}-Q & O_{N\times ۲N} & -e\\ O_{۲N\times N} & -I_{۲N\times ۲N} & O_{۲N\times ۱}\\ e^T & O_{۱\times ۲N} & ۰\end{pmatrix}.$$

در این حالت نیز به سادگی می‌توان ثابت کرد که ماتریس $\nabla\Psi(y)$ نیمه معین منفی است. □

لم ۳.۳.۳. تابع $\|(\nu+A\theta-b)^+\|^۲$ از $(\nu+A\theta-b)$ در (۲۲.۳)، روی $\mathbb{R}^N\times\mathbb{R}^{۲N}$ محدب و به‌طور پیوسته مشتق پذیر است.

برهان. به‌وضوح، $\|(\nu+A\theta-b)^+\|^۲=\sum_{k=۱}^{۲N}([(\nu+A\theta-b)_k]^+)^۲$ و به‌ازای $k=۱,\cdots,۲N$ داریم:

$$([(\nu+A\theta-b)_k)^+)^۲=\begin{cases}((\nu+A\theta-b)_k)^۲ & (\nu+A\theta-b)_k\geq ۰ \text{ اگر}\\ ۰ & \text{درغیر این‌صورت}\end{cases},$$

با توجه به [۹]، از تحدب و مشتق‌پذیری $(A\theta-b)_k(k=۱,\cdots,۲N)$ نتیجه حاصل می‌شود. هم‌چنین داریم:

$$\nabla(\|(v+A\theta-b)^+\|^۲)=\begin{bmatrix}۲A^T(v+A\theta-b)^+\\ ۲(v+A\theta-b)^+\end{bmatrix},$$

که این مطلب، اثبات را کامل می‌کند.

□

اکنون نتایج اصلی خود را به شرح زیر بیان می‌کنیم.

قضیه ۳.۳.۳. مدل شبکه عصبی (۲۰.۳) و (۲۱.۳) به مفهوم لیاپانوف پایدار است.

برهان. تابع لیاپانوف $E:\mathbb{R}^{۳N+۱}\to\mathbb{R}$ را به‌صورت زیر در نظر بگیرید:

$$E(y)=E_۱(y)+E_۲(y), \qquad (۴۴.۳)$$

که در آن $E_۱(y)=\|\Psi(y)\|^۲$ و $E_۲(y)=\frac{۱}{۲}\|y-y^*\|^۲$ است. با توجه به لم (۳.۳.۳)، می‌دانیم $E_۱(y)$ یک تابع مشتق‌پذیر است. بنابر (۲۲.۳) مشاهده می‌شود که

$$\frac{d\Psi}{dt}=\frac{\partial\Psi}{\partial y}\frac{dy}{dt}=\nabla\Psi(y)\Psi(y).$$

با محاسبه مشتق $E(t)$ نسبت به $y(t)$ در شبکه عصبی (۲۰.۳) و (۲۱.۳) داریم:

$$\begin{aligned}\frac{dE(y(t))}{dt}&=(\frac{d\Psi}{dt})^T\Psi+\Psi^T(\frac{d\Psi}{dt})+(y-y^*)^T\frac{dy(t)}{dt}\\ &=\Psi^T(\nabla\Psi(y)^T+\nabla\Psi(y))\Psi+(y-y^*)^T\Psi(y). \qquad (۴۵.۳)\end{aligned}$$

با استفاده از لم (۲.۳.۳) داریم:

$$\Psi^T(y)(\nabla\Psi(y)^T + \nabla\Psi(y))\Psi(y) \leq ۰, \forall y \neq y^*.$$

علاوه‌براین از تعریف ۲.۲ و لم ۳.۲ در [۳۸] داریم:

$$(y-y^*)^T(\Psi(y)-\Psi(y^*)) = (y-y^*)^T\Psi(y) \leq ۰, \forall y \neq y^*,$$

بنابراین

$$\frac{dE(y(t))}{dt} \leq ۰. \tag{۴۶.۳}$$

لذا شبکه عصبی (۲۰.۳) و (۲۱.۳) پایدار به مفهوم لیاپانوف است. □

لم ۴.۳.۳.

i) به ازای هر نقطه آغازین $y(t_۰) = (\theta(t_۰)^T, \nu(t_۰)^T, \vartheta(t_۰)^T)^T$، یک جواب پیوسته $y(t) = (\theta(t)^T, \nu(t)^T, \vartheta(t)^T)^T$، برای سیستم (۲۰.۳) و (۲۱.۳) وجود دارد.

ii) فرض کنید $y(t) = (\theta(t)^T, \nu(t)^T, \vartheta(t)^T)^T$ مسیر جواب سیستم (۲۰.۳) و (۲۱.۳) با نقطه آغازین $y(t_۰) = (\theta(t_۰)^T, \nu(t_۰)^T, \vartheta(t_۰)^T)^T$ باشد. اگر $\nu(t_۰) \geq ۰$ آن‌گاه $\nu(t) \geq ۰$.

برهان.

i) به‌راحتی قابل بررسی است که $Q\theta + d + A^T(\nu + A\theta - b)^+ + e\vartheta$، $(\nu + A\theta - b)^+ - \nu$ و $e^T\theta$ به‌صورت محلی، پیوسته لیپ شیتز هستند. با توجه به وجود جواب محلی معادلات دیفرانسیل معمولی در [۳۰]، شبکه عصبی (۲۰.۳) و (۲۱.۳) دارای یک جواب پیوسته یکتا $y(t)$، $t \in [t_۰, \eta]$ برای مقادیر $\eta > t_۰$ است.

با توجه به اثبات قضیه ۳.۳.۳، می‌دانیم که E یک تابع ناصعودی نسبت به t است، لذا

$$\frac{۱}{۲}\|y-y^*\|^۲ \leq E(y(t)) \leq E(y(t_۰)), \quad \forall t \geq t_۰.$$

این نشان‌دهنده این مطلب است که مسیر جواب شبکه عصبی ۲۰.۳ و ۲۱.۳ کران‌دار است. بنابراین $\eta \to +\infty$.

ii) برای هر نقطه آغازین $y(t_۰)$ با $\nu(t_۰) \geq ۰$ داریم:

$$\frac{d\nu}{dt} + \nu = (\nu + A\theta - b)^+,$$
$$\int_{t_۰}^{t}(\frac{d\nu}{dt} + \nu)e^s ds = \int_{t_۰}^{t} e^s(\nu + A\theta - b)^+ ds.$$

بنابراین

$$\nu(t) = e^{-(t-t_\circ)}\nu(t_\circ) + e^{-t}\int_{t_\circ}^{t} e^s(\nu + A\theta - b)^+ ds.$$

با توجه به این‌که $(\nu + A\theta - b)^+ \geq \circ$، لذا $\nu(t) \geq \circ$ به ازای هر $t \geq t_\circ$.

□

قضیه ۴.۳.۳. به ازای هر نقطه آغازین $y_\circ = (\theta(t_\circ)^T, \nu(t_\circ)^T, \vartheta(t_\circ)^T)^T \in R^{۳N+۱}$ مسیر جواب شبکه عصبی (۲۰.۳) و (۲۱.۳) به یک نقطه تعادل همگرا است. به‌ویژه زمانی که D^* که مجموعه نقاط بهینه مساله (۱۷.۳) و (۱۸.۳) و دوگان آن است دارای نقطه تعادل یکتا باشد، شبکه عصبی (۲۰.۳) و (۲۱.۳) به‌ازای هر نقطه آغازین
$y_\circ = (\theta(t_\circ)^T, \nu(t_\circ)^T, \vartheta(t_\circ)^T)^T \in R^{۳N+۱}$ پایدار مجانبی سراسری است.

برهان. با توجه به اثبات لم (۴.۳.۳) مسیر جواب $(\theta(t)^T, \nu(t)^T, \vartheta(t)^T)^T$ در شبکه عصبی (۲۰.۳) و (۲۱.۳) کران‌دار است. لذا دنباله‌ی صعودی مانند t_k به طوری‌که $t_k \to \infty$ وقتی $k \to \infty$ و یک نقطه‌ی حدی $(\bar\theta^T, \bar\nu^T, \bar\vartheta^T)^T$ وجود دارد به طوری‌که :

$$\lim_{k\to\infty} (\theta(t_k)^T, \nu(t_k)^T, \vartheta(t_k)^T)^T = (\bar\theta^T, \bar\nu^T, \bar\vartheta^T)^T.$$

با به‌کار بردن اصل تغییر ناپذیری لِسال، زمانی که $t \to \infty$، $\{(\theta(t)^T, \nu(t)^T, \vartheta(t)^T)^T \to M\}$ که در آن M بزرگ‌ترین مجموعه‌ی تغییر ناپذیر در $K = \{(\theta(t)^T, \nu(t)^T, \vartheta(t)^T)^T | \frac{dE(y(t))}{dt} = \circ\}$ است. از روابط (۲۰.۳) و (۲۱.۳) و (۴۶.۳) نتیجه می‌شود $\frac{d\theta}{dt} = \circ$، $\frac{d\nu}{dt} = \circ$ و $\frac{d\vartheta}{dt} = \circ$ اگر و فقط اگر $\frac{dE(y(t))}{dt} = \circ$. بنابراین $(\bar\theta^T, \bar\nu^T, \bar\vartheta^T)^T \in D^*$ که در آن $M \subseteq K \subseteq D^*$ است.

ثانیاً ثابت می‌کنیم مسیر جواب $(\theta(t)^T, \nu(t)^T, \vartheta(t)^T)^T$ همگرای سراسری به نقطه تعادل $(\bar\theta^T, \bar\nu^T, \bar\vartheta^T)^T$ است. تابع لیاپانوف را به‌شکل زیر تعریف می‌کنیم:

$$\bar E(y) = \|\Psi(y)\|^۲ + \frac{۱}{۲}\|y - \bar y\|^۲.$$

در رابطه (۴۴.۳)، $\theta^* = \bar\theta$، $\nu^* = \bar\nu$ و $\vartheta^* = \bar\vartheta$ را جایگزین می‌کنیم. با توجه به این‌که $\bar E(y)$ به‌طور پیوسته مشتق پذیر و $\bar E(\bar y) = \circ$ است، لذا

$$\lim_{k\to\infty} (\theta(t_k)^T, \nu(t_k)^T, \vartheta(t_k)^T)^T = (\bar\theta^T, \bar\nu^T, \bar\vartheta^T)^T.$$

پس داریم $\lim_{k\to\infty} \bar E(\theta(t_k)^T, \nu(t_k)^T, \vartheta(t_k)^T)^T = \bar E(\bar\theta^T, \bar\nu^T, \bar\vartheta^T)^T$. بنابراین، برای هر $\epsilon > \circ$ وجود دارد $q > \circ$ به‌طوری که برای تمامی $t \geq t_q$ داریم $\bar E(y(t)) < \epsilon$. به‌طور مشابه، می‌توان به‌دست آورد $\frac{d\bar E(y(t))}{dt} \leq \circ$. پس نتیجه می‌گیریم به‌ازای $t \geq t_q$،

$$\frac{۱}{۲}\|y(t) - \bar y\|^۲ \leq \bar E(y(t)) \leq \epsilon.$$

با توجه به مطالب بالا $\lim_{t\to\infty} \|y(t) - \bar y\| = \circ$ و در نتیجه $\lim_{t\to\infty} y(t) = \bar y$. بنابراین سیستم (۲۰.۳) و (۲۱.۳) همگرای سراسری به یک نقطه تعادل مثل $\bar y = (\bar\theta^T, \bar\nu^T, \bar\vartheta^T)^T$ است، که در آن $\bar\theta$ جواب بهینه مساله (۱۷.۳) و (۱۸.۳) است.

در حالت خاص اگر $D^* = \{(\theta^{*T}, \nu^{*T}, \vartheta^{*T})^T\}$، آن‌گاه شبکه عصبی (۲۰.۳) و (۲۱.۳) برای حل مساله (۱۷.۳) و (۱۸.۳)، پایدار مجانبی سراسری به نقطه تعادل $y^* = \{(\theta^{*T}, \nu^{*T}, \vartheta^{*T})^T\}$ است. در این‌جا D^* مجموعه جواب بهینه مساله (۱۷.۳) و (۱۸.۳) و دوگان آن است. □

لم ۵.۳.۳. با افزایش τ سرعت همگرایی شبکه عصبی (۲۰.۳) و (۲۱.۳) افزایش می‌یابد.

برهان. با توجه به روابط (۴۵.۳) ـ (۴۶.۳) به‌ازای هر $\tau > 0$ در (۲۰.۳) و (۲۱.۳) داریم:

$$\frac{dE(y)}{dt} \leq \tau \Psi(y)^T (\nabla \Psi(y)^T + \nabla \Psi(y)) \Psi(y) \leq 0.$$

آن‌گاه

$$E(y(t)) \leq E(y(t_0)) + \tau \int_{t_0}^{t} \Psi(y(s))^T (\nabla \Psi(y(s))^T + (\nabla \Psi(y(s))) \Psi(y(s)) ds.$$

چون $E(y) \geq \frac{1}{2} \|y - y^*\|^2$ که در آن y^* یک نقطه KKT مساله (۱۷.۳) و (۱۸.۳) است، داریم:

$$\|y(t) - y^*\|^2 \leq 2E(y(t_0)) + 2\tau \int_{t_0}^{t} \Psi(y(s))^T (\nabla \Psi(y(s))^T + (\nabla \Psi(y(s))) \Psi(y(s)) ds.$$

بنابراین با افزایش τ سرعت همگرایی $y(t)$ افزایش پیدا می‌کند.

□

۴.۳ مثال‌های عددی

برای نشان‌دادن کارایی شبکه عصبی مورد نظر، مثال‌هایی را در این بخش ارائه می‌دهیم. شبیه سازی این مثال ها با استفاده از نرم افزار $Matlab\ 7$ و حل کننده معادلات دیفرانسیل معمولی $ode45$ انجام شده است.

برای مشخص کردن مقدار مناسب برای پارامترهای C و ϵ از روش $Leave-One-Out$ در **اعتبارسنجی‌متقابل**[1] استفاده می‌کنیم. در این روش، از مجموعه داده‌های آموزشی یک مشاهده خارج شده و براساس بقیه مشاهدات، پارامترها برآورد می‌شود. سپس میزان خطای مدل برای مشاهده خارج شده، محاسبه می‌شود. از آن جایی که در این روش در هر مرحله از فرآیند CV فقط یکی از مشاهدات خارج می‌شود، تعداد مراحل تکرار فرآیند CV با تعداد داده‌های آموزشی برابر است. همان‌طور که در شکل (۵.۳) نشان داده شده است، در مرحله iام، مشاهده iام از مجموعه داده‌های آموزشی خارج شده و مدل توسط تابع $Interpolate$ برآورد شده و مقدار تخمینی برای متغیر پاسخ در مشاهده iام توسط مدل به‌دست می‌آید. مربع فاصله مقدار واقعی متغیر پاسخ $(y[i])$ از مقدار برآورد شده $(\mathrm{y\_out})$، نیز به عنوان برآورد خطای مدل در نظر گرفته شده و در انتها، میانگین خطای همه مدل‌ها به عنوان برآورد خطای کلی محاسبه می‌شود که به آن میانگین توان دوم خطای پیش‌بینی[2] می‌گویند. واضح است که هرچه مقدار $MSPE$ کمتر باشد، دقت برآورد بالاتر می‌رود.

```
Input:

x, {vector of length N with x-values of data points}

y, {vector of length N with y-values of data points}

Output:

err, {estimate for the prediction error}

Steps:

err ← 0

for i ← 1, . . . , N do

// define the cross-validation subsets

x_in ← (x[1], . . . , x[i - 1], x[i + 1], . . . , x[N])

y_in ← (y[1], . . . , y[i - 1], y[i + 1], . . . , y[N]

x_out ← x[i]

interpolate(x_in, y_in, x_out, y_out)

err ← err + (y[i] - y_out)^2

end for

err ← err/N
```

شکل ۵.۳: شبه کد اجرای روش $Leave-One-Out$ در اعتبار سنجی متقابل (CV)

[1] CV
[2] MSPE

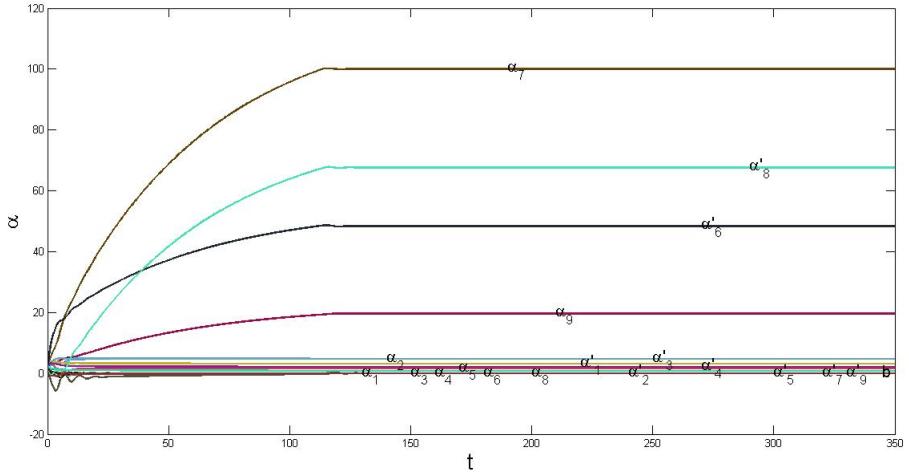
جدول ۱.۳: داده‌های رگرسیون

x	۵-	۳-	۲-	۱.۶	۳.۸	۱۰.۲	۱۱	۱۱.۵	۱۲.۷
y	۱.۶-	۱.۸	۱-	۱.۲-	۲.۲	۶.۸-	۹	۱۰	۱۰

مثال ۱.۴.۳. داده‌های رگرسیون جدول (۱.۳) را در نظر می‌گیریم.

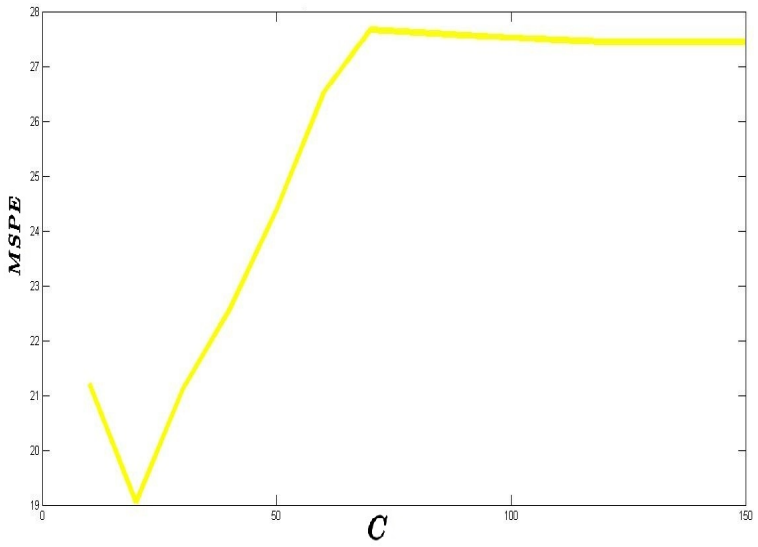
با توجه به غیرخطی بودن این مساله رگرسیون از کرنل تابع پایه شعاعی ($\Phi(x)^T\Phi(y) = K(x,y) = \exp(-\frac{\|x-y\|^2}{2\delta^2})$) برای رگرسیون استفاده می‌کنیم.

شکل (۶.۳) مسیرهای همگرایی $\alpha_1, \alpha_2 ..., \alpha_9, \alpha'_1, \alpha'_2, ..., \alpha'_9$ را نشان می‌دهد.

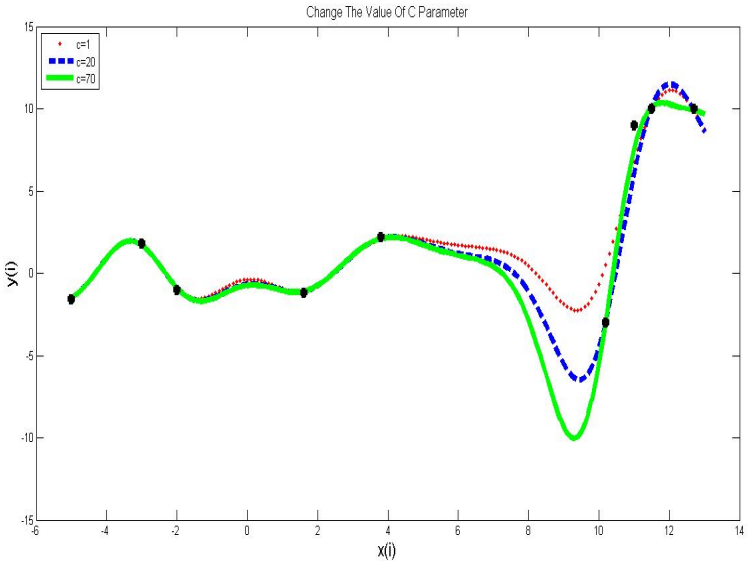


شکل ۶.۳: رفتار گذر $\alpha_1, \alpha_2 ..., \alpha_9, \alpha'_1, \alpha'_2, ..., \alpha'_9$ در شبکه عصبی (۲۰.۳) و (۲۱.۳)

در ادامه با تغییر یک پارامتر و ثابت نگه‌داشتن سایر پارامترها، رفتار $f(x)$ را بررسی می‌کنیم.
در شکل (۷.۳) مقدار $MSPE$ به‌ازای مقادیر مختلف C نشان داده شده است. شکل (۸.۳) نیز $f(x)$ را به‌ازای سه مقدار مختلف C نشان می‌دهد. با توجه به نمودار شکل (۷.۳)، به‌ازای مقدار $C = ۲۰$ دقت برآورد $f(x)$ که در شکل (۸.۳) نشان داده شده است، از دو مقدار دیگر C بیشتر است.

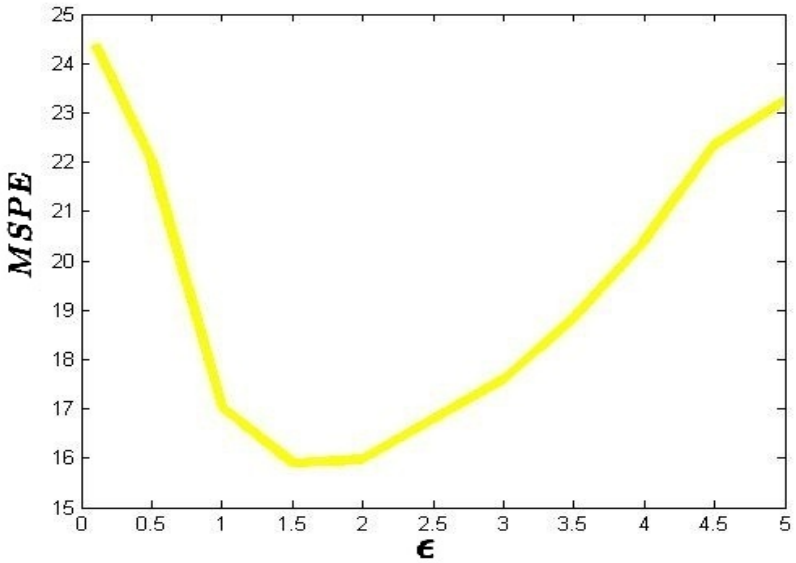


شکل ۷.۳: تغییرات $MSPE$ نسبت به مقادیر مختلف C

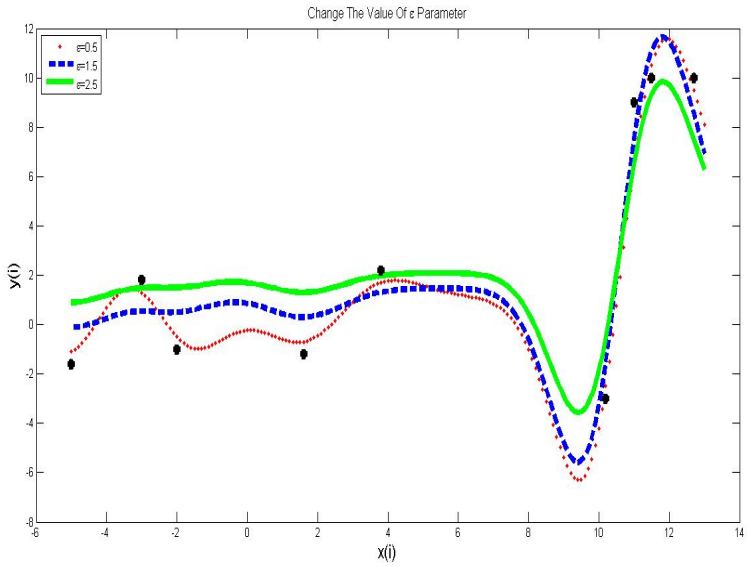


شکل ۸.۳: نتایج رگرسیون بردار پشتیبان به‌ازای سه مقدار مختلف C با استفاده از شبکه عصبی (۲۰.۳) و (۲۱.۳)

در شکل (۹.۳) مقدار $MSPE$ به‌ازای مقادیر مختلف ϵ نشان داده شده است. شکل (۱۰.۳) نیز $f(x)$ را به‌ازای مقادیر مختلف ϵ نشان می‌دهد. با توجه به نمودار شکل (۹.۳)، به‌ازای مقدار $\epsilon = ۱/۵$ دقت برآورد $f(x)$ که در شکل (۱۰.۳) نشان داده شده است، از دو مقدار دیگر ϵ بیشتر است.



شکل ۹.۳: تغییرات $MSPE$ نسبت به مقادیر مختلف ϵ



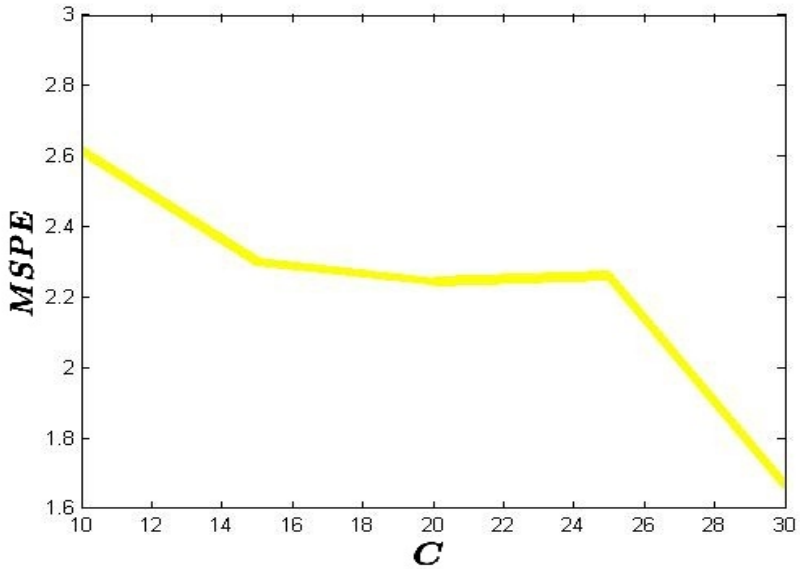
شکل ۱۰.۳: نتایج رگرسیون بردار پشتیبان به‌ازای سه مقدار مختلف ϵ با استفاده از شبکه عصبی (۲۰.۳) و (۲۱.۳)

مثال ۲.۴.۳. در این مثال مساله رگرسیون خطی را با شبکه عصبی (۲۰.۳) و (۲۱.۳) حل می کنیم. داده‌های جدول (۲.۳) را در نظر بگیرید.

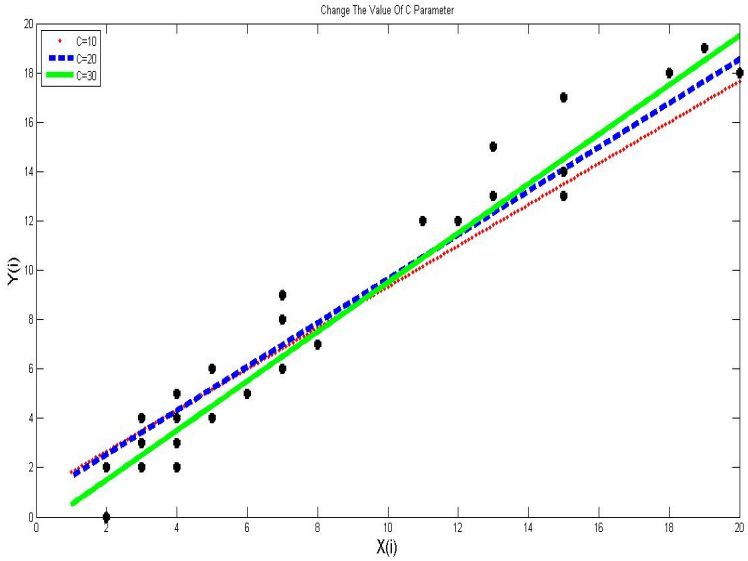
جدول ۲.۳: داده‌های رگرسیون

Y_i	X_i	i	Y_i	X_i	i
۱۷	۱۵	۱۶	۱۳	۱۵	۱
۱۲	۱۲	۱۷	۱۲	۱۱	۲
۳	۴	۱۸	۳	۳	۳
۴	۵	۱۹	۱۵	۱۳	۴
۲	۲	۲۰	۳	۳	۵
۱۹	۱۹	۲۱	۴	۳	۶
۱۴	۱۵	۲۲	۰	۲	۷
۱۲	۱۲	۲۳	۲	۳	۸
۸	۷	۲۴	۲	۴	۹
۴	۴	۲۵	۵	۴	۱۰
۱۳	۱۳	۲۶	۹	۷	۱۱
۱۸	۲۰	۲۷	۶	۷	۱۲
۲	۴	۲۸	۶	۵	۱۳
۵	۶	۲۹	۵	۶	۱۴
۷	۸	۳۰	۱۸	۱۸	۱۵

در شکل (۱۱.۳) مقدار $MSPE$ به‌ازای مقادیر مختلف C نشان داده شده است. شکل (۱۲.۳) نیز $f(x)$ را به‌ازای سه مقدار مختلف C نشان می‌دهد. با توجه به نمودار شکل (۱۱.۳)، به‌ازای مقدار $C = ۳۰$ دقت برآورد $f(x)$ که در شکل (۱۲.۳) نشان داده شده است، از دو مقدار دیگر C بیشتر است.

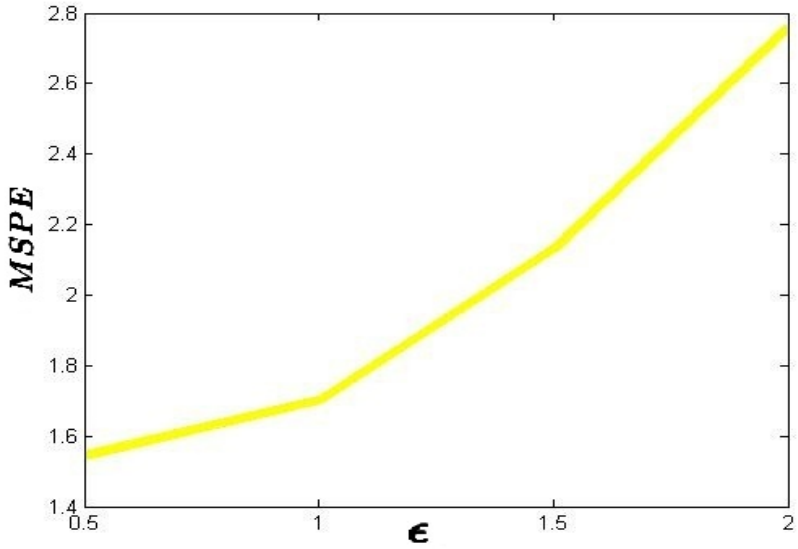


شکل ۱۱.۳: تغییرات $MSPE$ نسبت به مقادیر مختلف C

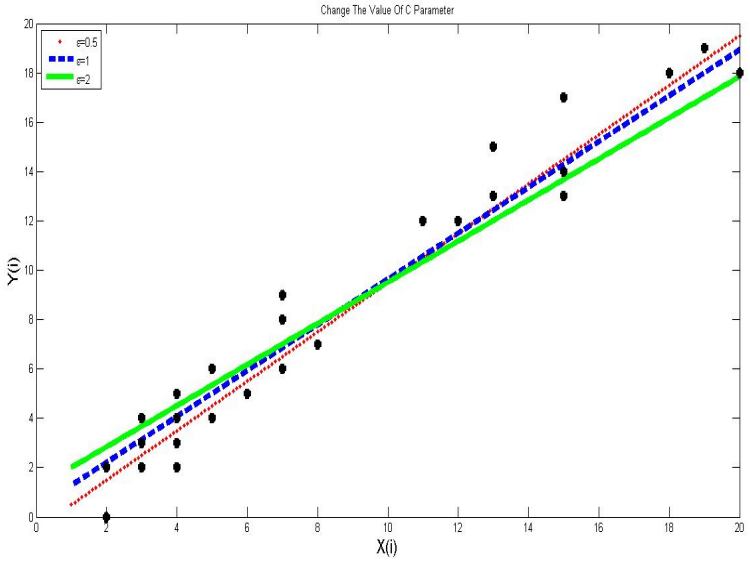


شکل ۱۲.۳: نتایج رگرسیون بردار پشتیبان به‌ازای سه مقدار مختلف C با استفاده از شبکه عصبی (۲۰.۳) و (۲۱.۳)

در شکل (۱۳.۳) مقدار $MSPE$ به‌ازای مقادیر مختلف ϵ نشان داده شده است. شکل (۱۴.۳) نیز $f(x)$ را به‌ازای مقادیر مختلف ϵ نشان می‌دهد. با توجه به نمودار شکل (۱۳.۳)، به‌ازای مقدار $\epsilon = ۰/۵$ دقت برآورد $f(x)$ که در شکل (۱۴.۳) نشان داده شده است، از دو مقدار دیگر ϵ بیشتر است.



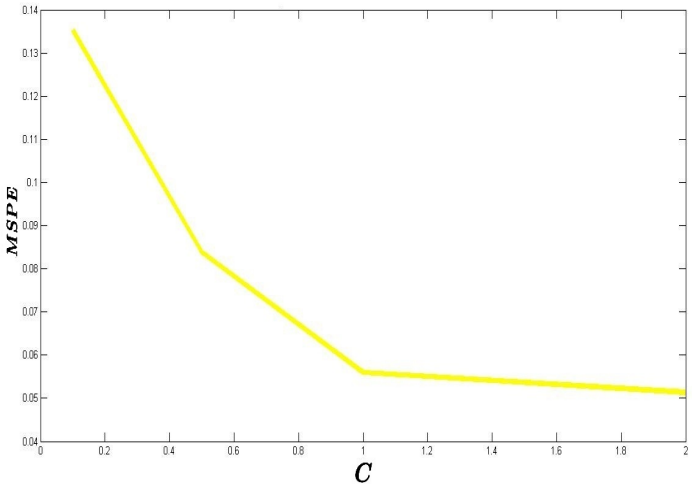
شکل ۱۳.۳: تغییرات $MSPE$ نسبت به مقادیر مختلف ϵ



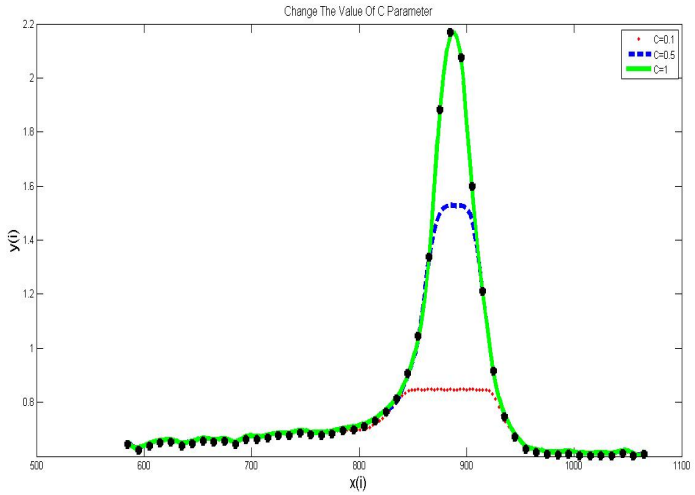
شکل ۱۴.۳: نتایج رگرسیون بردار پشتیبان به‌ازای سه مقدار مختلف ϵ با استفاده از شبکه عصبی (۲۰.۳) و (۲۱.۳)

مثال ۳.۴.۳. در این مثال، نتیجه رگرسیون را برای داده‌های رگرسیون تیتانیوم[۱] [۱۳] با استفاده از شبکه عصبی ارائه شده نشان می‌دهیم. کرنل استفاده شده مانند مثال (۱.۴.۳) است.

در شکل (۱۵.۳) مقدار $MSPE$ به‌ازای مقادیر مختلف C نشان داده شده است. شکل (۱۶.۳) نیز $f(x)$ را به‌ازای سه مقدار مختلف C نشان می‌دهد. با توجه به نمودار شکل (۱۵.۳)، به‌ازای مقدار $C = ۱$ دقت برآورد $f(x)$ که در شکل (۱۶.۳) نشان داده شده است، از دو مقدار دیگر C بیشتر است.



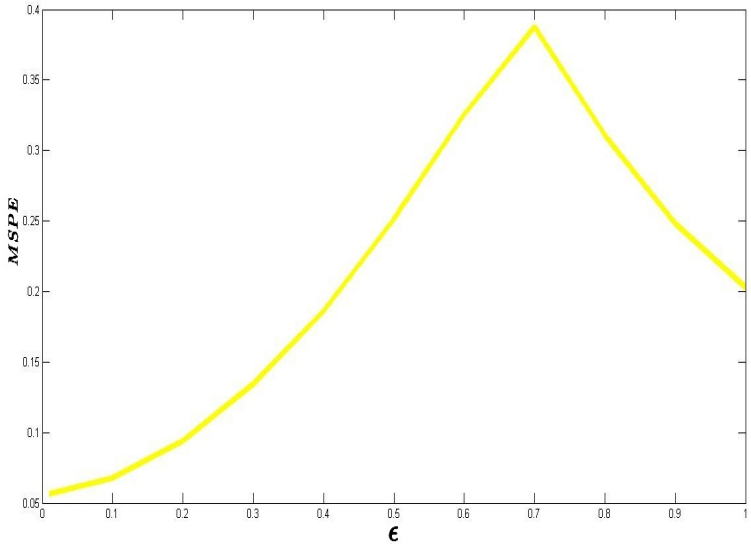
شکل ۱۵.۳: تغییرات $MSPE$ نسبت به مقادیر مختلف C



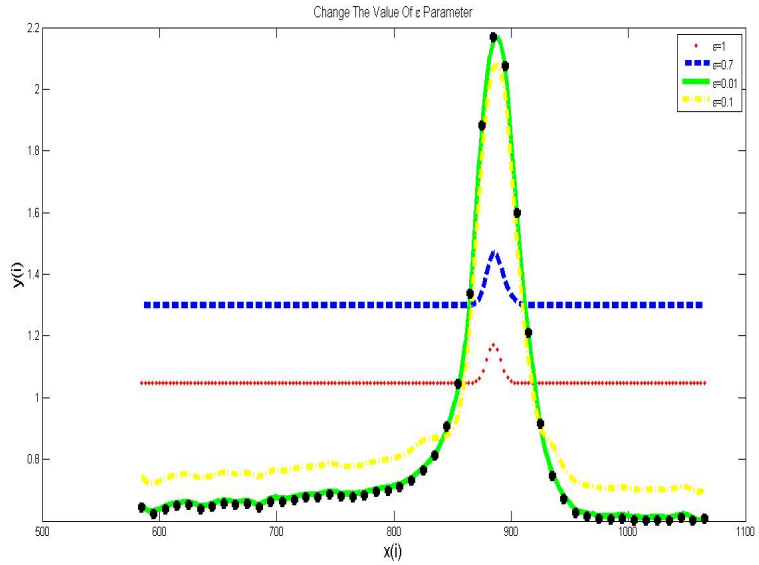
شکل ۱۶.۳: نتایج رگرسیون بردار پشتیبان به‌ازای سه مقدار مختلف C با استفاده از شبکه عصبی (۲۰.۳) و (۲۱.۳)

[۱] Titanium

در شکل (۱۷.۳) مقدار $MSPE$ به‌ازای مقادیر مختلف ϵ نشان داده شده است. شکل (۱۸.۳) نیز $f(x)$ را به‌ازای چهار مقدار مختلف ϵ نشان می‌دهد. با توجه به نمودار شکل (۱۷.۳)، به‌ازای مقدار $\epsilon = ۰/۰۱$ دقت برآورد $f(x)$ که در شکل (۱۸.۳) نشان داده شده است، از سه مقدار دیگر ϵ بیشتر است.



شکل ۱۷.۳: تغییرات $MSPE$ نسبت به مقادیر مختلف ϵ



شکل ۱۸.۳: نتایج رگرسیون بردار پشتیبان به‌ازای مقادیر مختلف ϵ با استفاده از شبکه عصبی (۲۰.۳) و (۲۱.۳)

فصل ۴

حل مسائل رگرسیون بردار پشتیبان تصادفی با قیدهای احتمالی

در این فصل شبکه عصبی برای حل مساله رگرسیون بردار پشتیبان با قیدهای احتمالی را ارائه می‌دهیم. ثابت می‌شود که نقطه تعادل شبکه عصبی مورد نظر با نقطه بهینه مساله درجه دوم رگرسیون بردار پشتیبان برابر است. وجود داشتن و منحصر به‌فرد بودن نقطه تعادل ثابت شده و با ساخت تابع لیاپانوف مناسب پایداری مجانبی اثبات می‌گردد. در نهایت با ارائه مثال‌های عددی کارایی این شبکه بررسی می‌شود.

۱.۴ مسائل رگرسیون بردار پشتیبان تصادفی با قیدهای احتمالی

۱.۱.۴ مسائل رگرسیون بردار پشتیبان تصادفی با قیدهای احتمالی در حالت خطی

همان‌طور که در مساله (۴.۳) و (۵.۳) در قبل نشان داده شد، مساله رگرسیون به‌صورت زیر نوشته می‌شود:

$$\text{minimize } \frac{۱}{۲}\|w\|^۲ + C\sum_{i=۱}^{n}(\zeta_i + \zeta_i'), \tag{۱.۴}$$

$$\text{subject to } \begin{cases} y_i - < w, x_i > -b \le \epsilon + \zeta_i, & i = ۱, ..., n, \\ < w, x_i > +b - y_i \le \epsilon + \zeta_i', & i = ۱, ..., n, \\ \zeta_i, \zeta_i' \ge ۰, & i = ۱, ..., n. \end{cases} \tag{۲.۴}$$

در رگرسیون عملی داده‌های آموزشی حاوی داده‌های ورودی و یا خروجی را به دلیل خطاهای نمونه‌گیری، خطاهای مدل‌سازی و یا خطاهای اندازه‌گیری، نمی‌توان به‌طور دقیق مشاهده کرد لذا آن‌ها را به‌صورت متغیر تصادفی بیان می‌کنند. فرض کنید نمونه‌های آموزشی $\{(X_۱, Y_۱), ..., (X_n, Y_n)\}$ متغیرهای تصادفی هستند، هم‌چنین $X_i = [X_i^۱, ..., X_i^m]^T$ بردار تصادفی در فضای ورودی با امید ریاضی $E(X_i) = [E(X_i^۱), ..., E(X_i^m)]^T$ و Y_i بردار تصادفی در فضای ورودی با امید ریاضی $E(Y_i)$ است. بنابراین قیود در رابطه (۲.۴) باید به‌صورت احتمالی نوشته شوند. لذا ابرصفحه بهینه رگرسیون با حل مساله‌ی زیر به‌دست می‌آید:

$$\text{minimize } \frac{۱}{۲}\|w\|^۲ + C\sum_{i=۱}^{n}(\zeta_i + \zeta_i'), \tag{۳.۴}$$

$$\text{subject to } \begin{cases} P_r(Y_i - w^T X_i - b \le \varepsilon + \zeta_i) \ge \delta, \\ P_r(w^T X_i + b - Y_i \le \varepsilon + \zeta_i') \ge \delta, \\ \zeta_i, \zeta_i' \ge ۰, \quad i = ۱, ..., n. \end{cases} \tag{۴.۴}$$

که در آن $\delta \in [۰, ۱]$.

با توجه به این‌که حل مساله بهینه سازی (۳.۴) و (۴.۴) با قیود نامساوی احتمالی دشوار است، لذا در ادامه شرایط کافی برای برقراری این نامساوی احتمالی را به‌دست می‌آوریم و مساله بهینه سازی احتمالی را تبدیل به یک مساله بهینه‌سازی درجه دوم می‌کنیم.

قضیه ۱.۱.۴. فرض کنید T متغیر تصادفی در بازه $[c, a]$ باشد آن‌گاه قضیه زیر برقرار است:

$$\left| P_r(T \le t) - \frac{a - E(T)}{a - c} \right| \le \frac{۱}{۲} + \frac{\left| t - \frac{c+a}{۲} \right|}{a - c}, \tag{۵.۴}$$

به‌ازای هر t در $[c, a]$.

برهان. مرجع [۱۰] را ببینید. □

ملاحظه ۱.۱.۴. با توجه به این واقعیت که $P_r(T \le t) = ۱ - P_r(T > t)$، از نامساوی (۵.۴) نامساوی معادل زیر را به‌ازای هر $t \in [c, a]$ به‌دست می‌آوریم:

$$\left| P_r(T \ge t) - \frac{E(T) - c}{a - c} \right| \le \frac{۱}{۲} + \frac{\left| t - \frac{c+a}{۲} \right|}{a - c}.$$

قضیه ۲.۱.۴. فرض کنید $X_i = [X_i^1, ..., X_i^m]^T$ بردار تصادفی با امید ریاضی $E(X_i) = [E(X_i^1), ..., E(X_i^m)]^T$ و Y_i بردار تصادفی با امید ریاضی $E(Y_i)$ باشد، آن گاه یک شرط کافی برای برقراری

$$P_r(Y_i - w^T X_i - b \leq \varepsilon + \zeta_i) \geq \delta,$$

عبارت است از :

$$E(Y_i) - w^T E(X_i) - b \leq \zeta_i - 2a\delta - \varepsilon.$$

هم‌چنین شرط کافی برای برقراری

$$P_r(w^T X_i + b - Y_i \leq \varepsilon + \zeta_i') \geq \delta,$$

عبارت است از

$$w^T E(X_i) + b - E(Y_i) \leq \zeta_i' - 2a\delta - \varepsilon,$$

که در آن $a > \varepsilon$.

برهان. مرجع [۸] را ببینید. □

بنابراین ابرصفحه بهینه رگرسیون را می‌توان با حل مساله زیر به‌دست آورد:

$$\text{minimize } \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\zeta_i + \zeta_{i'}), \tag{۶.۴}$$

$$\text{subject to } \begin{cases} E(Y_i) - w^T E(X_i) - b \leq \zeta_i - 2a\delta - \varepsilon, \\ w^T E(X_i) + b - E(Y_i) \leq \zeta_i' - 2a\delta - \varepsilon, \\ \zeta_i, \zeta_i' \geq 0, \ i = 1, ..., n. \end{cases} \tag{۷.۴}$$

که در آن ϵ پارامتر انحراف و C پارامتر جریمه است. برای ساختن دوگان مساله بهینه‌سازی (۶.۴) و (۷.۴) متغیرهای کمکی (α_i, α_i') را معرفی می‌کنیم. α_i متناظر محدودیت‌های شامل ζ_i و α_i' متناظر محدودیت‌های کمکی ζ_i' است. این متغیرها در حقیقت ضرایب لاگرانژ برای مساله اولیه هستند. مشابه روش‌های به‌کار رفته در طراحی رده‌بندی کننده‌های SV، مساله بهینه سازی مقید (۶.۴) و (۷.۴) را با تشکیل لاگرانژین متغیرهای اولیه حل می‌کنیم که تابع لاگرانژین آن به‌صورت زیر است:

$$L(w, b, \zeta, \zeta', \alpha, \alpha', \beta, \beta') = \frac{1}{2}w^T w + C\sum_{i=1}^{n}(\zeta_i + \zeta_i') - \sum_{i=1}^{n}(\beta_i\zeta_i + \beta_i'\zeta_i') + \sum_{i=1}^{n}\alpha_i[E(Y_i) -$$

$$w^T E(X_i) - b + (2a\delta + \varepsilon) - \zeta_i] + \sum_{i=1}^{n}\alpha_i'[w^T E(X_i) + b - E(Y_i) + (2a\delta + \varepsilon) - \zeta_i'].$$

تابع لاگرانژین $L_P(w,b,\zeta_i,\zeta_i',\alpha_i,\alpha_i',\beta_i,\beta_i')$ باید نسبت به متغیرهای اولیه w، b، ζ_i و ζ_i' می‌نیمم شود و نسبت به ضرایب لاگرانژ نامنفی α_i، α_i'، β_i و β_i' ماکزیمم گردد. بنابراین تابع لاگرانژین در نقطه جواب بهینه مساله (۶.۴) و (۷.۴) یعنی $(w^*,b^*,\zeta_i^*,\zeta_i'^*)$ دارای نقطه زینی است.

در جواب بهینه مشتقات جزئی نسبت به متغیرهای اولیه صفر می‌شود، یعنی :

$$\begin{aligned}
&\frac{\partial L}{\partial w} = w - \sum_{i=1}^{n}(\alpha_i-\alpha_i')E(x_i) = 0 \Rightarrow w = \sum_{i=1}^{n}(\alpha_i-\alpha_i')E(x_i), \qquad (8.4)\\
&\frac{\partial L}{\partial b} = \sum_{i=1}^{n}(\alpha_i-\alpha_i') = 0,\\
&\frac{\partial L}{\partial \zeta_i} = c-\alpha_i-\beta_i = 0, \qquad i=1,...,n,\\
&\frac{\partial L}{\partial \zeta_i'} = c-\alpha_i'-\beta_i' = 0, \qquad i=1,...,n.
\end{aligned}$$

بنابراین همانند مساله رگرسیون که در فصل قبل بیان شد، دوگان مساله (۶.۴) و (۷.۴) به‌شکل زیر نوشته می‌شود:

$$\begin{aligned}
\text{maximize} \quad & -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\alpha_i-\alpha_i')(\alpha_j-\alpha_j')(E(X_i))^TE(X_j) + (2a\delta+\varepsilon)\sum_{i=1}^{n}(\alpha_i+\alpha_i')\\
& + \sum_{i=1}^{n}(\alpha_i-\alpha_i')E(Y_i), \qquad (9.4)
\end{aligned}$$

$$\text{subject to} \quad \begin{cases} \sum_{i=1}^{n}(\alpha_i-\alpha_i') = 0, \\ \alpha_i,\alpha_i' \in [0,C],\ i=1,...,n. \end{cases} \qquad (10.4)$$

با توجه به‌شرط بهینگی (۸.۴) تابع تخمین $f(x)$ به‌صورت زیر بیان می‌شود:

$$f(x) = \sum_{i=1}^{n}(\alpha_i-\alpha_i')E(X_i)^TE(X) + b.$$

۲.۱.۴ حل مسائل رگرسیون بردار پشتیبان تصادفی با قیدهای احتمالی در حالت غیرخطی

در این بخش نیز مانند حالت‌های غیرخطی که در (۳.۲.۱) بیان شد، با نگاشت $\Phi:\mathbb{R}^m\to\mathbb{R}^{m_1}$ داده‌ها را به فضایی با بعد بیشتر می‌بریم. بنابراین ابر صفحه بهینه رگرسیون بردار پشتیبان در حالت غیرخطی با قیدهای احتمالی با حل مساله زیر به‌دست می‌آید:

$$\text{minimize } \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\zeta_i+\zeta_i'),$$

$$\text{subject to} \begin{cases} P_r(Y_i - w^T\Phi(X_i) - b \le \varepsilon + \zeta_i) \ge \delta, \\ P_r(w^T\Phi(X_i) + b - Y_i \le \varepsilon + \zeta_i') \ge \delta, \\ \zeta_i, \zeta_i' \ge ۰, \ i = ۱, ..., n. \end{cases}$$

که در آن $\Phi(X_i) = [\Phi^۱(X_i), ..., \Phi^{m_۱}(X_i)]^T$ یک بردار تصادفی است.

قضیه ۳.۱.۴. فرض کنید $\Phi(X_i) = [\Phi^۱(X_i), ..., \Phi^{m_۱}(X_i)]^T$ بردار تصادفی با امید ریاضی $E(\Phi(X_i)) = [E(\Phi^۱(X_i)), ..., E(\Phi^{m_۱}(X_i))]^T$ و Y_i یک بردار تصادفی با امید ریاضی $E(Y_i)$ باشد، آن‌گاه شرط کافی برای برقراری

$$P_r(Y_i - w^T\Phi(X_i) - b \le \varepsilon + \zeta_i) \ge \delta,$$

عبارت است از:

$$E(Y_i) - w^T E(\Phi(X_i)) - b \le \zeta_i - ۲a\delta - \varepsilon.$$

هم‌چنین یک شرط کافی برای برقراری

$$P_r(w^T\Phi(X_i) + b - Y_i \le \varepsilon + \zeta_i') \ge \delta,$$

به‌صورت

$$w^T E(\Phi(X_i)) + b - E(Y_i) \le \zeta_i' - ۲a\delta - \varepsilon,$$

می‌باشد که در آن $a > \varepsilon$.

برهان. مرجع [۸] را ببینید.

□

بنابراین ابرصفحه بهینه با حل مساله زیر به‌دست می‌آید:

$$\text{minimize } \frac{۱}{۲} w^T w + C \sum_{i=۱}^{n} (\zeta_i + \zeta_i'), \tag{۱۱.۴}$$

$$\text{subject to} \begin{cases} E(Y_i) - w^T E(\Phi(X_i)) - b \le \zeta_i - ۲a\delta - \varepsilon, \\ w^T E(\Phi(X_i)) + b - E(Y_i) \le \zeta_i' - ۲a\delta - \varepsilon, \\ \zeta_i, \zeta_i' \ge ۰, i = ۱, ..., n. \end{cases} \tag{۱۲.۴}$$

که در آن ϵ پارامتر انحراف و C پارامتر جریمه است. برای ساختن دوگان مساله بهینه‌سازی(۱۱.۴) و (۱۲.۴) متغیرهای کمکی (α_i, α_i') را معرفی می‌کنیم. α_i متناظر محدودیت‌های شامل ζ_i و α_i' متناظر محدودیت‌های کمکی ζ_i' است. این متغیرها در حقیقت ضرایب لاگرانژ برای مساله اولیه هستند. مشابه روش‌های به‌کار رفته در طراحی رده‌بندی کننده‌های SV، مساله مقید

بهینه سازی (۱۱.۴) و (۱۲.۴) را با تشکیل لاگرانژین متغیرهای اولیه حل می‌کنیم که تابع لاگرانژین آن به‌صورت زیر است:

$$L(w,b,\zeta,\zeta',\alpha,\alpha',\beta,\beta') = \frac{۱}{۲}w^Tw + C\sum_{i=۱}^{n}(\zeta_i+\zeta_i') - \sum_{i=۱}^{n}(\beta_i\zeta_i+\beta_i'\zeta_i') + \sum_{i=۱}^{n}\alpha_i[E(Y_i) -$$
$$w^TE(\phi(X_i)) - b + (۲a\delta+\varepsilon) - \zeta_i] + \sum_{i=۱}^{n}\alpha_i'[w^TE(\phi(X_i)) + b - E(Y_i) + (۲a\delta+\varepsilon) - \zeta_i'].$$

تابع لاگرانژین $L_P(w,b,\zeta_i,\zeta_i',\alpha_i,\alpha_i',\beta_i,\beta_i')$ باید نسبت به متغیرهای اولیه w، b، ζ_i و ζ_i' می‌نیمم شود و نسبت به ضرایب لاگرانژ نامنفی α_i، α_i'، β_i و β_i' ماکزیمم گردد. بنابراین تابع لاگرانژین در نقطه جواب بهینه مساله (۱۱.۴) و (۱۲.۴) یعنی $(w^*,b^*,\zeta_i^*,\zeta_i'^*)$ دارای نقطه زینی است.

در جواب بهینه مشتقات جزئی نسبت به متغیرهای اولیه صفر می‌شود، یعنی :

$$\frac{\partial L}{\partial w} = w - \sum_{i=۱}^{n}(\alpha_i-\alpha_i')E(\phi(X_i)) = ۰ \Rightarrow w = \sum_{i=۱}^{n}(\alpha_i-\alpha_i')E(\phi(X_i)), \qquad (۱۳.۴)$$
$$\frac{\partial L}{\partial b} = \sum_{i=۱}^{n}(\alpha_i-\alpha_i') = ۰,$$
$$\frac{\partial L}{\partial \zeta_i} = c-\alpha_i-\beta_i = ۰, \qquad i = ۱,...,n,$$
$$\frac{\partial L}{\partial \zeta_i'} = c-\alpha_i'-\beta_i' = ۰, \qquad i = ۱,...,n.$$

مشابه حالت خطی، دوگان مساله (۱۱.۴) و (۱۲.۴) به‌صورت زیر نوشته می‌شود:

$$\text{maximize } -\frac{۱}{۲}\sum_{i=۱}^{n}\sum_{j=۱}^{n}(\alpha_i-\alpha_i')(\alpha_j-\alpha_j')E(\Phi(X_i))^TE(\Phi(X_j)) +$$
$$(۲a\delta+\varepsilon)\sum_{i=۱}^{n}(\alpha_i+\alpha_i') + \sum_{i=۱}^{n}(\alpha_i-\alpha_i')E(Y_i), \qquad (۱۴.۴)$$

$$\text{subject to } \begin{cases} \sum_{i=۱}^{n}(\alpha_i-\alpha_i') = ۰, \\ \alpha_i,\alpha_i' \in [۰,C],\ i = ۱,...,n. \end{cases} \qquad (۱۵.۴)$$

با توجه به شرط بهینگی (۱۳.۴)، تابع تخمین $f(x)$ به‌صورت زیر بیان می‌شود:

$$f(x) = \sum_{i=۱}^{n}(\alpha_i-\alpha_i')E(\Phi(X_i))^TE(\Phi(X)) + b.$$

۳.۱.۴ حل مساله رگرسیون به کمک شبکه عصبی با قیدهای احتمالی

مساله درجه دوم (۱۴.۴) و (۱۵.۴) را می‌توان به‌شکل زیر نوشت:

$$\text{minimize } \frac{۱}{۲}\theta^TQ\theta + d^T\theta, \qquad (۱۶.۴)$$

$$\text{subject to } \begin{cases} A\theta - b \leq \circ, \\ e^T\theta = \circ. \end{cases} \tag{۱۷.۴}$$

که در آن :

$$Q_{۴n\times ۴n} = \begin{cases} E(\Phi(X_i))^T E(\Phi(X_j)) & ۱ \leq i,j \leq n \\ \circ & \text{در غیر این‌صورت} \end{cases},$$

$$d = (-E(Y_۱), -E(Y_۲), ..., -E(Y_n), \underbrace{-۲a\delta - \epsilon, -۲a\delta - \epsilon, \cdots, -۲a\delta - \epsilon}_{n}, \underbrace{\circ, \circ, \cdots, \circ}_{۲n})^T,$$

$$e = (\underbrace{۱, ۱, \cdots, ۱}_{n}, \underbrace{\circ, \circ, \cdots, \circ}_{۳n})^T \in \mathbb{R}^{۴n},$$

$$A = \begin{pmatrix} I_{۲n\times ۲n} & O_{۲n\times ۲n} \\ O_{۲n\times ۲n} & -I_{۲n\times ۲n} \end{pmatrix},$$

$$b = (\underbrace{C, C, \cdots, C}_{n}, \underbrace{۲C, ۲C, \cdots, ۲C}_{n}, \underbrace{C, C, \cdots, C}_{n}, \underbrace{\circ, \circ, \cdots, \circ}_{n})^T \in \mathbb{R}^{۴n},$$

$$\theta = (\alpha_۱ - \alpha'_۱, \alpha_۲ - \alpha'_۲, ..., \alpha_n - \alpha'_n, \alpha_۱ + \alpha'_۱, \alpha_۲ + \alpha'_۲ ..., \alpha_n + \alpha'_n, \alpha_۱ - \alpha'_۱, \alpha_۲ - \alpha'_۲, ..., \alpha_n - \alpha'_n,$$
$$\alpha_۱ + \alpha'_۱, \alpha_۲ + \alpha'_۲, ..., \alpha_n + \alpha'_n)^T.$$

۲.۴ معرفی شبکه عصبی

در این بخش ما یک شبکه عصبی برای حل مساله (۱۶.۴) و (۱۷.۴) ارائه می‌دهیم. ابتدا برای ساده‌سازی معماری شبکه لم زیر را بیان می‌کنیم.

لم ۱.۲.۴. θ^* یک جواب بهینه مساله (۱۶.۴) و (۱۷.۴) است اگر و فقط اگر وجود داشته باشد $u^* \geq \circ$ به‌طوری که $(\theta^{*T}, u^{*T})^T$ در شرایط زیر صدق کند:

$$(I - \widetilde{P})[Q\theta^* + d + A^T u^*] + \widetilde{Q}(e^T\theta^*) = \circ, \tag{۱۸.۴}$$

$$(u^* + A\theta^* - b)^+ - u^* = \circ, \tag{۱۹.۴}$$

که در آن $\widetilde{P} = e(e^Te)^{-۱}e^T$ ، $\widetilde{Q} = e(e^Te)^{-۱}$ و $(u)^+ = ([u_۱]^+, \cdots, [u_p]^+)^T$ که $[u_i]^+ = \max\{\circ, u_i\}$ است.

برهان. اگر θ^* یک جواب بهینه (۱۶.۴) و (۱۷.۴) باشد، بنابر شرایط KKT برای بهینه سازی محدب [۱۱]، $(\theta^{*T}, u^{*T}, v^{*T})^T$ و $u^* \geq 0$ وجود دارد به‌طوری که معادله‌های زیر برقرار باشند.

$$Q\theta^* + d + ev^* + A^T u^* = 0, \tag{۲۰.۴}$$

$$e^T\theta^* = 0, \tag{۲۱.۴}$$

$$A\theta^* \leq b, \quad u^* \geq 0, \quad (u^*)^T(A\theta^* - b) = 0. \tag{۲۲.۴}$$

واضح است که (۱۹.۴) و (۲۲.۴) معادل‌اند، یعنی :

$$(A\theta^* - b) \leq 0, \quad u^* \geq 0, \quad (u^*)^T(A\theta^* - b) = 0 \iff (u^* + A\theta^* - b)^+ - u^* = 0.$$

در ادامه ما بایستی ثابت کنیم جواب‌های (۱۸.۴) و جواب‌های (۲۰.۴) و (۲۱.۴) معادل‌اند. از (۲۰.۴) داریم:

$$-e^T(Q\theta^* + d + ev^* + A^T u^*) = 0.$$

از جمع رابطه بالا با (۲۱.۴) داریم:

$$e^T\theta^* - e^T(Q\theta^* + d) - e^T ev^* - e^T A^T u^* = 0.$$

بنابراین

$$ev^* = e(e^Te)^{-1}(e^T\theta^*) - e(e^Te)^{-1}e^T(Q\theta^* + d + A^T u^*). \tag{۲۳.۴}$$

با جایگذاری (۲۳.۴) در (۲۰.۴) داریم:

$$[I - e(e^Te)^{-1}e^T](Q\theta^* + d + A^T u^*) + e(e^Te)^{-1}(e^T\theta^*) = 0.$$

اکنون قرار دهید $\widetilde{P} = e(e^Te)^{-1}e^T$ و $\widetilde{Q} = e(e^Te)^{-1}$ آن‌گاه معادله‌ی بالا را می‌توان به شکل زیر نوشت:

$$(I - \widetilde{P})[Q\theta^* + d + A^T u^*] + \widetilde{Q}(e^T\theta^*) = 0.$$

برعکس اگر u^* وجود داشته باشد به‌طوری که $(\theta^{*T}, u^{*T})^T$ در معادله (۱۸.۴) صدق کند، با ضرب e^T در دو طرف (۱۸.۴)، داریم:

$$e^T(I - \widetilde{P})[Q\theta^* + d + A^T u^*] + e^T\widetilde{Q}(e^T\theta^*) = 0.$$

توجه شود که $e^T(I - \widetilde{P}) = 0$ و $e^T\widetilde{Q} = I$، بنابراین داریم:

$$e^T\theta^* = 0,$$
$$(I - \widetilde{P})(Q\theta^* + d + A^T u^*) = 0.$$

قرار دهید $v^* = -(e^Te)^{-1}e^T(Q\theta^* + d + A^Tu^*)$. بنابراین

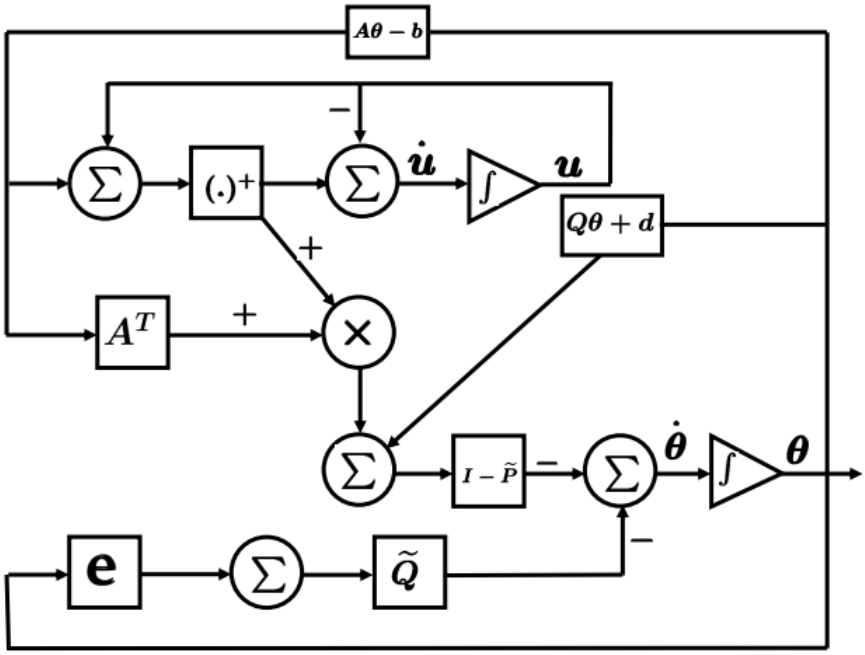
$$\begin{aligned} Q\theta^* + d + ev^* + A^Tu^* &= Q\theta^* + d + A^Tu^* - e(e^Te)^{-1}e^T(Q\theta^* + d + A^Tu^*) \\ &= (I - \widetilde{P})(Q\theta^* + d + A^Tu^*) = 0. \end{aligned}$$

در نتیجه $(\theta^{*T}, u^{*T}, v^{*T})^T$ وجود دارد به‌طوری که (۲۰.۴) و (۲۱.۴) در شرایط KKT صدق می‌کنند و این اثبات را کامل می‌کند. □

با توجه به لم ۱.۲.۴ ما یک شبکه عصبی کارا به‌صورت زیر برای حل مساله (۱۶.۴) و (۱۷.۴) پیشنهاد می‌کنیم:

$$\begin{cases} \dfrac{d\theta}{dt} = \eta\{-(I - \widetilde{P})[Q\theta + d + A^T(u + A\theta - b)^+] - \widetilde{Q}(e^T\theta)\}, \\ \dfrac{du}{dt} = \eta\{-u + (u + A\theta - b)^+\}, \end{cases} \tag{۲۴.۴}$$

که در آن η یک ضریب مقیاس است و نرخ همگرایی شبکه عصبی (۲۴.۴) را مشخص می‌کند. در این رساله $\eta = 1$ در نظر گرفته شده است. چگونگی پیاده‌سازی سخت افزاری شبکه عصبی (۲۴.۴) در شکل (۱.۴) نشان داده شده است.



شکل ۱.۴: یک نمودار ساده بلوکی شده برای شبکه عصبی (۲۴.۴)

ملاحظه ۱.۲.۴. با توجه به لم ۱.۲.۴ به راحتی فهمیده می‌شود که θ^* یک جواب بهینه (۱۶.۴) و (۱۷.۴) است اگر و فقط اگر $u^* \geq 0$ وجود داشته باشد به‌طوری که $(\theta^{*T}, u^{*T})^T$ یک نقطه تعادل شبکه (۲۴.۴) باشد. بنابراین زمانی که شبکه عصبی به نقطه تعادل همگرا است، مسیر $\theta(t)$ نیز به جواب بهینه مساله (۱۶.۴) و (۱۷.۴) همگراست.

۳.۴ تجزیه و تحلیل پایداری شبکه عصبی

در این بخش ما همگرایی سراسری شبکه عصبی (۲۴.۴) را ثابت می‌کنیم. فرض کنید Ω^e مجموعه نقاط تعادل شبکه عصبی (۲۴.۴) باشد.

لم ۱.۳.۴. فرض کنید $(\theta^{*T}, u^{*T})^T \in \Omega^e$ یک نقطه تعادل (۲۴.۴) است و

$$\begin{aligned} V(\theta,u) &= \frac{۱}{۲}\theta^T Q\theta + d^T\theta - (\frac{۱}{۲}\theta^{*T}Q\theta^* + d^T\theta^*) \\ &+ \frac{۱}{۲}\|(u+A\theta-b)^+\|^۲ - \frac{۱}{۲}\|(u^*+A\theta^*-b)^+\|^۲ \\ &- (\theta-\theta^*)^T(Q\theta^*+d+A^Tu^*) - (u-u^*)^Tu^* + \frac{۱}{۲}\|\theta-\theta^*\|^۲ + \frac{۱}{۲}\|u-u^*\|^۲. \end{aligned}$$

آن‌گاه:

$$(I)\ V(\theta,u) \geq \frac{۱}{۲}\|\theta-\theta^*\|^۲ + \frac{۱}{۲}\|u-u^*\|^۲,$$

$$(II)\ \frac{\mathrm{d}V}{\mathrm{d}t} \leq -\|(I-\widetilde{P})[Q\theta+d+A^T(u+A\theta-b)^+] + \widetilde{Q}(e^T\theta)\|^۲ - \frac{۱}{۲}\|u-(u+A\theta-b)^+\|^۲.$$

برهان. (I) به سادگی می‌توان بررسی کرد که $\frac{۱}{۲}\theta^TQ\theta + d^T\theta + \frac{۱}{۲}\|(u+A\theta-b)^+\|^۲$ یک تابع محدب مشتق پذیر است [۱۱] و

$$\nabla(\frac{۱}{۲}\theta^TQ\theta + d^T\theta + \frac{۱}{۲}\|(u+A\theta-b)^+\|^۲) = \begin{bmatrix} Q\theta+d+A^T(u+A\theta-b)^+ \\ (u+A\theta-b)^+ \end{bmatrix}.$$

بنابراین

$$\begin{aligned} \frac{۱}{۲}\theta^TQ\theta + d^T\theta - (\frac{۱}{۲}\theta^{*T}Q\theta^* + d^T\theta^*) + \frac{۱}{۲}\|(u+A\theta-b)^+\|^۲ - \frac{۱}{۲}\|(u^*+A\theta^*-d)^+\|^۲ \\ \geq (\theta-\theta^*)^T(Q\theta+d+A^Tu^*) + (u-u^*)^Tu^*. \end{aligned}$$

یعنی

$$V(\theta,u) \geq \frac{۱}{۲}\|\theta-\theta^*\|^۲ + \frac{۱}{۲}\|u-u^*\|^۲.$$

(II) اکنون مشتق V را نسبت به متغیر t محاسبه می‌کنیم،

$$\begin{aligned} \frac{dV}{dt} =& \frac{\partial V}{\partial\theta}\frac{d\theta}{dt} + \frac{\partial V}{\partial u}\frac{du}{dt} \\ =& -[Q\theta+d+A^T(u+A\theta-b)^+ - (Q\theta^*+d) - A^Tu^* + \theta - \theta^*]^T \end{aligned}$$

$$
\begin{aligned}
&\{(I-\widetilde{P})[Q\theta+d+A^T(u+A\theta-b)^+]+\widetilde{Q}(e^T\theta)\}\\
&-\frac{1}{2}[u-(u+A\theta-b)^+]^T[(u+A\theta-b)^+-2u^*+u]\\
=&-[Q(\theta-\theta^*)+\theta-\theta^*+A^T(u+A\theta-b)^+-A^Tu^*]^T\\
&\{(I-\widetilde{P})[Q(\theta-\theta^*)+A^T(u+A\theta-b)^+-A^Tu^*)]+\widetilde{P}(\theta-\theta^*)\}\\
&-\frac{1}{2}[u-(u+A\theta-b)^+]^T[u-(u+A\theta-b)^++2(u+A\theta-b)^+-2u^*]\\
=&-\{[Q(\theta-\theta^*)]^T(I-\widetilde{P})[Q(\theta-\theta^*)]\\
&+[Q(\theta-\theta^*)]^T(I-\widetilde{P})[A^T(u+A\theta-b)^+-A^Tu^*]\\
&+[Q(\theta-\theta^*)]^T\widetilde{P}(\theta-\theta^*)\\
&+(\theta-\theta^*)^T(I-\widetilde{P})[Q(\theta-\theta^*)]\\
&+(\theta-\theta^*)^T(I-\widetilde{P})[A^T(u+A\theta-b)^+-A^Tu^*]\\
&+(\theta-\theta^*)^T\widetilde{P}(\theta-\theta^*)\\
&+[A^T(u+A\theta-b)^+-A^Tu^*]^T(I-\widetilde{P})[Q(\theta-\theta^*)]\\
&+[A^T(u+A\theta-b)^+-A^Tu^*]^T(I-\widetilde{P})[A^T(u+A\theta-b)^+-A^Tu^*]\\
&+[A^T(u+A\theta-b)^+-A^Tu^*]^T\widetilde{P}(\theta-\theta^*)\}\\
&-\frac{1}{2}\|u-(u+A\theta-b)^+\|^2+[(u+A\theta-b)^+-u]^T[(u+A\theta-b)^+-u^*].
\end{aligned}
$$

با توجه به این‌که

$$(I-\widetilde{P})^2=I-\widetilde{P},\ \widetilde{P}^2=\widetilde{P},\ \widetilde{P}(I-\widetilde{P})=0,\ (u+A\theta-b)^+-u=A\theta-b+(-u+b-A\theta)^+$$

داریم:

$$
\begin{aligned}
\frac{dV}{dt}=\frac{dV}{dt}=&-[Q(\theta-\theta^*)]^T(I-\widetilde{P})^2[Q(\theta-\theta^*)]\\
&-2[Q(\theta-\theta^*)]^T(I-\widetilde{P})^2[A^T(u+A\theta-b)^+-A^Tu^*]\\
&-[A^T(u+A\theta-b)^+-A^Tu^*]^T(I-\widetilde{P})^2[A^T(u+A\theta-b)^+-A^Tu^*]\\
&-(\theta-\theta^*)^T\widetilde{P}^2(\theta-\theta^*)-(\theta-\theta^*)^T[Q(\theta-\theta^*)]\\
&-(\theta-\theta^*)^T[A^T(u+A\theta-b)^+-A^Tu^*]\\
&-\frac{1}{2}\|u-(u+A\theta-b)^+\|^2+[A\theta-b+(-u+b-A\theta)^+]^T[(u+A\theta-b)^+-u^*]\\
&=-\|(I-\widetilde{P})[Q(\theta-\theta^*)]+(I-\widetilde{P})[A^T(u+A\theta-b)^+-A^Tu^*]+\widetilde{P}(\theta-\theta^*)\|^2\\
&-(\theta-\theta^*)^T[Q(\theta-\theta^*)]-(\theta-\theta^*)^T[A^T(u+A\theta-b)^+-A^Tu^*]\\
&-\frac{1}{2}\|u-(u+A\theta-b)^+\|^2+(A\theta-b)^T(u+A\theta-b)^+-(A\theta-b)^Tu^*
\end{aligned}
$$

$$
\begin{aligned}
&+ [(-u+b-A\theta)^+]^T(u+A\theta-b)^+ - [(-u+b-A\theta)^+]^T u^* \\
=& - \|(I-\widetilde{P})[Q\theta+d+A^T(u+A\theta-b)^+] + \widetilde{Q}(e^T\theta)\|^2 \\
&- (\theta-\theta^*)^T[Q(\theta-\theta^*)] - (\theta-\theta^*)^T[A^T(u+A\theta-b)^+ - A^T u^*] \\
&- \frac{1}{2}\|u-(u+A\theta-b)^+\|^2 + (A\theta-b)^T[(u+A\theta-b)^+] - (A\theta-b)^T u^* \\
&+ [(-u+b-A\theta)^+]^T(u+A\theta-b)^+ - [(-u+b-A\theta)^+]^T u^* \\
=& - \|(I-\widetilde{P})[Q\theta+d+A^T(u+A\theta-b)^+] + \widetilde{Q}(e^T\theta)\|^2 \\
&- \frac{1}{2}\|u-(u+A\theta-b)^+\|^2 - (\theta-\theta^*)^T[Q(\theta-\theta^*)] - [(-u+b-A\theta)^+]^T u^* \\
&+ [A(\theta^*-\theta)+A(\theta-\theta^*)]^T u^* - (A\theta^*-b)^T u^* \\
&+ [A(\theta-\theta^*)+A(\theta^*-\theta)]^T(u+A\theta-b)^+ + (A\theta^*-b)^T(u+A\theta-b)^+.
\end{aligned}
$$

به راحتی می‌توان بررسی کرد که

$$
\begin{aligned}
&[(-u+b-A\theta)^+]^T(u+A\theta-b)^+ = 0, \\
&(A\theta^*-b)^T u^* = 0, \\
&[(-u+b-A\theta)^+]^T u^* \geq 0, \\
&(A\theta^*-b)^T[(u+A\theta-b)^+] \leq 0.
\end{aligned}
$$

چون ماتریس Q نیمه معین مثبت است، داریم:

$$(\theta-\theta^*)^T[Q(\theta-\theta^*)] \geq 0.$$

بنابراین

$$\frac{dV}{dt} \leq -\|(I-\widetilde{P})[Q\theta+d+A^T(u+A\theta-b)^+] + \widetilde{Q}(e^T\theta)\|^2 - \frac{1}{2}\|u-(u+A\theta-b)^+\|^2 \leq 0.$$

که این مطلب اثبات را کامل می‌کند. □

ملاحظه ۱.۳.۴. به آسانی می‌توان دید که $V(\theta,u)$ یک تابع لیاپانوف است. با توجه به اثبات لم ۱.۳.۴، داریم $\frac{dV}{dt} \leq 0$. بنابراین شبکه عصبی (۲۴.۴) پایدار به مفهوم لیاپانوف است.

قضیه ۱.۳.۴. به‌ازای هر نقطه اولیه $(\theta(0)^T, u(0)^T)^T \in \mathbb{R}^{8n}$، یک جواب پیوسته $(\theta(t)^T, u(t)^T)^T \in \mathbb{R}^{8n}$ از (۲۴.۴) برای $t \geq 0$ وجود دارد.

برهان. عملگر تصویر $(.)^+$ انقباضی است. واضح است که $(I-\widetilde{P})[Q\theta^*-d+A^T(u^*+A\theta^*-b)^+] + \widetilde{Q}(e^T\theta^*)$ و $(u^*+A\theta^*-b)^+ - u^*$ به‌طور محلی لیپ شیتز پیوسته هستند. بنابراین با توجه به قضیه وجود جواب محلی معادله دیفرانسیل معمولی [۳۰]، جواب یکتای پیوسته‌ای برای (۲۴.۴) در $[0,T)$ وجود دارد.

فرض کنید $[\circ,T)$ بازه ماکسیمال وجود جواب باشد. با توجه به لم ۱.۳.۴، می‌دانیم V نسبت به t تابع ناصعودی است، بنابراین

$$\frac{۱}{۲}\|\theta(t)-\theta^*\|^۲+\frac{۱}{۲}\|u(t)-u^*\|^۲\leq V(\theta(t),u(t))\leq V(\theta(\circ),u(\circ)),\quad \forall t\geq \circ. \qquad (۲۵.۴)$$

این نشان می‌دهد که مسیر (۲۴.۴) کران‌دار است، بنابراین $T=+\infty$ و این اثبات را کامل می‌کند. □

قضیه ۲.۳.۴. به ازای هر نقطه اولیه $(\theta(\circ)^T,u(\circ)^T)^T\in\mathbb{R}^{۸n}$، مسیر شبکه عصبی (۲۴.۴) به یک نقطه تعادل همگرا است. به‌ویژه زمانی که Ω^e نقطه تعادل یکتا داشته باشد، مسیر شبکه به‌ازای هر نقطه اولیه پایدار سراسری مجانبی است.

برهان. تعریف می‌کنیم

$$D(\theta,u)=\|(I-\widetilde{P})[Q\theta+d+A^T(u+A\theta-b)^+]+\widetilde{Q}(e^T\theta)\|^۲+\frac{۱}{۲}\|u-(u+A\theta-b)^+\|^۲.$$

آن‌گاه $D(\theta,u)=\circ$، اگر و فقط اگر $(\theta^T,u^T)^T$ یک نقطه تعادل شبکه عصبی (۲۴.۴) باشد. با توجه به اثبات قضیه (۱.۳.۴) مسیر $(\theta(t)^T,u(t)^T)^T$ از شبکه عصبی (۲۴.۴) کران‌دار است. بنابراین یک دنباله صعودی $\{t_n\}$ با $\lim_{n\to\infty}t_n\longrightarrow\infty$ و یک نقطه حدی به‌صورت $(\hat{\theta}^T,\hat{u}^T)^T$ که $\lim_{n\to\infty}\theta(t_n)\longrightarrow\hat{\theta}$ و $\lim_{n\to\infty}u(t_n)\longrightarrow\hat{u}$ وجود دارد. در ادامه ثابت می‌کنیم که $(\hat{\theta}^T,\hat{u}^T)^T\in\Omega^e$ و مسیر $(\theta(t)^T,u(t)^T)^T$ همگرای سراسری به نقطه تعادل $(\hat{\theta}^T,\hat{u}^T)^T$ است.

ابتدا ثابت می‌کنیم که $D(\hat{\theta},\hat{u})=\circ$، یعنی یک نقطه تعادل شبکه عصبی (۲۴.۴) است. اگر این‌چنین نباشد، یعنی $D(\hat{\theta},\hat{u})>\circ$، چون $D(\theta,u)$ نسبت به θ و u پیوسته است، به‌ترتیب $\circ<\varepsilon<\circ$، $q>\circ$ و یک ε همسایگی از $(\hat{\theta}^T,\hat{u}^T)^T$ وجود دارد به‌طوری که $D(\theta,u)>q$ به‌ازای تمام $(\theta^T,u^T)^T\in B((\hat{\theta},\hat{u}),\varepsilon)$. با توجه به این‌که $\lim_{n\to\infty}\theta(t_n)\longrightarrow\hat{\theta}$ و $\lim_{n\to\infty}u(t_n)\longrightarrow\hat{u}$، عدد صحیح مثبت N وجود دارد، به‌طوری که به ازای تمامی $n\geq N$، $\|\theta(t_n)-\hat{\theta}\|\leq\frac{۱}{۴}\varepsilon$ و $\|u(t_n)-\hat{u}\|\leq\frac{۱}{۴}\varepsilon$.

از شبکه (۲۴.۴) و کران‌داری $(\theta(t)^T,u(t)^T)^T$، مشتقات $\dot{\theta}$ و $\dot{u}$ را داریم که کران‌دار اند و کران بالای آن‌ها را با M با نشان می‌دهیم. در نظر بگیرید $t\in[t_n-\frac{\varepsilon}{۸M},t_n+\frac{\varepsilon}{۸M}]$، $n\geq N$ لذا داریم

$$\begin{aligned}\|\theta(t)-\hat{\theta}\|+\|u(t)-\hat{u}\|\leq&\|\theta(t)-\theta(t_n)\|+\|u(t)-u(t_n)\|+\|\theta(t_n)-\hat{\theta}\|+\|u(t_n)-\hat{u}\|\\=&\|\dot{\theta}(\xi_۱)\|\times|t-t_n|+\|\dot{u}(\xi_۲)\|\times|t-t_n|\\&+\|\theta(t_n)-\hat{\theta}\|+\|u(t_n)-\hat{u}\|\\\leq&۲M|t-t_n|+\frac{\varepsilon}{۲}\leq\varepsilon.\end{aligned}$$

یعنی برای هر $t\in[t_n-\frac{\varepsilon}{۸M},t_n+\frac{\varepsilon}{۸M}]$، $n\geq N$ داریم $(\theta^T,u^T)^T\in B((\hat{\theta},\hat{u}),\varepsilon)$. بنابراین به‌ازای هر $t\in[t_n-\frac{\varepsilon}{۸M},t_n+\frac{\varepsilon}{۸M}]$، $n\geq N$ داریم $D(\theta(t),u(t))>q$.

چون $\lim_{n\to\infty} t_n = \infty$، و اندازه لبگ مجموعه $t \in \cup_{n\geq N}[t_n - \frac{\varepsilon}{\lambda M}, t_n + \frac{\varepsilon}{\lambda M}]$ نامتناهی است بنابراین

$$\int_{\circ}^{\infty} D(\theta(t), u(t))\, dt = \infty. \qquad (۲۶.۴)$$

در حالی‌که

$$\begin{aligned}\int_{\circ}^{\infty} D(\theta(t), u(t))\, dt &= \lim_{s\to\infty} \int_{\circ}^{s} D(\theta(t), u(t))\, dt \\ &\leq -\lim_{s\to\infty} \int_{\circ}^{s} \dot{V}(\theta(t), u(t))\, dt \\ &= \lim_{s\to\infty} [V(\theta(\circ), u(\circ)) - V(\theta(s), u(s))] \\ &\leq V(\theta(\circ), u(\circ)),\end{aligned}$$

که با (۲۶.۴) تناقض دارد. لذا $D(\hat{\theta}, \hat{u}) = \circ$، یعنی $(\hat{\theta}^T, \hat{u}^T)^T \in \Omega^e$.

دوماً ثابت می‌کنیم مسیر $(\theta(t)^T, u(t)^T)^T$ همگرای سراسری به نقطه تعادل $(\hat{\theta}, \hat{u})$ است. برای این منظور تابع لیاپانوف را به‌شکل زیر تعریف می‌کنیم

$$\hat{V}(\theta, u) = \frac{۱}{۲}\theta^T Q\theta + d^T\theta - (\frac{۱}{۲}\hat{\theta}^T Q\hat{\theta} + d^T\hat{\theta}) + \frac{۱}{۲}\|(u + A\theta - b)^+\|^۲ - \frac{۱}{۲}\|(\hat{u} + A\hat{\theta} - b)^+\|^۲$$
$$- (\theta - \hat{\theta})^T(Q\hat{\theta} + d + A^T\hat{u}) - (u - \hat{u})^T\hat{u} + \frac{۱}{۲}\|\theta - \hat{\theta}\|^۲ + \frac{۱}{۲}\|u - \hat{u}\|^۲.$$

داریم: $\hat{V}(\hat{\theta}, \hat{u}) = \circ$. چون $\lim_{n\to\infty} \theta(t_n) \longrightarrow \hat{\theta}$ و $\lim_{n\to\infty} u(t_n) \longrightarrow \hat{u}$، $\forall \varepsilon > \circ$، وجود دارد یک t_k، به‌طوری که $\hat{V}(\theta(t_k), u(t_k)) < \varepsilon$. با توجه به لم ۱.۳.۴، داریم،

$$\frac{۱}{۲}\|\theta(t) - \hat{\theta}\|^۲ + \frac{۱}{۲}\|u(t) - \hat{u}\|^۲ \leq \hat{V}(\theta(t), u(t)),$$

و $\hat{V}$ ناصعودی است. بنابراین به ازای تمامی $t \geq t_k$، داریم

$$\frac{۱}{۲}\|\theta(t) - \hat{\theta}\|^۲ + \frac{۱}{۲}\|u(t) - \hat{u}\|^۲ \leq \hat{V}(\theta(t), u(t)) \leq \hat{V}(\theta(t_k), u(t_k)) \leq \varepsilon.$$

یعنی $\lim_{n\to\infty} \theta(t) = \hat{\theta}$، و $\lim_{n\to\infty} u(t) = \hat{u}$. بنابراین مسیر حالت شبکه عصبی (۲۴.۴) همگرای سراسری به نقطه تعادل $(\hat{\theta}^T, \hat{u}^T)^T$ است. به‌ویژه اگر $\Omega^e = \{(\theta^{*T}, u^{*T})^T\}$، آن‌گاه مسیر حالت $(\theta(t)^T, u(t)^T)^T$ به‌ازای هر نقطه اولیه $(\theta(\circ)^T, u(\circ)^T)^T$ با توجه به تجزیه و تحلیل بالا به نقطه تعادل $(\theta^{*T}, u^{*T})^T$ میل می‌کند. لذا شبکه عصبی پایدار مجانبی سراسری است. □

۱.۳.۴ مقایسه با برخی شبکه‌های عصبی موجود

در این بخش برای این‌که ببینیم شبکه عصبی (۲۴.۴) چگونه مساله (۱۶.۴) و (۱۷.۴) را حل می‌کند، این شبکه را با برخی از شبکه‌های عصبی موجود مقایسه می‌کنیم.

نوعی از شبکه عصبی به‌نام شبکه عصبی مدل گرادیان وجود دارد که برای استفاده از این شبکه یک مساله بهینه سازی مقید توسط یک مساله بهینه سازی نامقید تخمین زده می‌شود سپس تابع انرژی با استفاده از روش تابع جریمه ساخته می‌شود. مزیت مدل شبکه عصبی گرادیان این است که می‌توان مدل شبکه عصبی را به‌طور مستقیم توسط گرادیان تابع انرژی تعریف کرد ولی نقص آن در تضمین نشدن همگرایی به ویژه در مورد مجموعه‌های جواب بی‌کران است [۶۱]. علاوه بر این در شبکه عصبی گرادیان بر اساس تابع جریمه، برای هر پارامتر قابل تنظیم نیاز به پارامتر جریمه دارد. به‌عنوان نمونه با استفاده از روش جریمه، مساله بهینه‌سازی مقید (۱۶.۴) و (۱۷.۴) را می‌توان با مساله بهینه‌سازی نامقید زیر تخمین زد.

$$\text{minimize} \quad E_1(x) = \frac{1}{2}\theta^T Q\theta + d^T\theta + \frac{\gamma}{2}\left\{\sum_{k=1}^{4n}[(A\theta-b)_k^+]^2 + (e^T\theta)^2\right\},$$

که در آن γ پارامتر جریمه است. مدل شبکه عصبی ارائه شده برای مساله فوق به‌نام مدل شبکه عصبی کندی و چا[1] به‌شکل زیر است [۲۲]

$$\frac{dx}{dt} = -\nabla E_1(x) = -\left(Q\theta + d + \gamma[A^T(A\theta-b)^+ + e^T(e\theta)]\right). \qquad (۲۷.۴)$$

شبکه (۲۷.۴) به‌ازای مقادیر متناهی پارامتر جریمه قادر به پیدا کردن جواب دقیق مساله بهینه‌سازی نیست و به‌ازای مقادیر بسیار زیاد برای پارامتر جریمه، پیاده‌سازی این شبکه دشوار است [۲۴]. بنابراین این شبکه به‌ازای هر مقدار متناهی از پارامتر جریمه به تقریبی از جواب بهینه مساله (۱۶.۴) و (۱۷.۴) همگرا است. هم‌چنین می‌توان نشان داد که شبکه عصبی (۲۷.۴) در مساله بهینه‌سازی محدب، به یک جواب بهینه دقیق همگرای سراسری نیست. به‌عنوان نمونه مثال ۱.۴ را در [۳۴] ببینید.

ناظمی و همکار در [۳۶] شبکه عصبی گرادیانی به‌شکل زیر برای حل مساله (۱۶.۴) و (۱۷.۴) ارائه دادند.

$$\frac{d(y(t))}{dt} = -\nabla E_2(y(t)), \qquad (۲۸.۴)$$

$$y(0) = y_0, \quad y(t) = (x(t), u(t), v(t))^T, \qquad (۲۹.۴)$$

که در آن

$$E_2(y) = \frac{1}{2}\|\rho(y)\|^2,$$

$$\rho(y) = \begin{bmatrix} Q\theta + d + e^T v + A^T u \\ -e^T\theta \\ \phi_{\text{FB}}^{\varepsilon}(u, b - A\theta) \end{bmatrix} = 0,$$

[1] Kennedy and Chua's

و

$$\phi^{\varepsilon}_{\mathrm{FB}}(a,b) = (a+b) - \sqrt{a^{۲} + b^{۲} + \varepsilon}, \quad \varepsilon \to ۰_{+}.$$

همچنین ناظمی در [۳۵] مدل شبکه عصبی گرادیانی به شکل زیر برای حل مساله (۱۶.۴) و (۱۷.۴) بیان کرده است،

$$\frac{dy}{dt} = \Phi(y), \tag{۳۰.۴}$$

$$y(t_{۰}) = y_{۰}, \quad y(t) = (x(t), u(t), v(t))^{T}, \quad u(t_{۰}) > ۰, \tag{۳۱.۴}$$

که در آن

$$\Phi(y) = \begin{bmatrix} -\left(Q\theta + d + \frac{۱}{۲}A^{T}u^{۲} + e^{T}v\right) \\ \mathrm{diag}(u_{۱}, ..., u_{p})(A\theta - b) \\ e^{T}\theta \end{bmatrix}.$$

در مدل‌های (۲۹.۴)ـ(۲۸.۴) و (۳۰.۴)ـ(۳۱.۴) متغیرهای حالت $x \in \mathbb{R}^{۴n}$، $u \in \mathbb{R}^{۴n}$ و $v \in \mathbb{R}$ است. بنابراین این شبکه‌های عصبی دارای $۸n+۱$ نرون است. در حالی تعداد نرون‌های شبکه عصبی (۲۴.۴) $۸n$ است، که زمانی کارایی خود را نشان می‌دهد که تعداد قیود تساوی مساله بهینه‌سازی زیاد باشد. مهم‌تر این‌که مدل‌های (۲۹.۴)ـ(۲۸.۴) و (۳۰.۴)ـ(۳۱.۴) در حالتی که به‌ازای تمامی $x \in \mathbb{R}^{۴n}$ و $u_k \geq ۰$ ماتریس Q معین مثبت است، همگراست. لذا این مدل قادر به حل مسائل برنامه‌ریزی خطی که کاربرد آن‌را محدود می‌کند، نیست. در حالی‌که شبکه عصبی (۲۴.۴) وقتی که ماتریس Q به‌ازای تمامی $x \in \mathbb{R}^{۴n}$ و $u_k \geq ۰$ نیمه معین مثبت است، به‌جواب بهینه همگراست.

شبکه عصبی ارائه شده در این فصل دارای $۸n$ نرون است در حالی‌که شبکه عصبی که در بخش قبل با آن مساله رگرسیون بردار پشتیبان را حل کردیم، دارای $۸n+۱$ نرون بود. البته قابل ذکر است که در شبکه عصبی (۲۰.۳) و (۲۱.۳) ما نیازی به محاسبه ماتریس معکوس نداشتیم ولی در شبکه عصبی (۲۴.۴) بایستی ماتریس معکوس را محاسبه کرد.

۴.۴ مثال‌های عددی

برای نشان‌دادن کارایی شبکه عصبی مورد نظر، مثال‌هایی را در این بخش ارائه می‌دهیم. شبیه سازی این مثال ها با استفاده از نرم افزار $Matlab\ ۷$ و حل کننده معادلات دیفرانسیل معمولی $ode۴۵$ انجام شده است .

برای مشخص کردن مقدار مناسب برای پارامترهای C، ϵ، a و δ از روش $Leave-One-Out$ در **اعتبارسنجی‌متقابل** (CV) استفاده می‌کنیم.

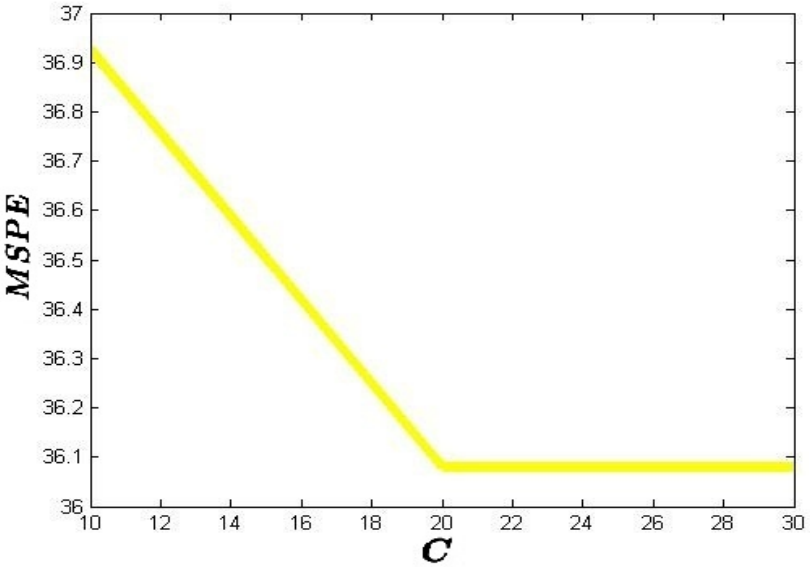
مثال ۱.۴.۴. داده‌های رگرسیون جدول (۱.۴) را در نظر می‌گیریم. می‌دانیم اگر $X_i \sim u(a,b)$ آن‌گاه $E(X_i) = \frac{a+b}{۲}$.

در ادامه با تغییر یک پارامتر و ثابت نگه‌داشتن سایر پارامترها، رفتار $f(x)$ را بررسی می‌کنیم.

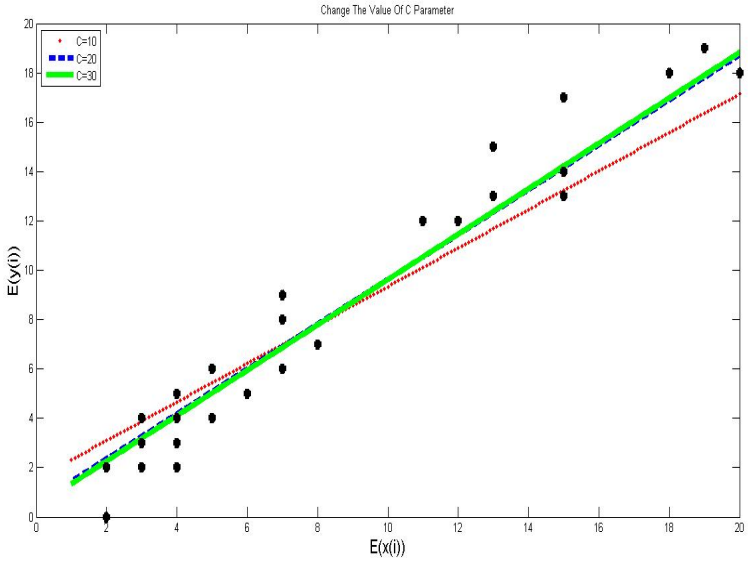
جدول ۱.۴: توزیع احتمال X_i و Y_i

i	X_i	Y_i	i	X_i	Y_i
۱	$u(۱۴,۱۶)$	$u(۱۲,۱۴)$	۱۶	$u(۱۳,۱۷)$	$u(۱۶,۱۸)$
۲	$u(۱۰,۱۲)$	$u(۱۱,۱۳)$	۱۷	$u(۱۱,۱۳)$	$u(۱۰,۱۴)$
۳	$u(۲,۴)$	$u(۲,۴)$	۱۸	$u(۲,۶)$	$u(۲/۵,۳/۵)$
۴	$u(۱۲,۱۴)$	$u(۱۴,۱۶)$	۱۹	$u(۳,۷)$	$u(۲,۶)$
۵	$u(۱,۵)$	$u(۱,۵)$	۲۰	$u(۱/۵,۲/۵)$	$u(۱۷/۷۵,۲/۲۵)$
۶	$u(۲/۵,۳/۵)$	$u(۳,۵)$	۲۱	$u(۱۸,۲۰)$	$u(۱۸,۲۰)$
۷	$u(۱,۳)$	$u(۰,۰)$	۲۲	$u(۱۴/۵,۱۵/۵)$	$u(۱۳,۱۵)$
۸	$u(۱/۵,۴/۵)$	$u(۱,۳)$	۲۳	$u(۱۰,۱۴)$	$u(۱۲/۵,۱۳/۵)$
۹	$u(۳/۵,۴/۵)$	$u(۱/۵,۲/۵)$	۲۴	$u(۶/۵,۷/۵)$	$u(۷,۹)$
۱۰	$u(۳,۵)$	$u(۴,۶)$	۲۵	$u(۲/۵,۵/۵)$	$u(۳/۵,۴/۵)$
۱۱	$u(۶,۸)$	$u(۸,۱۰)$	۲۶	$u(۱۱,۱۵)$	$u(۱۲/۵,۱۳/۵)$
۱۲	$u(۵,۹)$	$u(۵,۷)$	۲۷	$u(۱۹,۲۱)$	$u(۱۶,۲۰)$
۱۳	$u(۴,۶)$	$u(۴,۸)$	۲۸	$u(۳/۵,۴/۵)$	$u(۱/۲۵,۲/۷۵)$
۱۴	$u(۵,۷)$	$u(۳,۷)$	۲۹	$u(۴,۸)$	$u(۳,۷)$
۱۵	$u(۱۷,۱۹)$	$u(۱۷,۱۹)$	۳۰	$u(۷,۹)$	$u(۶,۸)$

در شکل (۲.۴) مقدار $MSPE$ به‌ازای مقادیر مختلف C نشان داده شده است. شکل (۳.۴) نیز $f(x)$ را به‌ازای سه مقدار مختلف C نشان می‌دهد. با توجه به نمودار شکل (۲.۴)، به‌ازای مقدار $C = ۳۰$ دقت برآورد $f(x)$ که در شکل (۳.۴) نشان داده شده است، از دو مقدار دیگر C بیشتر است.

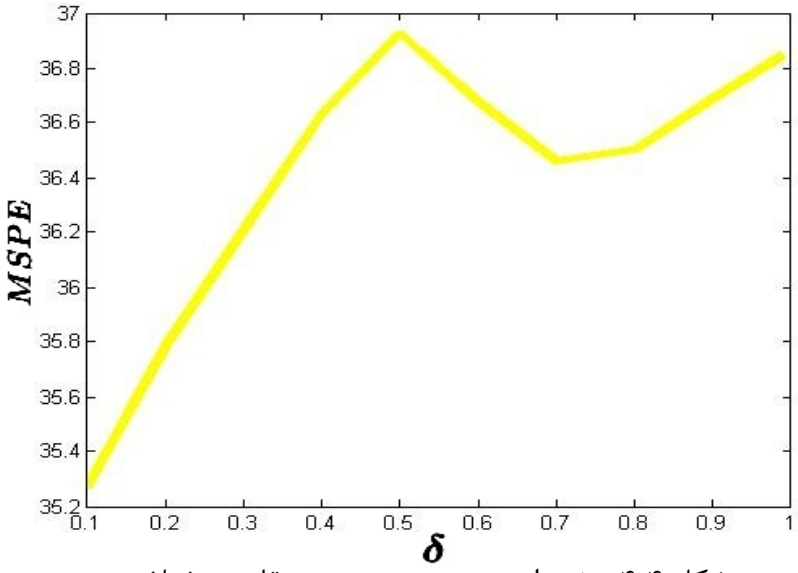


شکل ۲.۴: تغییرات $MSPE$ نسبت به مقادیر مختلف C

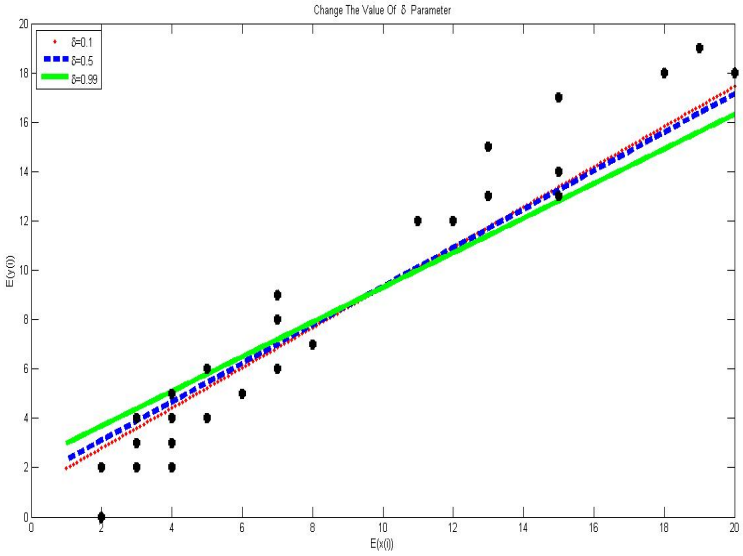


شکل ۳.۴: نتایج رگرسیون بردار پشتیبان به‌ازای سه مقدار مختلف C با استفاده از شبکه عصبی (۲۴.۴)

در شکل (۴.۴) مقدار $MSPE$ به‌ازای مقادیر مختلف δ نشان داده شده است. شکل (۵.۴) نیز $f(x)$ را به‌ازای سه مقدار مختلف δ نشان می‌دهد. با توجه به نمودار شکل (۴.۴)، به‌ازای مقدار $\delta = ۰/۱$ دقت برآورد $f(x)$ که در شکل (۵.۴) نشان داده شده است، از دو مقدار دیگر δ بیشتر است.

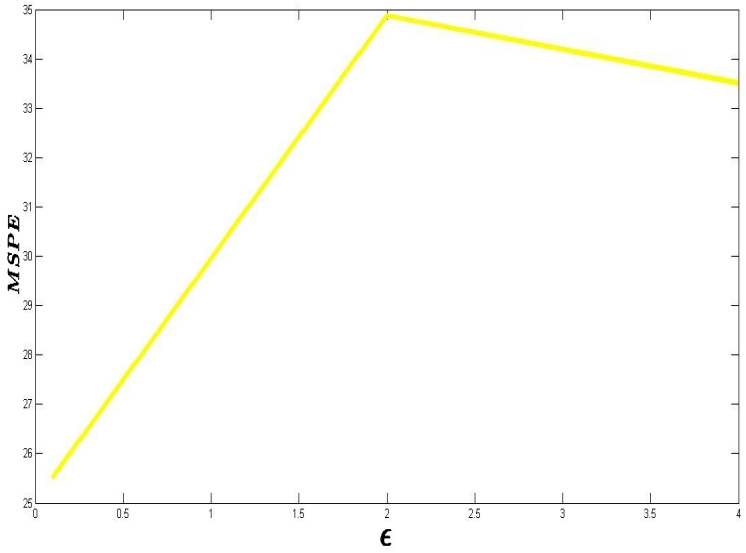


شکل ۴.۴: تغییرات $MSPE$ نسبت به مقادیر مختلف δ

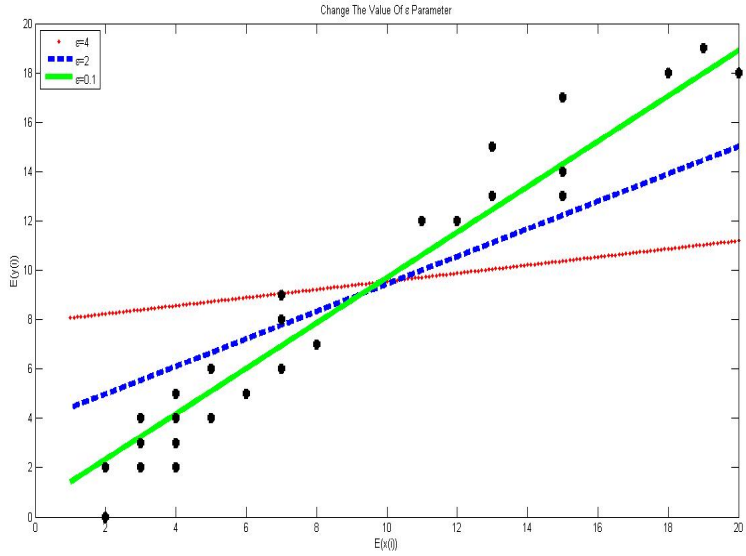


شکل ۵.۴: نتایج رگرسیون بردار پشتیبان به‌ازای سه مقدار مختلف δ با استفاده از شبکه عصبی (۲۴.۴)

در شکل (۶.۴) مقدار $MSPE$ به‌ازای مقادیر مختلف ϵ نشان داده شده است. شکل (۷.۴) نیز $f(x)$ را به‌ازای سه مقدار مختلف ϵ نشان می‌دهد. با توجه به نمودار شکل (۶.۴)، به‌ازای مقدار $\epsilon = ۰/۱$ دقت برآورد $f(x)$ که در شکل (۷.۴) نشان داده شده است، از دو مقدار دیگر ϵ بیشتر است.

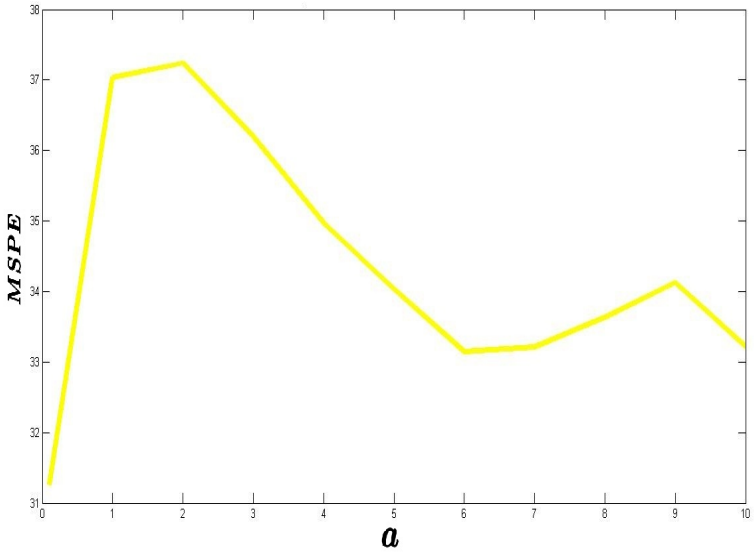


شکل ۶.۴: تغییرات $MSPE$ نسبت به مقادیر مختلف ϵ

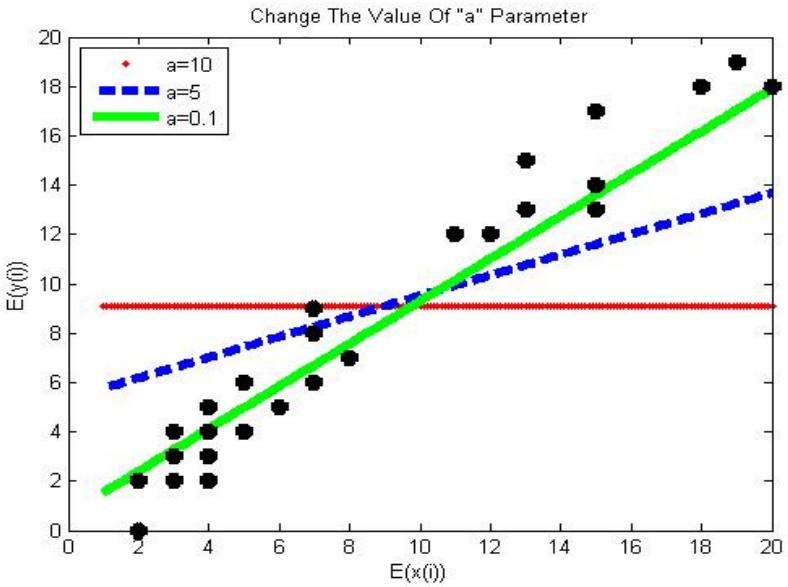


شکل ۷.۴: نتایج رگرسیون بردار پشتیبان به‌ازای سه مقدار مختلف ϵ با استفاده از شبکه عصبی (۲۴.۴)

در شکل (۸.۴) مقدار $MSPE$ به‌ازای مقادیر مختلف a نشان داده شده است. شکل (۹.۴) نیز $f(x)$ را به‌ازای سه مقدار مختلف a نشان می‌دهد. با توجه به نمودار شکل (۸.۴)، به‌ازای مقدار $a = ۰/۱$ دقت برآورد $f(x)$ که در شکل (۹.۴) نشان داده شده است، از دو مقدار دیگر a بیشتر است.



شکل ۸.۴: تغییرات $MSPE$ نسبت به مقادیر مختلف a



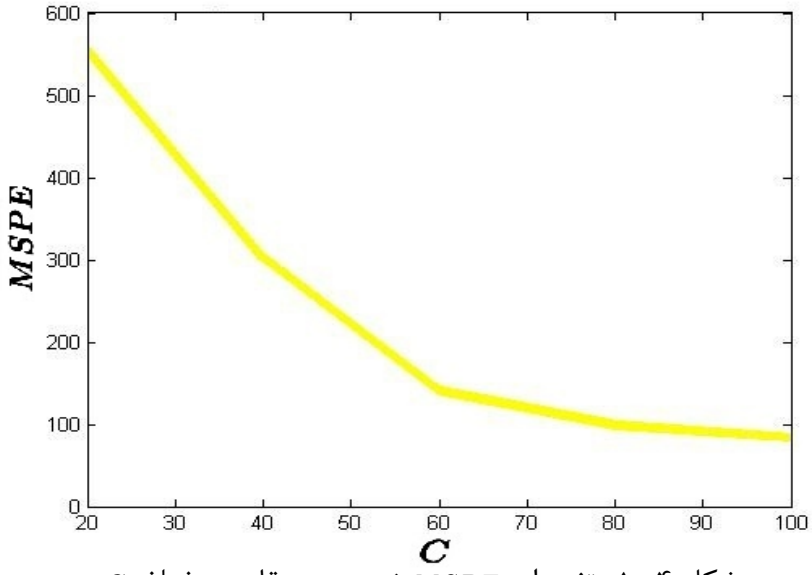
شکل ۹.۴: نتایج رگرسیون بردار پشتیبان به‌ازای سه مقدار مختلف a با استفاده از شبکه عصبی (۲۴.۴)

مثال ۲.۴.۴. نقاط رگرسیون جدول (۲.۴) را که دارای توزیع یکنواخت هستند را در نظر می‌گیریم. با توجه به غیرخطی بودن این مساله رگرسیون از کرنل تابع پایه شعاعی ($\Phi(x)^T\Phi(y) = K(x,y) = \exp(-\frac{\|x-y\|^۲}{۲\delta^۲})$) برای رگرسیون استفاده می‌کنیم.
در ادامه با تغییر یک پارامتر و ثابت نگه‌داشتن سایر پارامترها، رفتار $f(x)$ را بررسی می‌کنیم.

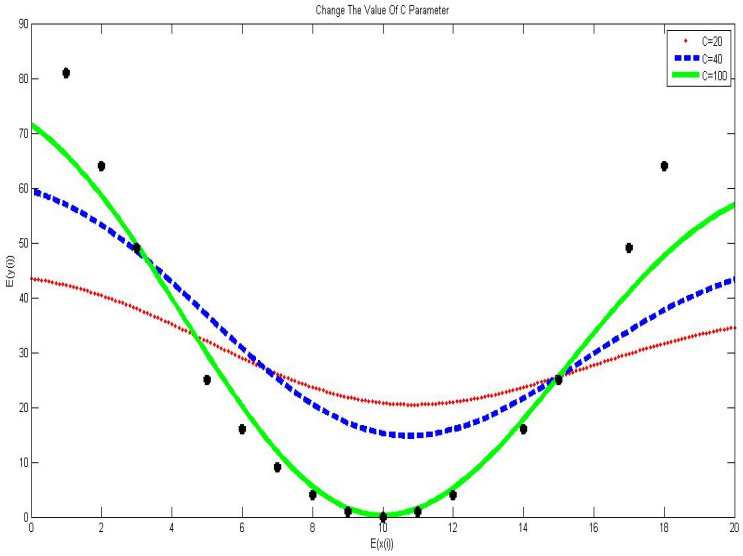
جدول ۲.۴: توزیع احتمال X_i و Y_i

i	X_i	Y_i	i	X_i	Y_i
۱	$u(۱۴,۱۶)$	$u(۲۴,۲۶)$	۱۶	$u(۸,۱۰)$	$u(۰,۲)$
۲	$u(۱۳,۱۷)$	$u(۲۳,۲۷)$	۱۷	$u(۱,۵)$	$u(۴۷,۵۱)$
۳	$u(۱۷,۱۹)$	$u(۶۳,۶۵)$	۱۸	$u(۷,۱۱)$	$u(۰,۲)$
۴	$u(۱,۳)$	$u(۶۲,۶۶)$	۱۹	$u(۱۶,۲۰)$	$u(۵۸,۷۰)$
۵	$u(۶,۸)$	$u(۸,۱۰)$	۲۰	$u(۶,۱۰)$	$u(۱,۷)$
۶	$u(۰,۲)$	$u(۸۰,۸۲)$	۲۱	$u(۲,۴)$	$u(۴۶,۵۲)$
۷	$u(۱۶,۱۸)$	$u(۴۸,۵۰)$	۲۲	$u(۱۱,۱۳)$	$u(۲,۶)$
۸	$u(۱۰,۱۲)$	$u(۰,۲)$	۲۳	$u(۱۵,۲۱)$	$u(۵۷,۷۱)$
۹	$u(۷,۹)$	$u(۷,۱)$	۲۴	$u(۵,۹)$	$u(۷,۱۱)$
۱۰	$u(۴,۶)$	$u(۲۲,۲۸)$	۲۵	$u(۰,۲)$	$u(۷۹,۸۳)$
۱۱	$u(۹,۱۳)$	$u(۰,۲)$	۲۶	$u(۰,۶)$	$u(۴۵,۵۳)$
۱۲	$u(۴,۸)$	$u(۱۵,۱۷)$	۲۷	$u(۱,۳)$	$u(۵۶,۶۲)$
۱۳	$u(۰,۴)$	$u(۶۱,۶۷)$	۲۸	$u(۱۲,۱۸)$	$u(۲۱,۲۹)$
۱۴	$u(۱,۳)$	$u(۶۰,۶۸)$	۲۹	$u(۹,۱۱)$	$u(۰,۰)$
۱۵	$u(۰,۴)$	$u(۵۹,۶۹)$	۳۰	$u(۱۳,۱۵)$	$u(۱۴,۱۸)$

در شکل (۱۰.۴) مقدار $MSPE$ به‌ازای مقادیر مختلف C نشان داده شده است. شکل (۱۱.۴) نیز $f(x)$ را به‌ازای سه مقدار مختلف C نشان می‌دهد. با توجه به نمودار شکل (۱۰.۴)، به‌ازای مقدار $C = ۱۰۰$ دقت برآورد $f(x)$ که در شکل (۱۱.۴) نشان داده شده است، از دو مقدار دیگر C بیشتر است.

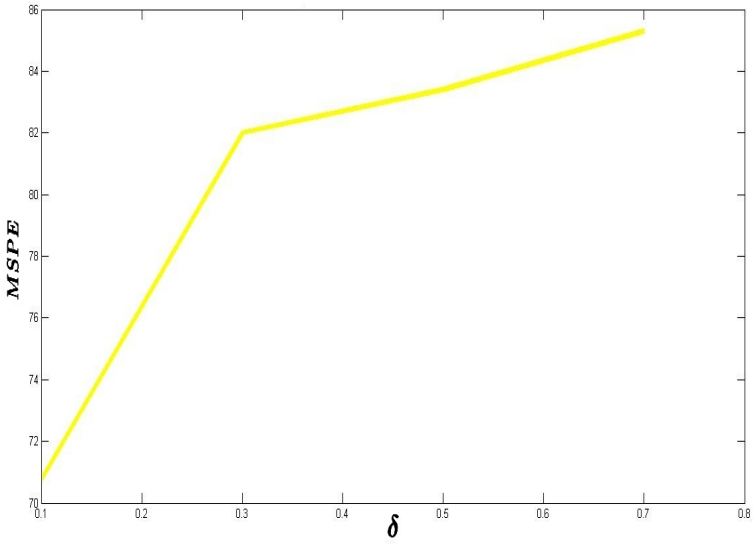


شکل ۱۰.۴: تغییرات $MSPE$ نسبت به مقادیر مختلف C

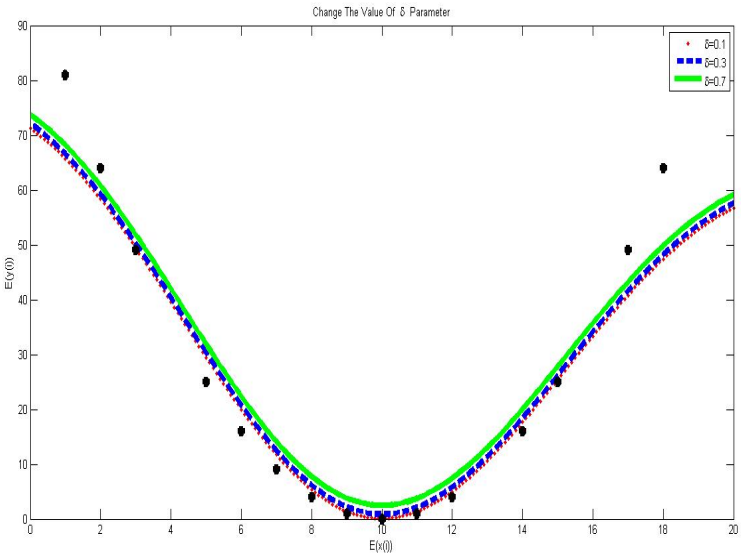


شکل ۱۱.۴: نتایج رگرسیون بردار پشتیبان به‌ازای سه مقدار مختلف C با استفاده از شبکه عصبی (۲۴.۴)

در شکل (۱۲.۴) مقدار $MSPE$ به‌ازای مقادیر مختلف δ نشان داده شده است. شکل (۱۳.۴) نیز $f(x)$ را به‌ازای سه مقدار مختلف δ نشان می‌دهد. با توجه به نمودار شکل (۱۲.۴)، به‌ازای مقدار $\delta = ۰/۱$ دقت برآورد $f(x)$ که در شکل (۱۳.۴) نشان داده شده است، از دو مقدار دیگر δ بیشتر است.

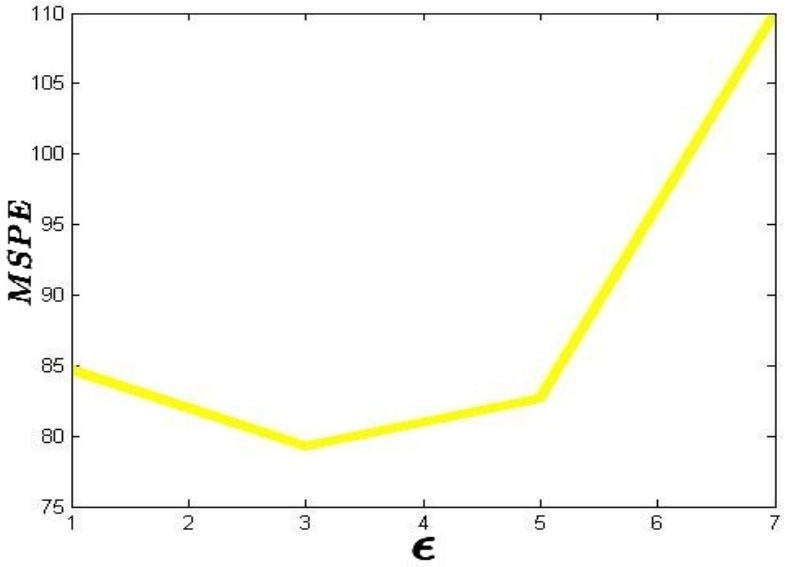


شکل ۱۲.۴: تغییرات $MSPE$ نسبت به مقادیر مختلف δ

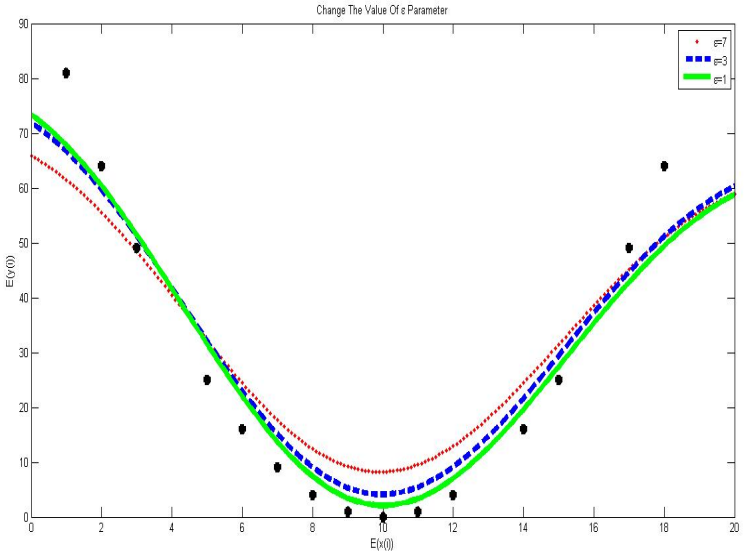


شکل ۱۳.۴: نتایج رگرسیون بردار پشتیبان به‌ازای سه مقدار مختلف δ با استفاده از شبکه عصبی (۲۴.۴)

در شکل (۱۴.۴) مقدار $MSPE$ به‌ازای مقادیر مختلف ϵ نشان داده شده است. شکل (۱۵.۴) نیز $f(x)$ را به‌ازای سه مقدار مختلف ϵ نشان می‌دهد. با توجه به نمودار شکل (۱۴.۴)، به‌ازای مقدار $\epsilon = ۳$ دقت $f(x)$ که در شکل (۱۵.۴) نشان داده شده است، از دو مقدار دیگر ϵ بیشتر است.

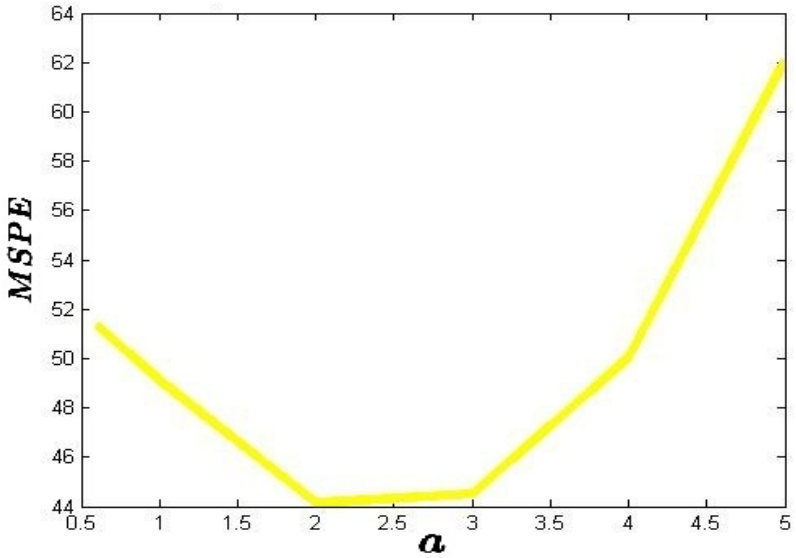


شکل ۱۴.۴: تغییرات $MSPE$ نسبت به مقادیر مختلف ϵ

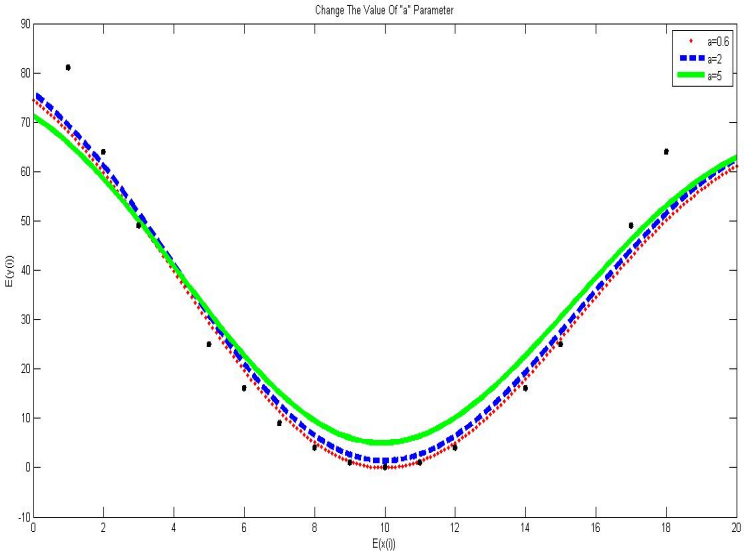


شکل ۱۵.۴: نتایج رگرسیون بردار پشتیبان به‌ازای سه مقدار مختلف ϵ با استفاده از شبکه عصبی (۲۴.۴)

در شکل (۱۶.۴) مقدار $MSPE$ به‌ازای مقادیر مختلف a نشان داده شده است. شکل (۱۷.۴) نیز $f(x)$ را به‌ازای سه مقدار مختلف a نشان می‌دهد. با توجه به نمودار شکل (۱۶.۴)، به‌ازای مقدار $a = ۲$ دقت برآورد $f(x)$ که در شکل (۱۷.۴) نشان داده شده است، از دو مقدار دیگر a بیشتر است.



شکل ۱۶.۴: تغییرات $MSPE$ نسبت به مقادیر مختلف a

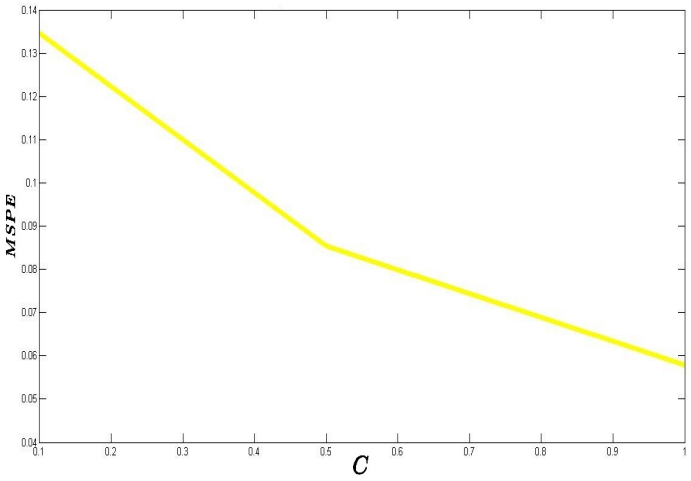


شکل ۱۷.۴: نتایج رگرسیون بردار پشتیبان به‌ازای سه مقدار مختلف a با استفاده از شبکه عصبی (۲۴.۴)

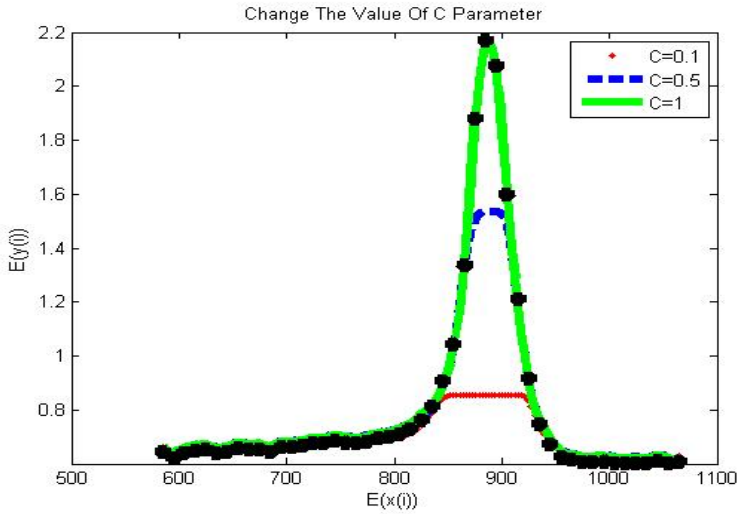
مثال ۳.۴.۴. در این مثال نقاط تیتانیوم مثال (۳.۴.۳) را طوری در نظر می‌گیریم که X_iها و Y_iها امید ریاضی نقاط تصادفی باشند.

در ادامه همانند مثال‌های قبل با تغییر یک پارامتر و ثابت نگه‌داشتن سایر پارامترها، رفتار $f(x)$ را بررسی می‌کنیم.

در شکل (۱۸.۴) مقدار $MSPE$ به‌ازای مقادیر مختلف C نشان داده شده است. شکل (۱۹.۴) نیز $f(x)$ را به‌ازای سه مقدار مختلف C نشان می‌دهد. با توجه به نمودار شکل (۱۸.۴)، به‌ازای مقدار $C = ۱$ دقت برآورد $f(x)$ که در شکل (۱۹.۴) نشان داده شده است، از دو مقدار دیگر C بیشتر است.

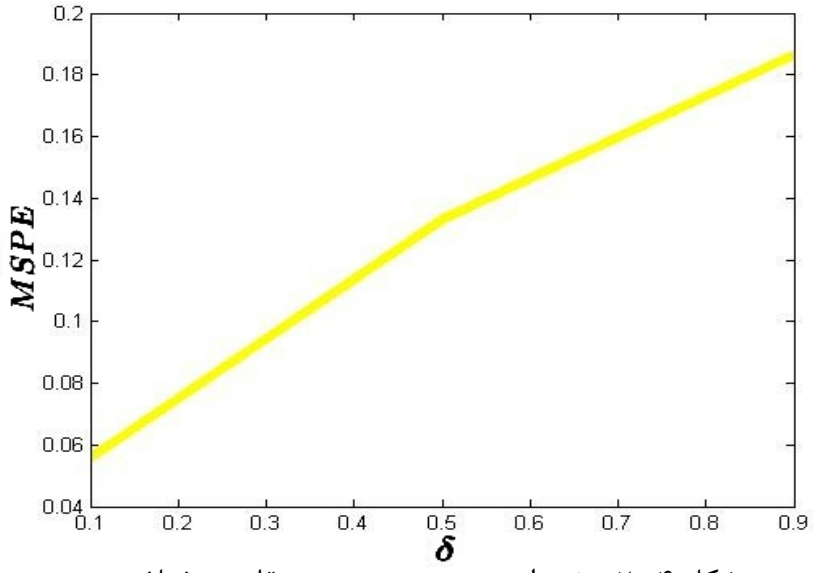


شکل ۱۸.۴: تغییرات $MSPE$ نسبت به مقادیر مختلف C

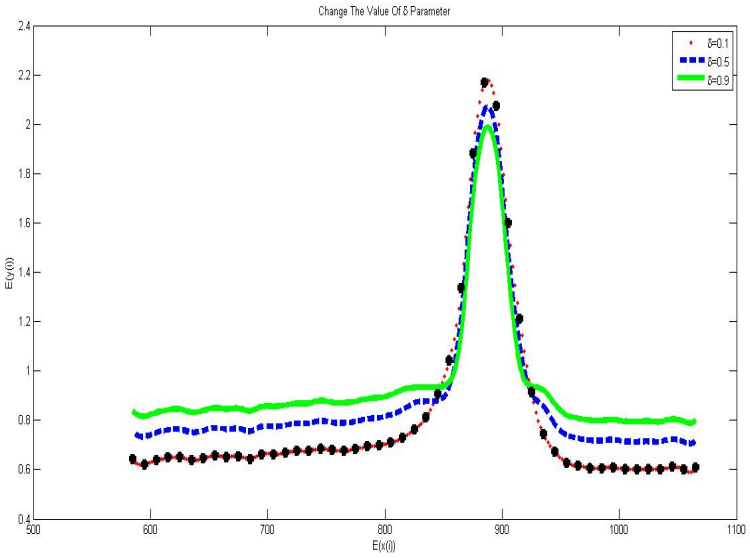


شکل ۱۹.۴: نتایج رگرسیون بردار پشتیبان به‌ازای سه مقدار مختلف C با استفاده از شبکه عصبی (۲۴.۴)

در شکل (۲۰.۴) مقدار $MSPE$ به‌ازای مقادیر مختلف δ نشان داده شده است. شکل (۲۱.۴) نیز $f(x)$ را به‌ازای سه مقدار مختلف δ نشان می‌دهد. با توجه به نمودار شکل (۲۰.۴)، به‌ازای مقدار $\delta = ۰/۱$ دقت برآورد $f(x)$ که در شکل (۲۱.۴) نشان داده شده است، از دو مقدار دیگر δ بیشتر است.

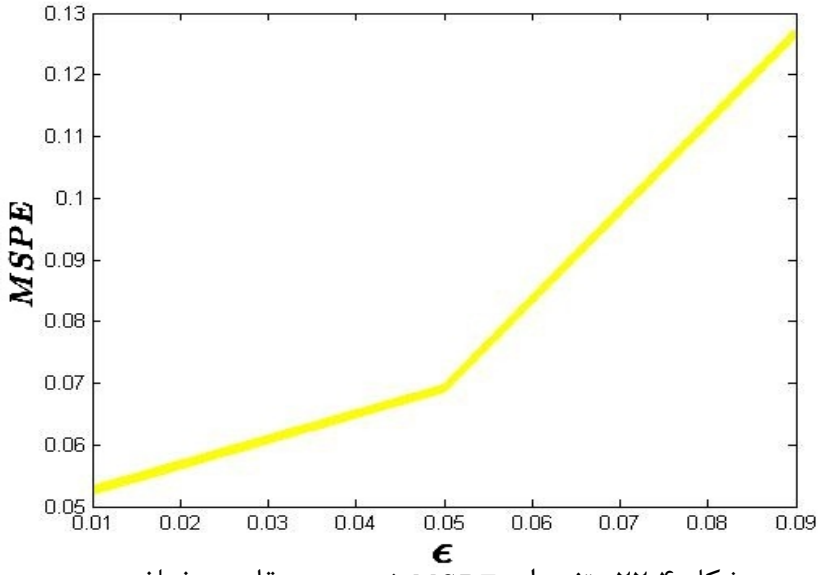


شکل ۲۰.۴: تغییرات $MSPE$ نسبت به مقادیر مختلف δ

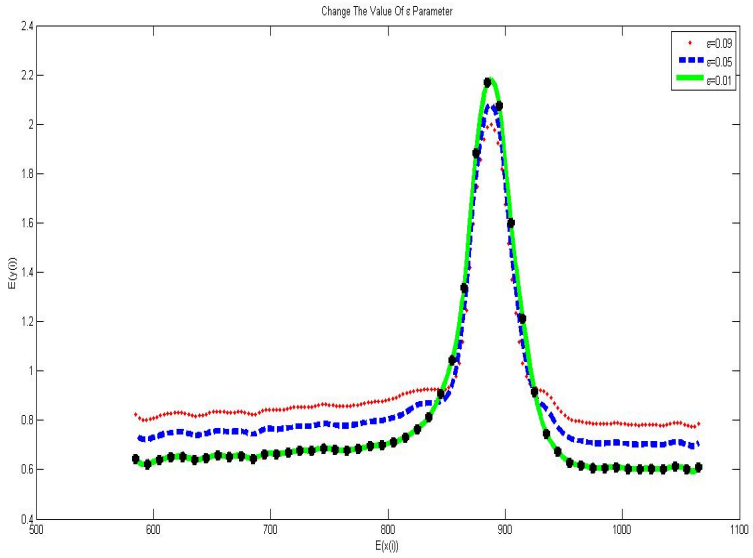


شکل ۲۱.۴: نتایج رگرسیون بردار پشتیبان به‌ازای سه مقدار مختلف δ با استفاده از شبکه عصبی (۲۴.۴)

در شکل (۲۲.۴) مقدار $MSPE$ به‌ازای مقادیر مختلف ϵ نشان داده شده است. شکل (۲۳.۴) نیز $f(x)$ را به‌ازای سه مقدار مختلف ϵ نشان می‌دهد. با توجه به نمودار شکل (۲۲.۴)، به‌ازای مقدار $\epsilon = ۰/۰۱$ دقت برآورد $f(x)$ که در شکل (۲۳.۴) نشان داده شده است، از دو مقدار دیگر ϵ بیشتر است.

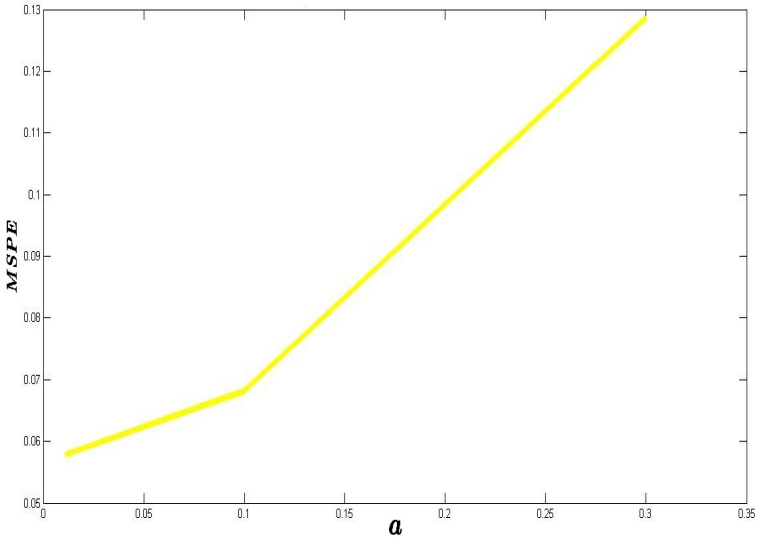


شکل ۲۲.۴: تغییرات $MSPE$ نسبت به مقادیر مختلف ϵ

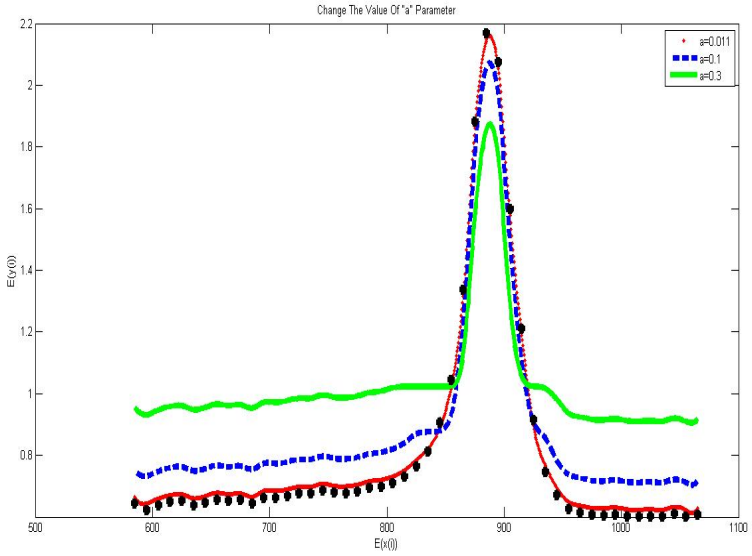


شکل ۲۳.۴: نتایج رگرسیون بردار پشتیبان به‌ازای سه مقدار مختلف ϵ با استفاده از شبکه عصبی (۲۴.۴)

در شکل (۲۴.۴) مقدار $MSPE$ به‌ازای مقادیر مختلف a نشان داده شده است. شکل (۲۵.۴) نیز $f(x)$ را به‌ازای سه مقدار مختلف a نشان می‌دهد. با توجه به نمودار شکل (۲۴.۴)، به‌ازای مقدار $a = ۰/۰۱۱$ دقت برآورد $f(x)$ که در شکل (۲۵.۴) نشان داده شده است، از دو مقدار دیگر a بیشتر است.



شکل ۲۴.۴: تغییرات $MSPE$ نسبت به مقادیر مختلف a



شکل ۲۵.۴: نتایج رگرسیون بردار پشتیبان به‌ازای سه مقدار مختلف a با استفاده از شبکه عصبی (۲۴.۴)

نتایج و پیشنهادات

با استفاده از دانش برنامه‌نویسی رایانه می‌توان ساختار داده‌ای طراحی کرد که همانند یک نرون عمل نماید. سپس با ایجاد شبکه‌ای از این نرون‌های مصنوعی به هم پیوسته، ایجاد یک الگوریتم آموزشی برای شبکه و اعمال این الگوریتم به شبکه آن را آموزش داد. این شبکه‌ها کارایی بسیار بالایی از خود نشان داده‌اند. گستره کاربرد این مدل‌های ریاضی بر گرفته از عملکرد مغز انسان، بسیار وسیع می‌باشد که به عنوان چند نمونه کوچک می‌توان استفاده از این ابزار ریاضی در پردازش سیگنال‌های بیولوژیکی، مخابراتی و الکترونیکی تا کمک در نجوم و فضانوردی را نام برد. ما در این رساله یک مدل شبکه عصبی برای حل مساله رگرسیون بردار پشتیبان بر اساس شرایط بهینگی KKT ارائه دادیم. شبکه عصبی پیشنهاد شده ساختاری ساده دارد. علاوه بر آن نشان دادیم که شبکه عصبی ارائه شده پایدار لیاپانوف و همگرای سراسری به جواب بهینه مساله رگرسیون بردار پشتیبان است. همچنین کارایی مدل شبکه عصبی ارائه شده با حل مثال‌های متنوع نشان داده شد. همچنین یک مدل شبکه عصبی اصلاح شده برای حل مساله رگرسیون بردار پشتیبان با قیدهای احتمالی ارائه دادیم. نشان دادیم که مدل شبکه عصبی ارائه شده پایدار مجانبی و همگرای سراسری به جواب بهینه مساله رگرسیون بردار پشتیبان با قیدهای احتمالی است. با ارائه مثال‌هایی کارایی مدل شبکه عصبی پیشنهاد شده تایید شد.

با توجه به این که تاکنون مدل شبکه عصبی برای حل مساله رگرسیون بردار پشتیبان با قیدهای احتمالی ارائه نشده است، در این رساله برای اولین بار یک مدل شبکه عصبی مبتنی بر شرایط KKT برای حل این مسئله بیان شد. با توجه به کاربردهای بسیار گسترده رگرسیون بردار پشتیبان با قیدهای احتمالی نتایج بدست آمده در این رساله می‌تواند مورد توجه قرار گیرد.

به عنوان پیشنهادی برای کارهای آتی می‌توان به ارائه مدل‌هایی از شبکه عصبی برای مسائلی چون مسائل کلاس‌بندی که در آن نقاط، متغیرهایی تصادفی هستند و یا مسائل رگرسیون بردار پشتیبان با متغیرهای بازه‌ای که در آن به‌جای یک نقطه با بازه‌ای از نقاط سرو کار داریم و یا مسائل رگرسیون بردار پشتیبان با نقاط خروجی تصادفی اشاره کرد.

مراجع

[۱] ارجمندزاده، آمنه. ماشین بردار پشتیبان و رده بندی داده های بازه ای. *دانشگاه تربیت معلم سبزوار، دانشکده علوم پایه، پایان‌نامه کارشناسی‌ارشد* ، ۱۳۸۸.

[۲] کبودیان، جهانشاه، رحمتی، محمد، و همایون پور، محمد مهدی. استفاده از ماشین بردار پشتیبان (svm) در سه مساله شناسایی الگو. *اولین کنفرانس بین المللی فناوری اطلاعات و دانش، تهران* ، ۱۳۸۲.

[۳] ساده، محمدجواد. رده‌بندی در داده‌های ریزآرایه‌ها. *پایان‌نامه کارشناسی ارشد، دانشگاه فردوسی مشهد* ، ۱۳۸۴.

[۴] ضمیری، آذر. شبکه‌های عصبی مصنوعی. *پایان‌نامه کارشناسی ارشد، دانشگاه فردوسی مشهد* ، ۱۳۸۷.

[۵] عباس، قماشی لنگرودی. حل مسائل بهینه‌سازی با استفاده از شبکه‌های عصبی تصویر. *دانشگاه تربیت معلم سبزوار، دانشکده علوم پایه، پایان‌نامه کارشناسی‌ارشد* ، ۱۳۸۳.

[6] Hebb, D.O. the organization of behavior: A neuropsychological theory. new york: John wiley and sons, inc., 1949. *Science Education*, 34:336–337, 1950.

[7] Widrow, B. generalization and information storage in networks of adaline neurons. in self organizing systems. *Spartan Books*, 1959.

[8] Abaszade, M. and Effati, S. Stochastic support vector regression with probabilistic constraints. *Applied Intelligence*, 48:243–256, 2018.

[9] Mordecai, A. *Nonlinear programming : analysis and methods / Mordecai Avriel*. Englewood Cliffs (N.J.) : Prentice-Hall, 1976.

[10] Barnett, N.S., Dragomir, S.S. and Agarwal, R.P. Some inequalities for probability, expectation, and variance of random variables defined over a finite interval. *Computers and Mathematics with Applications*, 43:1319–1357, 2002.

[11] Bazaraa, M.S., Sherali, H.D., and Shetty, C.M. Nonlinear programming, theory and algorithms (2nd edn), 2013.

[12] Bouzerdoum, A. and Pattison, T.R. Neural network for quadratic optimization with bound constraints. *IEEE Transactions on Neural Networks*, 4:293–304, 1993.

[13] Dierckx, P. Curve and surface fitting with splines, chapter univariate splines, 1993.

[14] Effati, S. and Baymani, M. A new nonlinear neural network for solving convex nonlinear programming problems. *Applied Mathematics and Computation*, 168:1370 – 1379, 2005.

[15] Effati, S., Ghomashi, A. and Nazemi, A.R. Application of projection neural network in solving convex programming problems. *Applied Mathematics and Computation*, 188:1103 – 1114, 2007.

[16] Effati, S. and Nazemi, A.R. Neural network models and its application for solving linear and quadratic programming problems. *Applied Mathematics and Computation*, 172:305 – 331, 2006.

[17] Effati, S., Ghomashi, A. and Abbasi, M. A novel recurrent neural network for solving mlcps and its application to linear and quadratic programming. *Asia-Pacific Journal of Operational Research*, 28:523–541, 2011.

[18] Forti, M., Nistri, P., and Quincampoix, M. Convergence of neural networks for programming problems via a nonsmooth lojasiewicz inequality. *IEEE Transactions on Neural Networks*, 17:1471–1486, 2006.

[19] Friesz, T.L., Bernstein, D., Mehta, N.J., Tobin, R.L. and Ganjalizadeh,S. Day-to-day dynamic network disequilibria and idealized traveler information systems. *Operations Research*, 42:1120–1136, 1994.

[20] Hopfield, J.J. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79:2554–2558, 1982.

[21] Hopfield, J.J. and Tank, D.W. "neural" computation of decisions in optimization problems. *Biological cybernetics*, 52:141–152, 1985.

[22] Kennedy, M.P. and Chua, L.O. Neural networks for nonlinear programming. *IEEE Transactions on Circuits and Systems*, 35:554–562, 1988.

[23] Khalil, H.K. and Grizzle, J.W. *Nonlinear systems*, volume 3. Prentice hall Upper Saddle River, NJ, 2002.

[24] Lillo, W.E., Loh, M.H., Hui, S. and Zak, S.H. On solving constrained optimization problems with neural networks: a penalty method approach. *IEEE Transactions on Neural Networks*, 4(6):931–940, 1993.

[25] Luenberger, D. and Ye, Y. *Linear and Nonlinear Programming*. Springer US, 2008.

[26] Maa, C.Y. and Shanblatt, M.A. Linear and quadratic programming neural network analysis. *IEEE transactions on neural networks*, 3:580–594, 1992.

[27] Maa, C.Y. and Schanblatt, M.A. A two-phase optimization neural network. *IEEE transactions on neural networks*, 3:1003—1009, 1992.

[28] Maugeri, A. and Zichichi, A. *Variational analysis and applications*. New York, Springer, 2005.

[29] Warren, S.M. and Walter, P. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5:115–133, 1943.

[30] Miller, R.K. and Michel, A.N. Ordinary differential equations. academic press, new york et. al. 1982.

[31] Minsky, M.L. and Perceptrons, S.P. An introduction to computational geometry. *Mit Press*, 1972.

[32] Moghaddas, M. and Eshaghnejad, M. A recurrent neural network for solving non-convex nonlinear optimization problem. *The 44th Annual Iranian Mathematics Conference, Mashhad, Iran,*, 2013.

[33] Moghaddas, M. and Effati, S. A novel recurrent neural network based on ncp function for solving convex nonlinear optimization problems. *The 5th Iranian Conference on Applied Mathematics, Hamadan, Iran,*, 2013.

[34] Nazemi, A.R. Solving general convex nonlinear optimization problems by an efficient neurodynamic model. *Engineering Applications of Artificial Intelligence*, 26:685–696, 2013.

[35] Nazemi, A.R. A neural network model for solving convex quadratic programming problems with some applications. *Engineering Applications of Artificial Intelligence*, 32:54–62, 2014.

[36] Nazemi, A.R. and Nazemi, M. A gradient-based neural network method for solving strictly convex quadratic programming problems. *Cognitive Computation*, 6:484–495, 2014.

[37] Nazemi, A.R. and Omidi, F. A capable neural network model for solving the maximum flow problem. *Journal of Computational and Applied Mathematics*, 236:3498 – 3513, 2012.

[38] Nazemi, A.R. A dynamic system model for solving convex nonlinear optimization problems. *Communications in Nonlinear Science and Numerical Simulation*, 17:1696 – 1705, 2012.

[39] Perko, L. *Differential equations and dynamical systems*. Springer, 3nd ed, 2001.

[40] Quarteroni, A., Sacco, R. and Saleri, F. *Numerical mathematics*, volume 37. Springer-Verlag, Berlin, second edition, 2007.

[41] Rodriguez-Vazquez, A., Castro, R., Rueda, A., Huertas, J.L. and Sanchez-Sinencio, E. Nonlinear switched-capacitor 'neural network' for optimisation problems. *Circuits and Systems, IEEE Transactions on*, 37:384 – 398, 1990.

[42] Rosenblatt, F. The perception: a probabilistic model for information storage and organization in the brain, neurocomputing: foundations of research, 1988.

[43] Rosenblatt, F. Principles of neurodynamics. perceptrons and the theory of brain mechanisms. Technical report, Cornell Aeronautical Lab Inc Buffalo NY, 1961.

[44] Rumelhart, D.E., Hinton, G.E. and Williams, R.J. Learning representations by back-propagating errors. *nature*, 323:533–536, 1986.

[45] Schölkopf, B., Smola, A.J. and et al. Learning with kernels: support vector machines, regularization. *Optimization, and Beyond. MIT press*, 1(2), 2002.

[46] Tank, D.W. and Hopfield, J.J. Simple 'neural' optimization networks: An a/d converter, signal decision circuit, and a linear programming circuit. *IEEE Transactions on Circuits and Systems*, CAS-33:533–541, 1986.

[47] Tao, Q., Cao, J. and Sun, D. A simple and high performance neural network for quadratic programming problems. *Applied Mathematics and Computation*, 124:251–260, 2001.

[48] Tax, D. One-class classification: Concept learning in the absence of counter-examples. 2002.

[49] Ortega, T.M. and Rheinboldt, W.C. *Iterative solution of nonlinear equations in several variables*. Academic Press, New York, 1970.

[50] Vapnik, V. Statistical learning theory wiley. *New York*, 1:624, 1998.

[51] Werbos, P.J. Beyond regression: New tools for prediction and analysis in the behavioral sciences. ph. d. thesis, harvard university, cambridge, ma, 1974. 1974.

[52] Widrow, B. and Hoff, M.E. Adaptive switching circuits (pp. 96–104). In *Institute of Radio Engineers (IRE) WESCON Convention record*, 1960.

[53] Widrow, B. and Lehr, M.A. 30 years of adaptive neural networks: perceptron, madaline, and backpropagation. *Proceedings of the IEEE*, 78:1415–1442, 1990.

[54] Widrow, B. and Winter, R. Neural nets for adaptive filtering and adaptive pattern recognition. *Computer*, 21:25–39, 1988.

[55] Xia, Y. A new neural network for solving linear and quadratic programming problems. *IEEE transactions on neural networks*, 7:1544–1548, 1996.

[56] Xia, Y., Feng, G., and Wang, J. A recurrent neural network with exponential convergence for solving convex quadratic program and related linear piecewise equations. *Neural Networks*, 17:1003–1015, 2004.

[57] Xia, Y., Feng, G., and Wang, J. A novel recurrent neural network for solving nonlinear optimization problems with inequality constraints. *IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council*, 19:1340–53, 2008.

[58] Xia, Y. and Wang, J. A recurrent neural network for solving linear projection equations. *Neural Networks*, 13:337–350, 2000.

[59] Xu, H. Projection neural networks for solving constrained convex and degenerate quadratic problems. *2010 IEEE International Conference on Intelligent Computing and Intelligent Systems*, 3:91–96, 2010.

[60] Xue, X. and Bian, W. A project neural network for solving degenerate convex quadratic program. *Neurocomputing*, 70:2449–2459, 2007.

[61] Yang, Y. and Cao, J. A feedback neural network for solving convex constraint optimization problems. *Applied Mathematics and Computation*, 201(1):340 – 350, 2008.

[62] Xia, Y. and Wang, J. A general methodology for designing globally convergent optimization neural networks. *IEEE Transactions on Neural Networks*, 9:1331–1343, 1998.

[63] Xia, Y. and Wang, J. A general projection neural network for solving monotone variational inequalities and related optimization problems. *IEEE Transactions on Neural Networks*, 15:318–328, 2004.

[64] Xia, Y. and Wang, J. A recurrent neural network for nonlinear convex optimization subject to nonlinear inequality constraints. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 51:1385–1394, 2004.

[65] Xia, Y. and Wang, J. A recurrent neural network for solving nonlinear convex programs subject to linear constraints. *IEEE Transactions on Neural Networks*, 16:379–386, 2005.

[66] Xia, Y., Leung, H. and Wang, J. A projection neural network and its application to constrained optimization problems. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 49:447–458, 2002.

[67] Zabczyk, J. Mathematical control theory. *An Introduction*, 1992.

[68] Zhang, G., Shen, Y., Yin, Q. and Sun, J. Global exponential periodicity and stability of a class of memristor-based recurrent neural networks with multiple delays. *Inf. Sci.*, 232:386–396, 2013.

واژه‌نامه فارسی به انگلیسی

الگوریتم پس انتشار ..... Back propagation algorithm
اعتبار سنجی متقابل ..... Cross validation
ϵ ـ حساسیت ..... ϵ - insensitive
ابرصفحه بهینه جدا کننده ..... Optimum separating hyperplane
آدلاین ..... Adeline
اصل تغییرناپذیری لِسال ..... LaSalle principle of invariance
آکسون ..... Axon
اکیداً یکنوا ..... Strictly monotone
آموزش ..... Learning
انقباضی ..... Nonexpansive
به‌طور شعاعی بی‌کران ..... Radially unbounded
بهینه‌سازی محدب ..... Convex optimization
بهینه‌سازی نامحدب ..... Nonconvex optimization
پاسخ سرپرست ..... Supervisor response
پایانه بعد از سیناپسی ..... Postsynaptic terminal
پایانه قبل از سیناپسی ..... Presynaptic terminal
پایدار مجانبی سراسری ..... Globally asymptotically stable
پایداری به مفوم لیاپانوف ..... Stability in the sense of Lyapunov
پایداری نمایی سراسری ..... Globally exponential stability
پرسپترون ..... Perceptron
پیوسته لیپ‌شیتز ..... Lipschitz continuous
تابع پایه شعاعی ..... Radial basis function
تابع فعال‌سازی ..... Activation function
تابع مخاطره ..... Risk function
تحریک شدن ..... Firing
تعمیم تصویر ..... Generalization portrait

تولید کننده ..... Generator
جواب شدنی ..... Feasible solution
حاشیه ..... Margin
دسته‌بندی دوتایی ..... Binary classification
دندریت ..... Dendrite
دوگانی موضعی ..... Local duality
رگرسیون ϵـ بردار پشتیبان ..... ϵ - SV Regression
سرپرست ..... Supervisor
سیستم دینامیکی ..... Dynamic system
سیناپس ..... Synapse
شبکه عصبی بازگشتی ..... Feedback (Recurrent) neural network
شبکه عصبی پیش‌خورد ..... Feed forward neural network
شبکه عصبی سوئیچ‌ـخازن ..... Switched-Capacitor neural network
شبکه عصبی مصنوعی ..... Artificial neural network
شرایط کافی مرتبه دوم ..... Second-order sufficient conditions
شکاف سیناپسی ..... Synaptic cleft
شناسایی الگو ..... Pattern recognition
ضرب‌گر لاگرانژ ..... Lagrangian multiplier
عملگر تصویر ..... Projection operator
عنصر خطی تطبیقی ..... Adaptive linear element
فضای ویژگی ..... Feature space
قضیه نقطه ثابت بروئر ..... Brouwer's fixed point theorem
قویاً یکنوا ..... Strongly monotone
ماتریس هسین ..... Hessian matrix
مادلاین ..... Madaline
ماشین بردار پشتیبان ..... Support vector machine
متغیر دوگان ..... Dual variable
مجموعه تغییر ناپذیر ..... Invariant set
مجموعه سطح ..... Level set
مساله دوگان لاگرانژ ..... Lagrangian dual problem
مسأله مکمل خطی ..... Linear complementarity problem (LCP)
مسأله مکمل غیرخطی ..... Nonlinear complementarity problem (NCP)
میانگین توان دوم خطای پیش‌بینی ..... Mean squared prediction error
معین مثبت قوی ..... Strongly positive definite

معین مثبت ..... Positive definite
نامساوی وردشی ..... Variational inequality
نرون ..... Neuron
نقطه تعادل تنها ..... Isolated equilibrium point
نقطه تعادل ..... Equilibrium point
نقطه منظم ..... Regular point
نقطه مینیمم محلی اکید ..... Strict local minimum point
نیمه معین مثبت ..... Positive semidefinite
هسته ..... Kernel
همگرای سراسری ..... Globally convergent
یکنوا ..... Monotone

واژه‌نامه انگلیسی به فارسی

شبکه عصبی مصنوعی ..... Artificial neural network
دسته‌بندی دوتایی ..... Binary classification
ϵ ـ حساسیت ..... ϵ - insensitive
رگرسیون ϵـ بردار پشتیبان ..... ϵ - SV Regression
تابع فعال‌سازی ..... Activation function
عنصر خطی تطبیقی ..... Adaptive linear element
آدلاین ..... Adeline
آکسون ..... Axon
الگوریتم پس انتشار ..... Back propagation algorithm
قضیه نقطه ثابت بروئر ..... Brouwer's fixed point theorem
اعتبار سنجی متقابل ..... Cross validation
بهینه‌سازی محدب ..... Convex optimization
دندریت ..... Dendrite
متغیر دوگان ..... Dual variable
سیستم دینامیکی ..... Dynamic system
نقطه تعادل ..... Equilibrium point
جواب شدنی ..... Feasible solution
فضای ویژگی ..... Feature space
شبکه عصبی پیش‌خورد ..... Feed forward neural network
شبکه عصبی بازگشتی ..... Feedback (Recurrent) neural network
تحریک شدن ..... Firing
تعمیم تصویر ..... Generalization portrait
تولید کننده ..... Generator
پایدار مجانبی سراسری ..... Globally asymptotically stable
همگرای سراسری ..... Globally convergent
پایداری نمایی سراسری ..... Globally exponential stability

ماتریس هسین ..... Hessian matrix
مجموعه تغییر ناپذیر ..... Invariant set
نقطه تعادل تنها ..... Isolated equilibrium point
هسته ..... Kernel
مساله دوگان لاگرانژ ..... Lagrangian dual problem
ضرب‌گر لاگرانژ ..... Lagrangian multiplier
اصل تغییرناپذیری لِسال ..... LaSalle principle of invariance
آموزش ..... Learning
مجموعه سطح ..... Level set
مسأله مکمل خطی ..... Linear complementarity problem (LCP)
پیوسته لیپ‌شیتز ..... Lipschitz continuous
دوگانی موضعی ..... Local duality
مادلاین ..... Madaline
حاشیه ..... Margin
میانگین توان دوم خطای پیش‌بینی ..... Mean squared prediction error
یکنوا ..... Monotone
نرون ..... Neuron
بهینه‌سازی نامحدب ..... Nonconvex optimization
انقباضی ..... Nonexpansive
مسأله مکمل غیرخطی ..... Nonlinear complementarity problem (NCP)
ابرصفحه بهینه جدا کننده ..... Optimum separating hyperplane
شناسایی الگو ..... Pattern recognition
پرسپترون ..... Perceptron
معین مثبت ..... Positive definite
نیمه معین مثبت ..... Positive semidefinite
پایانه بعد از سیناپسی ..... Postsynaptic terminal
پایانه قبل از سیناپسی ..... Presynaptic terminal
عملگر تصویر ..... Projection operator
تابع پایه شعاعی ..... Radial basis function
به‌طور شعاعی بی‌کران ..... Radially unbounded
نقطه منظم ..... Regular point
تابع مخاطره ..... Risk function
شرایط کافی مرتبه دوم ..... Second-order sufficient conditions
پایداری به مفوم لیاپانوف ..... Stability in the sense of Lyapunov

نقطه مینیمم محلی اکید ..... Strict local minimum point
اکیداً یکنوا ..... Strictly monotone
قویاً یکنوا ..... Strongly monotone
معین مثبت قوی ..... Strongly positive definite
پاسخ سرپرست ..... Supervisor response
سرپرست ..... Supervisor
ماشین بردار پشتیبان ..... Support vector machine
شبکه عصبی سوئیچ‌ـخازن ..... Switched-Capacitor neural network
سیناپس ..... Synapse
شکاف سیناپسی ..... Synaptic cleft
نامساوی وردشی ..... Variational inequality

پیوست

۱. مقدمه‌ای بر بهینه‌سازی غیرخطی

در پیوست رساله مطالبی را که خواننده احیاناً برای مطالعه بهتر رساله نیاز به‌آن پیدا می‌کند، آورده شده است.

تعریف ۱.۱. [۲۵] تابع $f : \mathbb{R}^n \to \mathbb{R}$ روی مجموعه محدب و ناتهی $S \subseteq \mathbb{R}^n$ محدب نامیده می‌شود اگر برای هر $x, y \in S$ و $\alpha \in (0, 1)$ داشته باشیم

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha) f(y).$$

تابع f روی S به طور اکید محدب نامیده می‌شود اگر نامساوی بالا زمانی‌که $x \neq y$ است، به‌طور اکید برقرار باشد.

یک مسأله برنامه‌ریزی غیرخطی در حالت کلی به‌صورت زیر در نظر گرفته می‌شود

$$\text{minimize} \quad f(x), \tag{۳۲}$$

$$\text{subject to} \quad g_i(x) \leq 0, \tag{۳۳}$$

$$h_j(x) = 0, \tag{۳۴}$$

$$x \in X \subset \mathbb{R}^n, \tag{۳۵}$$

که $f : \mathbb{R}^n \to \mathbb{R}$، $g_i : \mathbb{R}^n \to \mathbb{R}$ $(i = 1, ..., m)$ و $h_j : \mathbb{R}^n \to \mathbb{R}$ $(j = 1, ..., p)$ توابع دو بار مشتق پذیر پیوسته و X زیر مجموعه‌ای باز از $\mathbb{R}^n$ است. برای تسهیل در نمادگذاری، توابع برداری $\boldsymbol{g} = (g_1, g_2, ..., g_m)^T$ و $\boldsymbol{h} = (h_1, h_2, ..., h_p)^T$ را معرفی می‌کنیم و (۳۲)ـ(۳۵) را به‌صورت زیر می‌نویسیم

$$\text{minimize} \quad f(x), \tag{۳۶}$$

$$\text{subject to} \quad g(x) \leq 0, \tag{۳۷}$$

$$h(x) = 0, \tag{۳۸}$$

$$x \in X \subset \mathbb{R}^n, \tag{۳۹}$$

که محدودیت‌های $\boldsymbol{g}(x) \leq ۰$ و $\boldsymbol{h}(x) = ۰$، محدودیت‌های تابعی و محدودیت $x \in X \subset \mathbb{R}^n$، محدودیت مجموعه‌ای نامیده می‌شود. هرگاه توابع برداری f و $\boldsymbol{g}$ روی X محدب باشند و $\boldsymbol{h}$ آفین باشد یعنی $\boldsymbol{h} = Ax - b$، $A \in \mathbb{R}^{p \times n}$، $rank(A) = p$ $(۰ \leq p < n)$ و $b \in \mathbb{R}^p$ باشد، آن‌گاه (۳۶)ـ(۳۹) یک مسأله بهینه‌سازی محدب و در غیر این‌صورت یک مسأله بهینه‌سازی نامحدب نامیده می‌شود.

تعریف ۲.۱. [۲۵] نقطه x جواب شدنی مسأله (۳۶)ـ(۳۹) نامیده می‌شود هرگاه در محدودیت‌های مسأله صدق کند. نقطه شدنی x یک نقطه منظم نامیده می‌شود اگر بردارهای گرادیان $\nabla g_i(x)$، $i \in \{i | g_i(x) = ۰\}$ و $\nabla h_j(x)$، $j = ۱, ..., p$ مستقل خطی باشند.

قضیه ۱.۱. (شرایط کافی مرتبه دوم) [۲۵] فرض کنید x^* یک نقطه شدنی برای مسأله (۳۶)ـ(۳۹) باشد. اگر $\mu \in \mathbb{R}^m$ و $\lambda \in \mathbb{R}^p$ وجود داشته باشند به‌طوری که

$$\mu \geq ۰,$$
$$\mu^T \boldsymbol{g}(x^*) = ۰,$$
$$\nabla f(x^*) + \lambda^T \nabla \boldsymbol{h}(x^*) + \mu^T \nabla \boldsymbol{g}(x^*) = ۰,$$

و ماتریس هسین

$$\nabla^۲ L(x^*, \lambda, \mu) = \nabla^۲ f(x^*) + \sum_{i=۱}^{m} \lambda_i \nabla^۲ g_i(x^*) + \sum_{j=۱}^{p} \mu_j \nabla^۲ h_j(x^*),$$

بر زیرفضای

$$M(x^*) = \{d | \nabla \boldsymbol{h}(x^*) d = ۰, \nabla g_j(x^*) d = ۰, \forall j \in J(x^*)\},$$

که در آن

$$J(x^*) = \{j | g_j(x^*) = ۰, \mu_j > ۰\},$$

معین مثبت باشد، آن‌گاه x^* یک نقطه مینیمم محلی اکید برای مسأله(۳۶)ـ(۳۹) است.

مسأله دوگان لاگرانژ

مسأله برنامه‌ریزی غیرخطی (۳۶)ـ(۳۹) را در نظر بگیرید، این مسأله را مسأله اولیه (P) نیز می‌نامیم. مسأله دوگان لاگرانژ(D) به‌صورت زیر تعریف می‌شود

$$\max \quad \theta(u, v), \tag{۴۰}$$
$$\text{subject to} \quad u \geq ۰, \tag{۴۱}$$

که در آن

$$\theta(u,v)=\inf_x\left\{f(x)+\sum_{i=۱}^{m}u_ig_i(x)+\sum_{j=۱}^{p}v_jh_j(x),\ x\in X\right\},$$
$$u=(u_۱,u_۲,...,u_m)^T,\quad v=(v_۱,v_۲,...,v_p)^T,$$

u_i، iــامین مؤلفه از u، متغیر دوگان یا ضرب‌گر لاگرانژ مربوط به قید $g_i(x)\le ۰$ می‌باشد و v_j، jــامین مؤلفه v، متغیر دوگان یا ضرب‌گر لاگرانژ مربوط به قید $h_j(x)=۰$ می‌باشد.

مسأله دوگان لاگرانژ (۴۰) و (۴۱) در شکل برداری به‌صورت زیر نوشته می‌شود

$$\text{maximize}\ \ \theta(u,v), \tag{۴۲}$$
$$\text{subject to}\ \ u\ge ۰, \tag{۴۳}$$

که در آن

$$\theta(u,v)=\inf_x\left\{f(x)+u^t\boldsymbol{g}(x)+v^t\boldsymbol{h}(x),\ x\in X\right\},$$

هم چنین $f:\mathbb{R}^n\to\mathbb{R}$ ، $\boldsymbol{g}=(g_۱,...,g_m)^T:\mathbb{R}^n\to\mathbb{R}^m$ و $\boldsymbol{h}=(h_۱,...,h_p):\mathbb{R}^n\to\mathbb{R}^p$ هستند.

قضیه ۲.۱. (دوگانی موضعی) [۲۵] فرض کنید مسأله

$$\begin{aligned}&\text{minimize}\ \ f(x),\\ &\text{subject to}\ \ h(x)=۰,\\ &x\in X\subset\mathbb{R}^n,\end{aligned}$$

که $\boldsymbol{h}(x)\in\mathbb{R}^m$ و $f,\boldsymbol{h}\in C^۲$، دارای یک جواب موضعی در x^* با مقدار متناظر r^* و ضریب لاگرانژ v^* است. هم‌چنین فرض کنید که x^* یک نقطه منظم برای قیود است و هسین لاگرانژ متناظر $\nabla^۲L(x^*,v^*)$ معین مثبت است. در این‌صورت مسأله دوگان

$$\max\ \theta(v),$$

که در آن

$$\theta(v)=\inf_x\left\{f(x)+v^t\boldsymbol{h}(x),\ x\in X\right\},$$

دارای یک جواب موضعی در v^* با مقدار متناظر r^* است و x^* جواب متناظر با v^* در تعریف θ است.

به‌سادگی می‌توان نتایج فوق را به مسائلی که علاوه بر قیود تساوی قیود نامساوی هم دارند تعمیم داد [۲۵].

۲. سیستم‌های دینامیکی

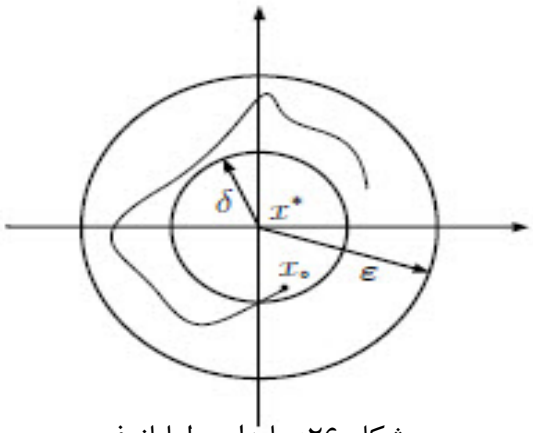
تعریف ۱.۲. [۳۹] سیستم دینامیکی زیر را در نظر بگیرید:

$$\dot{x} = f(x(t)),\ x(t_\circ) = x_\circ \in \mathbb{R}^n, \qquad (۴۴)$$

که در آن f یک تابع برداری از $\mathbb{R}^n$ به $\mathbb{R}^n$ می‌باشد. x^* یک نقطه تعادل (۴۴) نامیده می‌شود اگر $f(x^*) = \circ$. اگر همسایگی $\Omega^* \subseteq \mathbb{R}^n$ از x^* وجود داشته باشد که $f(x^*) = \circ$ و $\forall x \in \Omega^* \backslash \{x^*\}$، $f(x) \neq \circ$، آن‌گاه x^* یک نقطه تعادل تنها نامیده می‌شود.

تعریف ۲.۲. (پایداری به مفهوم لیاپانوف) [۳۹] فرض می‌کنیم که $x(t)$ یک جواب (۴۴) باشد، نقطه تعادل تنها x^*، پایدار به مفهوم لیاپانوف است اگر برای هر $x_\circ = x(t_\circ)$ و هر $\varepsilon > \circ$، یک $\delta > \circ$ وجود داشته باشد به‌طوری که اگر $\|x(t_\circ) - x^*\| < \delta$ آن‌گاه

$$\|x(t) - x^*\| < \varepsilon,\ \forall t \geq t_\circ.$$



شکل ۲۶: پایداری لیاپانوف.

تعریف ۳.۲. [۳۹] سیستم دینامیکی (۴۴) همگرای سراسری به مجموعه جواب‌های (۴۴) که با Ω^* نمایش داده می‌شود، گفته می‌شود اگر برای هر نقطه آغازین دلخواه، مسیر $x(t)$ در رابطه زیر صدق کند:

$$\lim_{t\to\infty} dist(x(t), \Omega^*) = \circ,$$

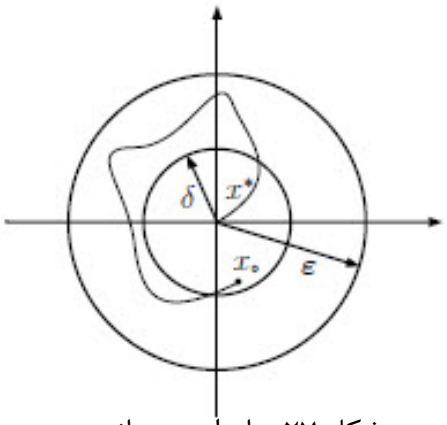
که در آن

$$dist\,(x, \Omega^*) = \inf_{y\in\Omega^*} \|x - y\|\,,$$

و $\|.\|$ نرم $l_۲$ را نشان می‌دهد.

تعریف ۴.۲. [۳۹] سیستم دینامیکی (۴۴) در نقطه تعادل x^* که یکتا می‌باشد، پایدار مجانبی سراسری نامیده می‌شود اگر x^* به مفهوم لیاپانوف پایدار باشد و در شرط زیر صدق کند:

$$\lim_{t\to\infty} x(t) = x^*.$$



شکل ۲۷: پایداری مجانبی.

تعریف ۵.۲. [۳۹] سیستم دینامیکی (۴۴) در نقطه x^* دارای پایداری نمایی سراسری با نرخ β است اگر برای هر نقطه آغازین دلخواه، مسیر $x(t)$ در رابطه زیر صدق کند

$$\|x(t) - x^*\| \leq \alpha \|x(t_\circ) - x^*\| e^{-\beta(t-t_\circ)}, \ \forall t \geq t_\circ,$$

که در آن α و β ثابت‌های مثبت و مستقل از نقاط آغازین هستند.
واضح است که پایداری نمایی سراسری پایداری مجانبی سراسری را نتیجه می‌دهد.

تعریف ۶.۲. [۳۹] مجموعه $M \subseteq \mathbb{R}^n$ یک مجموعه تغییر ناپذیر نسبت به سیستم (۴۴) گفته می‌شود اگر $x(t_\circ) \in M$ و $t_\circ \geq \circ$، به ازای هر $t \geq t_\circ$، $x(t) \in M$ باشد.

تعریف ۷.۲. [۳۹] فرض می‌کنیم که $\Omega \subset \mathbb{R}^l$ یک مجموعه بسته و محدب باشد که ممکن است بیکران باشد. $P_\Omega(x)\colon \mathbb{R}^l \to \Omega$ یک عملگر تصویر نامیده می‌شود و به‌صورت زیر تعریف می‌شود

$$P_\Omega(x) = \arg\min_{v \in \Omega} \|x - v\|,$$

که $\|.\|$ نرم l_2 را نشان می‌دهد.

تعریف ۸.۲. [۳۹] تابع $F : E \subset \mathbb{R}^n \to \mathbb{R}^n$ روی E در شرط لیپ شیتز صدق می‌کند (پیوسته لیپ شیتز است) اگر ثابت L وجود داشته باشد به‌طوری که برای هر دو نقطه $x, y \in E$ داشته باشیم:

$$\|F(x) - F(y)\| \leq L \|x - y\|.$$

تابع F پیوسته لیپ شیتز محلی روی E نامیده می‌شود اگر برای نقطه $x_\circ \in E$، یک εـ همسایگی مانند $N_\varepsilon(x_\circ) \subset E$ و یک ثابت $L_\circ$ وجود داشته باشد به‌طوری که برای هر دو نقطه $x, y \in N_\varepsilon(x_\circ)$ داشته باشیم:

$$\|F(x) - F(y)\| \leq L_\circ \|x - y\|.$$

قضیه ۱.۲. [۳۹] تابع $F : E \subset \mathbb{R}^n \to \mathbb{R}^n$ را در نظر بگیرید. اگر $F \in C^1(E)$، آن‌گاه F روی E لیپ شیتز محلی است.

تعریف ۹.۲. نگاشت $F : \mathbb{R}^n \to \mathbb{R}^n$

الف) یکنوا نامیده می‌شود اگر برای هر دو نقطه $x, y \in \mathbb{R}^n$ داشته باشیم

$$\langle x - y, F(x) - F(y) \rangle \geq 0.$$

ب) اکیداً یکنوا نامیده می‌شود اگر برای هر دو نقطه متمایز $x, y \in \mathbb{R}^n$ داشته باشیم

$$\langle x - y, F(x) - F(y) \rangle > 0.$$

پ) قویاً یکنوا نامیده می‌شود اگر یک ثابت $\beta > 0$ وجود داشته باشد که برای هر دو نقطه $x, y \in \mathbb{R}^n$ داشته باشیم

$$\langle x - y, F(x) - F(y) \rangle \geq \beta \|x - y\|^2.$$

که $\langle \cdot, \cdot \rangle$ ضرب داخلی را نشان می‌دهد.

تعریف ۱۰.۲. [۴۹] اگر $g : \Omega_1 \subset \mathbb{R}^l \to \mathbb{R}$، آن‌گاه هر مجموعه غیر تهی به شکل زیر

$$L(r) = \{u \in \Omega_1 | g(u) \leq r,\ r \in \mathbb{R}\},$$

یک مجموعه سطح از g نامیده می‌شود.

تعریف ۱۱.۲. [۴۹] یک ماتریس $n \times n$، $M(x)$ که عناصر آن m_{ij}، $i = 1, ..., n$ و $j = 1, ..., n$ توابعی هستند که روی مجموعه $S \subset \mathbb{R}^n$ تعریف شده‌اند، روی S نیمه معین مثبت نامیده می‌شود اگر

$$v^T M(x) v \geq 0,\ \forall v \in \mathbb{R}^n,\ \forall x \in S.$$

ماتریس $M(x)$ روی S معین مثبت قوی نامیده می‌شود اگر عددی مثبت مانند α وجود داشته باشد به‌قسمی‌که:

$$v^T M(x) v \geq \alpha \|v\|^2,\ \forall v \in \mathbb{R}^n,\ \forall x \in S.$$

اگر $\gamma(x)$ کوچکترین مقدار ویژه ماتریس $M(x)$ باشد سه حالت زیر را می‌توان نتیجه گرفت:

۱. $M(x)$ روی S نیمه معین مثبت است اگر و فقط اگر به ازای هر $x \in S$، $\gamma(x) \geq 0$.

۲. $M(x)$ روی Sمعین مثبت است اگر و فقط اگر به ازای هر $x \in S$، $\gamma(x) > 0$.

۳. $M(x)$ روی S قویاً معین مثبت است اگر و فقط اگر به ازای هر $x \in S$، $\gamma(x) \geq \alpha > 0$.

تعریف ۱۲.۲. [۴۹] فرض می‌کنیم $V: \mathbb{R}^n \to \mathbb{R}$ یک تابع به‌طور پیوسته مشتق پذیر باشد. $V(x)$ به‌طور شعاعی بی‌کران گفته می‌شود اگر

$$\|x\| \to \infty \Rightarrow V(x) \to \infty.$$

لم ۱.۲. [۴۹] فرض کنیم که $g : \Omega_1 \subset \mathbb{R}^l \to \mathbb{R}$ باشد که در آن Ω_1 بی‌کران است. آن‌گاه همه مجموعه‌های سطح g کراندار می باشند اگر وفقط اگر

$$\lim_{k\to\infty} g(u^k) = +\infty,$$

که در آن $\{u^k\} \subset \Omega_1$ و $\lim_{k\to\infty} \|u^k\| = +\infty$.

لم ۲.۲. [۴۹] فرض کنیم که $g : \Omega_1 \subset \mathbb{R}^l \to \mathbb{R}$ روی مجموعه بسته Ω_1 پیوسته باشد آن‌گاه همه مجموعه‌های سطح g کراندار می‌باشند اگر و فقط اگر مجموعه مینیمم کننده‌های g غیرتهی و کراندار باشند.

قضیه ۲.۲. (قضیه نقطه ثابت بروئر) [۴۹] تابع پیوسته $f : \Omega \to \Omega$، که Ω یک مجموعه فشرده و محدب می‌باشد حداقل یک نقطه ثابت دارد.

قضیه ۳.۲. [۴۹] فرض کنیم که تابع $F : K \subset \mathbb{R}^l \to \mathbb{R}^l$ روی K به‌طور پیوسته مشتق‌پذیر باشد، آن‌گاه F روی K یکنوا (اکیداً یکنوا) است اگر و فقط اگر ماتریس ژاکوبین آن $\nabla F(x)$ برای هر $x \in K$ نیمه‌معین مثبت (معین مثبت) باشد.

قضیه ۴.۲. [۴۹] فرض کنیم که تابع $F(x)$ روی K به‌طور پیوسته مشتق‌پذیر باشد و ژاکوبین $F(x)$ قویاً معین مثبت باشد آن‌گاه $F(x)$ قویاً یکنوا می‌باشد.

قضیه ۵.۲. [۳۹] در سیستم دینامیکی (۴۴) فرض کنیم که f یک تابع پیوسته از $\mathbb{R}^n$ به $\mathbb{R}^n$ باشد، آن‌گاه برای هر $t_\circ \geq \circ$ و $x_\circ \in \mathbb{R}^n$ ، یک جواب محلی $x(t)$ به ازای $t \in [t_\circ, \tau)$ که $\tau > t_\circ$ وجود دارد. علاوه براین اگر f در $x_\circ$ پیوسته لیپ شیتز محلی باشد آن‌گاه جواب یکتاست و اگر f در $\mathbb{R}^n$ پیوسته لیپ شیتز باشد آن‌گاه τ می‌تواند تا $+\infty$ توسعه داده شود.

تعریف ۱۳.۲. [۳۹] اگر یک جواب محلی تعریف شده در بازه $[t_\circ, \tau)$ نتواند به یک جواب محلی روی یک بازه بزرگتر $[t_\circ, \tau_1)$ که $\tau_1 > t_\circ$ می باشد گسترش یابد، آن‌گاه یک جواب ماکسیمال نامیده می‌شود و بازه $[t_\circ, \tau)$، بازه ماکسیمال وجود جواب نامیده می‌شود. بازه ماکسیمال وجود جواب مربوط به $x_\circ$ اغلب به‌صورت $[t_\circ, \tau(x_\circ))$ تعریف می‌شود.

قضیه ۶.۲. [۶۷] فرض کنیم $f : \mathbb{R}^n \to \mathbb{R}^n$ یک تابع پیوسته باشد. اگر $x(t)$، $t \in [t_\circ, \tau(x_\circ))$ یک جواب ماکسیمال باشد و $\tau(x_\circ) < +\infty$ آن‌گاه

$$\lim_{t\to\tau(x_\circ)} \|x(t)\| = +\infty.$$

قضیه ۷.۲. (قضیه پایداری لیاپانوف) [۲۳] فرض کنید که $x = ۰$ یک نقطه تعادل سیستم (۴۴) باشد و $V: \mathbb{R}^n \to \mathbb{R}$ یک تابع به‌طور پیوسته مشتق پذیر باشد به‌طوری‌که:

۱. $V(۰) = ۰$

۲. به ازای هر $x \in \mathbb{R}^n \backslash \{۰\}$، $V(x) > ۰$

۳. به ازای هر $x \in \mathbb{R}^n \backslash \{۰\}$، $\dot{V}(x) \leq ۰$

آن‌گاه $x = ۰$ نقطه پایداری سیستم خواهد بود.

قضیه ۸.۲. (قضیه پایداری مجانبی) [۲۳] تحت شرایط قضیه ۷.۲. اگر $V(\cdot)$ در شرایط زیر صدق کند

۱. $V(۰) = ۰$

۲. به ازای هر $x \in \mathbb{R}^n \backslash \{۰\}$، $V(x) > ۰$

۳. به ازای هر $x \in \mathbb{R}^n \backslash \{۰\}$، $\dot{V}(x) < ۰$

آن‌گاه $x = ۰$ پایدار مجانبی خواهد بود.

قضیه ۹.۲. (قضیه پایداری مجانبی سراسری) [۲۳] تحت شرایط قضیه ۷.۲. اگر $V(\cdot)$ به‌طور شعاعی بی‌کران باشد آن‌گاه نقطه $x = ۰$ پایدار مجانبی سراسری است.

قضیه ۱۰.۲. (اصل تغییرناپذیری لِسال) [۲۳] فرض کنیم یک تابع به‌طور پیوسته مشتق پذیر باشد، همچنین فرض کنیم

۱. $M \subset \mathbb{R}^n$ یک مجموعه فشرده و پایدار نسبت به جواب سیستم (۴۴) باشد.

۲. به ازای $x \in M$، $\dot{V}(x) \leq ۰$.

۳. E مجموعه همه نقاط M باشد به‌طوری که $\dot{V}(x) = ۰$.

۴. N بزرگترین مجموعه پایدار در E باشد.

آن‌گاه به ازای هر $x(t_۰) \in M$ که $t_۰ \geq ۰$ می‌باشد داریم:

$$\lim_{t \to \infty} dist(x(t), N) = ۰.$$

۳. نامساوی‌های وردشی

مقدمه‌ای بر نامساوی وردشی

نظریه نامساوی وردشی اولین بار در سال ۱۹۶۶ به وسیله هارتمن[۱] و استمپکیا[۲] به عنوان ابزاری برای مطالعه معادلات دیفرانسیل جزئی با کاربردهایی در مکانیک معرفی شد [۲۸]. این نامساوی‌ها با بعد نامتناهی بودند. نظریه نامساوی وردشی با بعد متناهی در سال ۱۹۸۰ که به وسیله دافِرموس[۳] معرفی شد، دافِرموس متوجه شد که شرایط تعادل شبکه ترافیک که به وسیله اسمیت[۴] در سال ۱۹۷۹ بیان شد ساختار نامساوی وردشی دارد. نامساوی‌های وردشی در اقتصاد، علم مدیریت، تحقیق در عملیات و هم‌چنین در مهندسی کاربردهای فراوانی دارد.

تعریف ۱.۳. فرض کنید $F : K \subset \mathbb{R}^n \to \mathbb{R}^n$ تابعی پیوسته روی مجموعه بسته و محدب K باشد، مسأله نامساوی وردشی که آن را با نماد $VI(F, K)$ نشان می‌دهند یافتن یک بردار $x^* \in K$ است به‌طوری که

$$F(x^*)^T(x - x^*) \geq 0, \quad \forall x \in K,$$

یا به‌طور معادل

$$\langle F(x^*), x - x^* \rangle \geq 0, \quad \forall x \in K, \tag{۴۵}$$

که در آن $\langle \cdot, \cdot \rangle$ ضرب داخلی در فضای اقلیدسی n بعدی است.

از دیدگاه هندسی، نامساوی وردشی (۴۵) بیان می‌کند که $F(x^*)$ بر ناحیه شدنی K در نقطه x^* عمود است (شکل ۲۸).

قضیه ۱.۳. [۱۱] مسأله بهینه‌سازی زیر را در نظر بگیرید:

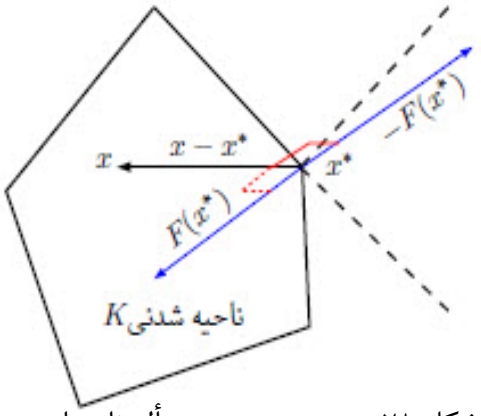
$$\begin{aligned} &\text{minimize} \quad f(x), \\ &\text{subject to} \quad x \in K, \end{aligned}$$

که در آن f یک تابع محدب و به‌طور پیوسته مشتق پذیر و K یک مجموعه محدب و غیرتهی در $\mathbb{R}^n$ است آن‌گاه x^* جواب بهینه مسأله فوق است اگر و تنها اگر x^* جواب مسأله نامساوی وردشی

$$\nabla f(x^*)^T (x - x^*) \geq 0, \quad \forall x \in K$$

باشد.

[1] Hartman
[2] Stampacchia
[3] Dafermos
[4] Smith



شکل ۲۸: تعبیر هندسی مسأله نامساوی وردشی.

قضیه ۲.۳.۲. [۵] فرض کنید تابع $F : K \subset \mathbb{R}^n \to \mathbb{R}^n$ روی K به‌طور پیوسته مشتق‌پذیر باشد و ماتریس ژاکوبین F متقارن و نیمه معین مثبت باشد، آن‌گاه تابع محدب حقیقی مقدار $f : K \to \mathbb{R}$ وجود دارد که در رابطه

$$\nabla f(x) = F(x),$$

صدق می‌کند و x^* جواب نامساوی وردشی $VI(F, K)$ است اگر و فقط اگر x^* جواب بهینه مسأله زیر باشد:

$$\begin{aligned} &\text{minimize} \quad f(x), \\ &\text{subject to} \quad x \in K, \end{aligned}$$

تعریف ۲.۳.۲. [۵] فرض می‌کنیم $\mathbb{R}^n_+$ ناحیه نامنفی روی $\mathbb{R}^n$ را نشان دهد و فرض کنیم $F : \mathbb{R}^n \to \mathbb{R}^n$ باشد. مسأله مکمل غیرخطی روی $\mathbb{R}^n_+$ یک سیستم از معادلات و نامعادلات است به صورت زیر است :
یافتن $x^* \geq 0$ به‌طوری که

$$F(x^*) \geq 0, \quad F(x^*)^T x^* = 0. \tag{۴۶}$$

هرگاه نگاشت F آفین باشد یعنی $F(x) = Mx + b$ که در آن M یک ماتریس $n \times n$ و b یک بردار $n \times 1$ می‌باشد مسأله (۴۶) یک مسأله مکمل خطی نامیده می‌شود.

قضیه زیر رابطه بین مسأله مکمل و مسأله نامساوی وردشی را نشان می‌دهد.

قضیه ۳.۳.۲. [۵] مسأله نامساوی وردشی $VI(F, \mathbb{R}^n_+)$ و مسأله مکمل (۴۶) جواب‌های یکسان دارند.

Aabstract

Support vector machines are powerful tools for data classification and regression. In recent years, many fast algorithms have been developed for support vector machines. Support vector machines are used in many military and engineering applications, real-time issues, such as classification in complex electromagnetic environments, diagnostics in medicine, and so on. Given that numerical convergence methods require more calculations and can not meet our needs in real-time problems. One promising process for training real-time support vector machines is the use of recurrent neural networks based on periodic execution.

In this thesis, we introduce two neural networks to solve support vector problems such as support vector regression problem and stochastic support vector regression with probabilistic constraints . We prove that the equilibrium points of the neural networks are equivalent to the optimal solution of the support vector problems. The stability and convergence of the proposed neural networks will also be demonstrated. We confirm the theoretical results by providing examples.

Keywords: Support vector, Neural network, Optimization, Regression, Random variable .

Faculty Of Mathematical Sciences

PhD Thesis in Control and Optimization

Solving a class of support vector machine problems using dynamic optimization methods

By: Amir Feizi

Supervisor:

Dr Alireza Nazemi

Advisor:

Dr Mohammad Reza Rabiei

February 2021