



دانشگاه تهران

دانشکده سامانه‌های هوشمند

یادگیری ماشین

تمرین چهارم

استاد درس

دکتر سامان هراتی زاده

زمان تحویل: ۱۰ دی

پاییز ۱۴۰۲

یادگیری ماشین

تمرین «۴»



دستیاران آموزشی

حمید نعمتی

پرستو پیله‌ور

امیرمحمد کویش‌پور

دکتر سامان هراتی‌زاده

دانشگاه تهران - دانشکده سامانه‌های هوشمند

نیم‌سال اول ۱۴۰۳-۱۴۰۲

ددلاین: ساعت ۲۳:۵۹ | ۱۴۰۲/۱۰/۱۰

>>

فهرست

- ۱..... یافتن معادله خط جداساز
- ۱..... Soft Margin Classifier
- ۱..... Kernel Trick
- ۲..... Multiple Classes
- ۲..... SVM++
- ۲..... طراحی تابع‌های منطقی
- ۳..... شبکه عصبی با تابع فعالساز Sigmoid
- ۳..... MLP (Regression)
- ۴..... MLP (Classification)
- ۵..... Sequential Minimal Optimization (SMO)

شکل‌ها

- ۱..... شکل ۱: نمایش نواحی جداساز
- ۳..... شکل ۱: دادگان آموزش
- ۳..... شکل ۲: ساختار شبکه عصبی

یادگیری ماشین

تمرین «۴»



ددلاین: ساعت ۲۳:۵۹ | ۱۴۰۲/۱۰/۱۰

دستیاران آموزشی

حمید نعمتی

پرستو پیله‌ور

امیرمحمد کویش‌پور

دکتر سامان هراتی‌زاده

دانشگاه تهران - دانشکده سامانه‌های هوشمند

نیم‌سال اول ۱۴۰۳-۱۴۰۲

یافتن معادله خط جداساز

در این سوال هدف حل یک مسئله SVM به صورت دستی است.

داده‌های گروه اول: $\{(-1, 0)\}$

داده‌های گروه دوم: $\{(0, 1), (1, 0)\}$

۱. با تشکیل یک دستگاه معادلات چند معادله و چند مجهول معادله خط جداساز^۱ و همچنین معادله خطوط support vector را بیابید.

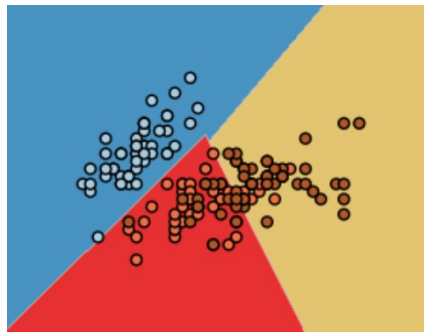
۲. در مورد SVM Dual Problem مطالعه کرده و علت وجود آن را مختصر توضیح دهید و سپس به این شیوه نیز معادله خط جداساز را بیابید.

Soft Margin Classifier

۱. معادله‌ای که Soft Margin Hyperplane قرار است بهینه‌سازی شود در زیر به شما داده شده است. توضیح دهید ضریب C چه نقشی در بهینه‌سازی دارد؟ همچنین برای خطای طبقه‌بندی یک حد بالا به دست بیاورید (می‌توانید از کتابخانه استفاده کنید).

$$\text{Minimize } \frac{1}{2} \|W\|^2 + C \sum_{i=1}^n \xi_i$$

۲. با استفاده از SVM مجموعه داده data.csv را طبقه‌بندی کنید، مقادیر زیر را برای پارامتر C آزمایش نمایید و هر بار همانند شکل زیر خطوط و نواحی جداساز را نمایش دهید. $C = \{0.1, 1, 10, 100, 1000\}$



شکل ۱: نمایش نواحی جداساز

Kernel Trick

۱. داده‌های زیر به صورت خطی جدا پذیر نیستند؟ یک Transformation Function دلخواه ارائه دهید که باعث شود داده‌ها خطی جداپذیر شوند سپس با استفاده از روش ضرایب لاگرانژ W و b اندازه حاشیه را پیدا کنید.

جدول ۱: نمونه دادگان

برچسب	دادگان
1	1
1	-1
-1	0

۲. کرنلی به صورت $K(x_i, x_j) = \tanh(ax_i^T x_j + b)$ تعریف شده است که به کرنل Sigmoid معروف است. تحقیق کنید که آیا Mercer Condition برای این کرنل برقرار است یا خیر؟ سپس ثابت کنید که رابطه زیر برقرار است و کوچکترین مقدار K را بیابید $\|\varphi(x_i) - \varphi(x_j)\|^2 \leq K$ (φ یک transformation function است).

$$\|\varphi(x_i) - \varphi(x_j)\|^2 \leq K$$

¹ Decision Boundry

یادگیری ماشین

تمرین «۴»



ددلاین: ساعت ۲۳:۵۹ | ۱۴۰۲/۱۰/۱۰

دستیاران آموزشی

حمید نعمتی

پرستو پیله‌ور

امیرمحمد کوش پور

دکتر سامان هراتی‌زاده

دانشگاه تهران - دانشکده سامانه‌های هوشمند

نیم‌سال اول ۱۴۰۳-۱۴۰۲

۳. ثابت کنید برای یک کرنل معتبر، رابطه زیر همیشه برقرار است.

$$K(x, y)^2 \leq K(x, x)K(y, y)$$

۴. در مورد کرنل RBF و Polynomial تحقیق کنید و بیان کنید که هر کدام برای چه مجموعه داده‌ای مناسب‌تر هستند.

۵. به کمک Naïve Bayse و SVM این مجموعه داده را طبقه‌بندی کنید، همچنین از کرنل‌های Linear، Polynomial و RBF در SVM بهره بگیرید. برای اینکار از Grid Search استفاده کرده و برای هر کرنل سه مقدار ۰/۱ و ۰/۵ و ۰/۹ را به عنوان مقدار Kernel coefficient تست کنید. Confusion Matrix را برای مدل‌های منتخب نمایش دهید و روی نتایج بحث کنید (توجه: می‌توانید از کتابخانه استفاده کنید و برای تبدیل متن به ویژگی می‌توانید از تابع CountVectorizer استفاده کنید).

Multiple Classes

۱. SVM برای طبقه‌بندی دو کلاس طراحی شده است. سه روش که آن را بتوان برای طبقه‌بندی چند کلاسه استفاده کرد ارائه دهید سپس یکی را به دلخواه انتخاب کرده و روی مجموعه داده‌گان MNIST پیاده‌سازی کنید. برای افزایش دقت مدل ترکیب هایپرپارامترهای مختلف را با استفاده از تابع Random Search امتحان کنید (می‌توانید از کتابخانه استفاده کنید). بهترین مدل را انتخاب کرده و خطوطی که این مدل یافته است را نمایش دهید. آیا ممکن است که ناحیه indecision وجود داشته باشد؟ برای داده‌های که در این ناحیه هستند چگونه برجسب انتخاب می‌کنید؟

SVM++

۱. در مورد مرتبه زمانی یافتن SVM تحقیق کنید و آن را با Classifier های دیگری که تا کنون یاد گرفته‌اید، مقایسه کنید.

۲. در مورد Relevance Vector Machine تحقیق کرده و نتایج را در گزارش خود ارائه دهید. SVM چه مشکلاتی دارد که باعث به وجود آمدن مدل‌های پیشرفته‌تر شده است؟ مزایای RVM بر SVM را لیست کنید. تحلیل خود را بر مزایا بودن موارد آورده شده، لازم است که ارائه کنید.

۳. در این بخش با SVR آشنا می‌شوید. ابتدا از Yahoo Finance داده Close Price را برای سهام Google دریافت کنید سپس از قیمت ۱۰ روز گذشته برای پیشبینی قیمت روز بعد استفاده کنید. این کار را با مدل Linear Regression و SVM انجام دهید. خطا را به صورت MAE و MSE محاسبه نموده و نتایج این دو مدل را باهم مقایسه کنید (می‌توانید از کتابخانه استفاده کنید).

طراحی تابع‌های منطقی

تابع‌های منطقی AND، OR و XOR را برای دو ورودی و با مقادیر علامت‌دار^۱ در نظر بگیرید. به اینصورت که مقادیر ورودی و خروجی می‌توانند مقدارهای ۱ و -۱ را اتخاذ کنند.

۱. کدامیک از این توابع منطقی را می‌توان به شکل طبقه‌بندی خطی و به صورت زیر نوشت؟ مقادیر w_1 ، w_2 و بایاس b را برای توابع مربوطه حساب کنید. کدامیک از این توابع منطقی را نمی‌توان به صورت طبقه‌بندی خطی بیان کرد؟

$$f(x; w) = \text{sign}(w_1x_1 + w_2x_2 + b)$$

۲. برای هر کدام از این توابع منطقی که آن را به صورت طبقه‌بندی خطی نمی‌توان بیان کرد از پرسپترون^۲ دو لایه استفاده کرده و مقادیر بایاس و بردارهای وزن را برای آن محاسبه کنید.

$$f(x; w) = \text{sign}(w^{\text{out}}T_h + b^{\text{out}})$$

$$h = [h_1, h_2]^T$$

$$h_1 = \text{sign}(w^{(1)T}x + b^{(1)})$$

$$h_2 = \text{sign}(w^{(2)T}x + b^{(2)})$$

برای حل این قسمت لازم است مقادیر سه بردار وزن $w^{(1)}$ ، $w^{(2)}$ و مقادیر بایاس $b^{(1)}$ ، $b^{(2)}$ محاسبه شوند.

¹ Signed

² Perceptron

یادگیری ماشین

تمرین «۴»



ددلاین: ساعت ۲۳:۵۹ | ۱۴۰۲/۱۰/۱۰

دستیاران آموزشی

حمید نعمتی

پرستو پیله‌ور

امیرمحمد کویش‌پور

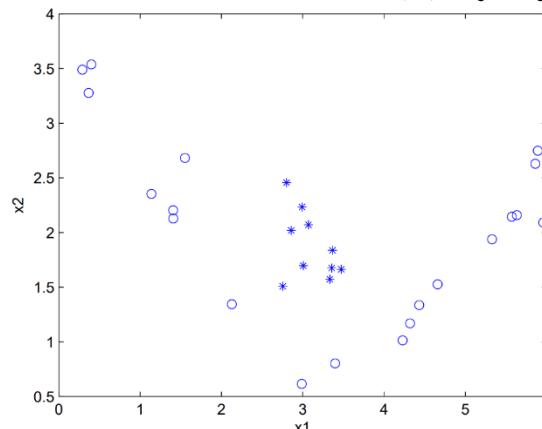
دکتر سامان هراتی‌زاده

دانشگاه تهران - دانشکده سامانه‌های هوشمند

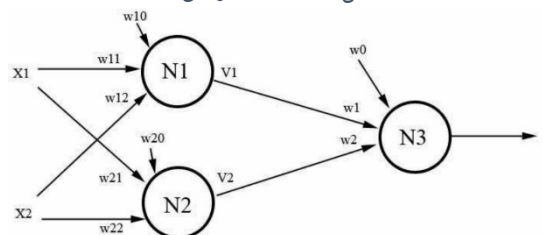
نیم‌سال اول ۱۴۰۳-۱۴۰۲

شبکه عصبی با تابع فعال‌ساز Sigmoid

داده‌های آموزش شکل ۱ را در نظر گرفته که در آن «*» مقدار یک و «O» مقدار صفر را نشان می‌دهند. همچنین شبکه از تابع فعال‌ساز Sigmoid استفاده می‌کند. هدف انتخاب بردارهای وزن و مقادیر بایاس برای شبکه عصبی شکل ۲ به گونه‌ای است که این شبکه قادر باشد دادگان آموزش مورد نظر را طبقه‌بندی کند.



شکل ۲: دادگان آموزش



شکل ۳: ساختار شبکه عصبی

۱. ابتدا مرزهای تصمیم^۱ را برای N_1 و N_2 رسم کنید (برای مثال؛ برای نورون N_1 ، مرز تصمیم عبارت است از خط $w_{11}x_1 + w_{12}x_2 = 0$).
۲. در شکل بعدی که دستگاه محور مختصات V_1 و V_2 را دارد، $\{V_1(x_1, x_2), V_2(x_1, x_2)\}$ چند نقطه آموزشی را رسم کرده و مرز تصمیم را به گونه‌ای ارائه دهید که در نهایت شبکه عصبی داده‌های آموزشی را به درستی دسته‌بندی نماید.
۳. بردارهای وزن و مقادیر بایاس انتخاب شده را اعلام کنید.

MLP (Regression)

- هدف این تمرین پیش‌بینی قیمت خانه برحسب ویژگی‌های آن است. اطلاعات مربوط به این داده‌ها در فایل data_description.txt و خود داده‌ها در فایل house_prices.csv قرار دارند (دقت کنید که برای «پیاده‌سازی شبکه عصبی» لازم است که «حتماً» از «پایتورچ و نامپای» استفاده کنید، در غیر اینصورت فقط می‌توانید از «ماژول‌های داخلی پایتون» استفاده کنید).
۱. ابتدا پیش‌پردازش‌های لازم برای آماده‌سازی دیتا را انجام داده و آن را به اختصار توضیح دهید.
 ۲. به صورت تصادفی ۸۰ درصد داده‌ها را به عنوان داده آموزشی و ۲۰ درصد را به عنوان داده آزمون در نظر بگیرید. شبکه عصبی چند لایه‌ای را طراحی کنید که قیمت خانه را پیش‌بینی کند. تعداد لایه و تعداد نورون و تابع فعال‌ساز مناسب برای هر لایه را مشخص کنید و آنها را تحلیل کنید.

¹ Decision Boundaries

یادگیری ماشین

تمرین «۴»



ددلاین: ساعت ۲۳:۵۹ | ۱۴۰۲/۱۰/۱۰

دستیاران آموزشی

حمید نعمتی

پرستو پیله‌ور

امیرمحمد کویش‌پور

دکتر سامان هراتی‌زاده

دانشگاه تهران - دانشکده سامانه‌های هوشمند

نیم‌سال اول ۱۴۰۳-۱۴۰۲

۳. با ثابت در نظر گرفتن پارامترهای به دست آمده در قسمت (۲) و با در نظر گرفتن MSE به عنوان تابع $loss$ ، مقادیر $metric$ های MSE و MAE را در هر اپیک^۱ برای داده آموزشی و داده آزمون در یک نمودار رسم کنید (یک نمودار برای معیار MSE و یک نمودار برای MAE). برای داده آزمون نمودار مقادیر پیش‌بینی شده بر حسب مقادیر واقعی را رسم نمایید. تعداد اپیک بهینه در این قسمت را مشخص کنید.
۴. با ثابت در نظر گرفتن پارامترهای به دست آمده در قسمت (۲) و با در نظر گرفتن MAE به عنوان تابع $loss$ ، مقادیر $metric$ های MSE و MAE را در هر اپیک برای داده آموزشی و داده آزمون در یک نمودار رسم کنید (یک نمودار برای معیار MSE و یک نمودار برای MAE). برای داده آزمون نمودار مقادیر پیش‌بینی شده بر حسب مقادیر واقعی را رسم نمایید. تعداد اپیک بهینه در این قسمت را مشخص کنید.
۵. ابتدا روابط ریاضی MSE و MAE را بنویسید و سپس نتایج قسمت (۳) و (۴) را با هم مقایسه کرده و توضیح دهید.
۶. با ثابت در نظر گرفتن پارامترهای به دست آمده در قسمت (۲) و با در نظر گرفتن MAE به عنوان تابع $loss$ ، مقادیر $metric$ های MSE و MAE را در هر اپیک برای داده آموزشی و داده آزمون در یک نمودار رسم کنید (یک نمودار برای معیار MSE و یک نمودار برای MAE). برای داده آزمون نمودار مقادیر پیش‌بینی شده بر حسب مقادیر واقعی را رسم نمایید. تعداد اپیک بهینه در این قسمت را مشخص کنید.
۷. ابتدا روابط ریاضی MSE و MAE را بنویسید و سپس نتایج قسمت (۳) و (۴) را با هم مقایسه کرده و توضیح دهید.

MLP (Classification)

- یک طبقه‌بند با استفاده از شبکه MLP برای دسته‌بندی دادگان [sonar](#) طراحی کنید. این مجموعه دادگان الگوی برخورد سیگنال سونار به سیلندر در حالت‌های مختلف و هدف دسته‌بندی جنس سیلندر (آهن یا سنگ بودن) است. توضیحات بیشتر در لینک قرار داده شده است. دقت کنید که برای «پایده‌سازی شبکه عصبی» لازم است که «حتمًا» از «پایتورچ و نامپای» استفاده کنید، در غیر اینصورت فقط می‌توانید از «ماژول‌های داخلی پایتون» استفاده کنید.
۱. مجموعه داده را با استفاده از تقسیم‌بندی مناسب، به دادگان آموزش ارزیابی و آزمون تقسیم کنید. درباره‌ی روش تقسیم‌بندی انتخابی خود و مزایای آن توضیح دهید. پیش‌پردازش‌های مناسب را برای مجموعه داده در نظر بگیرید. مدل MLP خود را با حداقل ۲ لایه بسازید. معماری شبکه ساخته شده را ذکر کنید.
 ۲. نمودار تغییرات دقت و خطای مدل را در هر اپیک برای دادگان آموزش و ارزیابی نمایش دهید.
 ۳. مقادیر خطا، دقت و ماتریس درهم‌ریختگی دادگان آزمون را گزارش کنید.
 ۴. درباره‌ی معیار خطای مورد استفاده در آموزش شبکه و جایگزین‌های موجود و برتری معیار انتخابی بحث کنید.
 ۵. آیا معیار دقت^۲ استفاده شده عیار مناسبی برای توانمندی شبکه در دسته‌بندی این مجموعه داده است؟ مقدار معیارهای منتخب دیگر را به دست آورید.
 ۶. حال برای آموزش این مدل از روش $stochastic\ mini\ batch\ based$ در سه حالت مختلف با اندازه $batch$ های ۱۲۸، ۶۴ و ۳۲ استفاده کنید. اندازه $batch$ متناسب را اعلام و علت انتخاب آن را توضیح دهید.
 ۷. توابع فعال‌ساز را در لایه‌های ماقبل آخر تغییر داده و نتایج را گزارش کنید. مزایا و معایب توابع فعال‌ساز انتخابی را نسبت به یکدیگر بیان کنید (از توابع $Tanh$, $Sigmoid$, $ReLU$ استفاده کنید).
 ۸. افزودن لایه به شبکه تأثیری در خروجی دارد؟ فرضیه خود را با افزودن تعداد لایه‌های مختلف بررسی نمایید.
 ۹. پارامترهای بهترین شبکه به دست آمده با استفاده از تغییر هاپرپارامترهای مربوطه در بخش‌های قبلی را گزارش کنید. آیا باز هم می‌توان این شبکه را بهبود بخشید؟

¹ Epoch

² Accuracy

یادگیری ماشین

تمرین «۴»



ددلاین: ساعت ۲۳:۵۹ | ۱۴۰۲/۱۰/۱۰

دستیاران آموزشی

حمید نعمتی

پرستو پیله‌ور

امیرمحمد کویش‌پور

دکتر سامان هراتی‌زاده

دانشگاه تهران - دانشکده سامانه‌های هوشمند

نیم‌سال اول ۱۴۰۳-۱۴۰۲

۱۰. آیا شبکه با نورون و لایه‌های کمتر، در ایپاک‌های کمتری بیش‌پرازش^۱ می‌شود؟ فرضیه‌ی خود را آزمایش کرده و علت را توضیح دهید.

امتیازی: Sequential Minimal Optimization (SMO)

هدف از این سوال دسته‌بندی دادگان با الگوریتم SVM و استفاده از SMO به عنوان بهینه‌ساز است. مجموعه‌دادگان قرار داده شده در مسیر SMO شامل سه مجموعه و هر مجموعه شامل دو ویژگی و برچسبی با دو کلاس ۰ و ۱ است. تعداد ۵۱ و ۸۶۳ نمونه در مجموعه‌داده اول و دوم و ۲۱۱ و ۲۰۰ نمونه برای بخش آموزش و اعتبارسنجی در مجموعه‌داده سوم است. در حالت ساده با فرموله کردن SVM به مسئله زیر می‌رسیم:

$$\min_{w,b} \left(\frac{1}{2} \|W\|^2 \right) \quad \text{s.t.} \quad y_i(W^T x_i + b) - 1 \geq 0, \quad \forall (x_i, y_i) \in \mathcal{D}$$

با عوض کردن فرض‌های مسئله، نرم کردن^۲ مرزها و پذیرفتن خطا به رابطه زیر می‌رسیم:

$$\min_{w,b,\xi} \left(\frac{1}{2} \|W\|^2 \right) + C \sum_{i=1}^N \xi_i \quad \text{s.t.} \quad y_i(W^T x_i + b) + \xi_i - 1 \geq 0, \quad \forall (x_i, y_i) \in \mathcal{D} \quad \xi_i \geq 0$$

پس از اعمال لاگرانژین به سراغ مسئله دوگان^۳ می‌رویم:

$$L(a) = \sum_{i=1}^N a_i - \frac{1}{2} \sum_{i,j=1}^N a_i a_j y_i y_j \langle x_i, x_j \rangle \quad \text{s.t.} \quad 0 \leq a_i \leq C, \quad \sum_{i=1}^N a_i y_i = 0$$

روش حل بهینه‌سازی متوالی کمینه^۴ برای الگوریتم SVM به همراه تمرین (Simplified SMO Algorithm.pdf) پیوست شده است.

۱. الگوریتم SVM را بر مجموعه‌داده اول اجرا کرده و نتیجه (مرز تصمیم) را همراه با داده‌ها ترسیم کنید. پارامتر C را بین مقادیر ۰.۱، ۱، ۱۰، ۱۰۰ تغییر داده و نتایج را تحلیل کنید (الگوریتم SMO به دو مقدار tolerance و max_passes نیاز دارد که آن‌ها را در کل این سوال به ترتیب ۰.۰۰۱ و ۵ در نظر بگیرید).
۲. کرنل گوسی را پیاده‌سازی کرده و برای بررسی صحت پیاده‌سازی مقادیر $x_1 = [1; 2; 1]$ و $x_2 = [0; 4; -1]$ و $\sigma = 2.0$ را به عنوان ورودی کرنل داده و خروجی مورد انتظار باید حدود ۲۳۴۶۵۲.۰ باشد.
۳. الگوریتم را با کرنل گوسی و پارامترهای $\sigma = 0.1$ و $C = 1$ بر مجموعه‌داده دوم پیاده‌سازی کرده و مرز تصمیم را همراه با داده‌ها ترسیم کنید.
۴. الگوریتم را با کرنل گوسی بر مجموعه‌داده سوم پیاده‌سازی کرده و مقادیر σ و C را در بازه $[30, 10, 3, 1, 0.3, 0.1, 0.03, 0.01]$ تغییر دهید. با بررسی دقت^۵ در مجموعه‌داده اعتبارسنجی بهترین پارامترها را انتخاب کنید.
۵. برای بخش «۴» مرز تصمیم بهترین پارامترها را به همراه داده‌ای آموزش و اعتبارسنجی جداگانه رسم کنید.

توجه لازم است تا خروجی‌های را با دو پیاده‌سازی موجود برای SMO در گیت‌هاب^۶ اجرا کرده و نتایج کار خود را با نتایج پروژه‌های موجود مقایسه کنید (توجه کنید تنها می‌توانید از پروژه‌های ۵ سال اخیر برای آموزش، ارزیابی و مقایسه نتایج استفاده کنید).

¹ Overfit

² Norm

³ Dual

⁴ Simplified SMO Algorithm

⁵ Accuracy

⁶ Github