به نام خدا



دانشگاه صنعتی امیرکبیر

(پلی‌تکنیک تهران)

درس یادگیری ماشین

استاد ناظرفرد

تمرین سوم

علیرضا مازوچی

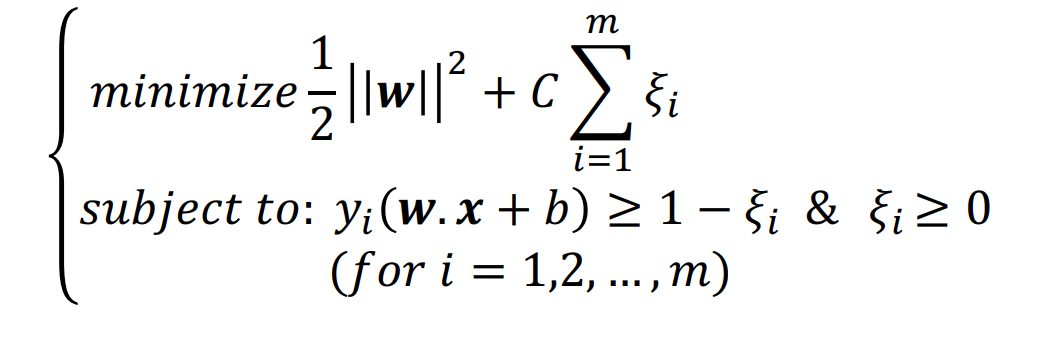
۴۰۰۱۳۱۰۷۵

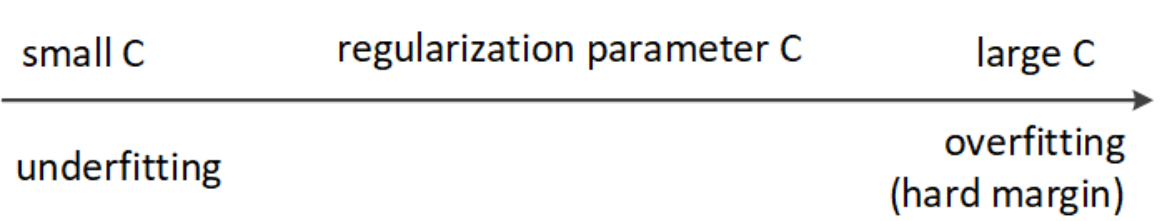
بخش اول: پرسش‌های تشریحی

سوال ۱

الف) نادرست؛ روش SVM‌ یک روش non-parametric است. چراکه نمی‌توان برای این روش و نسبت به تعداد اعضای مجموعه، تعداد مشخصی پارامتر معرفی کنیم و بسته به شرایط مسئله و تعداد داده‌هایی که بردار پیشتیبان محسوب می‌شوند، پارامتر‌های مدل تعیین خواهند شد.

ب) درست؛ در حالت soft margin و با تعیین پارامتر منظم‌ساز مناسب تابع خطا SVM‌، دو هدف وجود دارد: افزایش حاشیه میان دو کلاس و قرار گرفتن داده‌های آموزشی در کلاس صحیح. طبیعتا با وجود این دو هدف به صورت همزمان مدل تعمیم‌پذیری مناسب خود را حفظ می‌کند و دچار بیش‌برازش نمی‌شود. به عنوان مثال به دنبال خط جداکننده‌ای نمی‌گردد که به قیمت پیش‌بینی درست داده‌ها، حاشیه بسیار پایینی برای مدل به وجود آورد. طبیعتا اگر پارامتر منظم‌ساز در رابطه زیر زیاد باشد، امکان بیش‌برازش وجود دارد:





ج) نادرست؛ در الگوریتم SVM، عملا تعدادی از داده‌ها (بردار‌های پشتیبان) خط جداکننده را تعیین می‌کنند. پس در شرایطی که داده‌های مرزی دارای خطا باشند خروجی تحت تاثیر قرار خواهد گرفت و بی‌تاثیر نخواهد بود؛ بدیهی است که با استفاده از soft-margin و پارامتر C مناسب می‌توان این تاثیر را کم کرد.

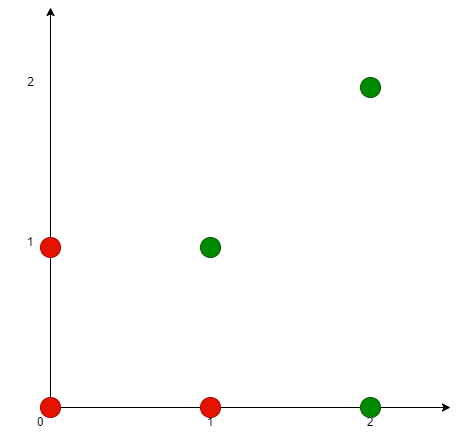
د) نادرست؛ خطای صفر برای بعضی از مجموعه‌های آموزشی و با هر الگوریتمی (هرچند بیش‌برازش شده!) امکان‌پذیر نیست. چراکه ممکن است دو داده با کلاس متفاوت به دلیل وجود خطا یا سایر دلایل دقیقا در یک نقطه قرار گرفته باشند. طبیعتا دسته‌بند برای آن نقطه اگر هر تصمیمی بگیرد، خطا صفر نخواهد شد!

ه) نادرست؛ امکان تخصیص وزن منفی به یک دسته‌بند وجود دارد. باتوجه به آنکه وزن یک دسته‌بند از رابطه زیر حاصل می‌شود، اگر خطای دسته‌بندی بالای 50٪ باشد، این دسته‌بند وزن منفی می‌گیرد:

و) درست؛ در الگوریتم آدابوست داده‌ای که در یک گام به اشتباه پیش‌بینی شود، وزنی بیش از پیش می‌گیرد؛ در نتیجه اگر داده‌ نویزی در محل کلاس اشتباهی قرار بگیرد، وزن زیادی را در طی گام‌های متوالی خواهد گرفت و عملکرد الگوریتم را دچار مشکل می‌کند.

سوال ۲

الف) بله به وضوح جداپذیرند.



ب)

باتوجه به تصویر قسمت الف، چهار نقطه (0,1)، (1,0)، (1,1) و (2,0) بردار‌های پشتیبان هستند و داریم:

معادله خط جداکننده:

ج) خیر؛ اندازه حاشیه ثابت می‌ماند. چراکه دو بردار پشتیبان برای کلاس مثبت و دو بردار پشتیبان برای کلاس منفی وجود دارد که اگر یکی از داده‌ها حدف شود،‌ یک کلاس دو داده‌اش باقی می‌ماند که به واسطه آن و تک داده باقی‌مانده همچنان خط جداکننده بدون تغییر می‌ماند.

د) این گزاره صحیح است. فرض کنید این گزاره غلط باشد؛ یعنی مثالی وجود داشته باشد که در آن با حذف یکی از بردار‌های پشتیان اندازه حاشیه کاهش پیدا کند. حالتی که این بردار پشتیبان حذف شده است را درنظر بگیرید. در این حالت اگر بردار پشتیبان را اضافه کنیم، طبیعتا باید اندازه حاشیه افزایش یابد ولی چنین چیزی امکان ندارد. چراکه داده‌های فعلی در بیشترین فاصله ممکن از هم قرار گرفته‌اند و وجود یک داده‌ی جدید نه تنها نمی‌تواند این حاشیه را بیشتر کند بلکه ممکن است خود در داخل این حاشیه قرار بگیرد و لازم شود تا حاشیه کوچکتری درنظر گرفته شود. پس به تناقض می‌خوریم و فرض اولیه اثبات می‌شود.

سوال ۳

الف) خیر. با دیدن تصویر زیر مشخص است.

D:\Univercity\Machine Learning\HW3\ML_HW3\MachineLearning-HW-HW3-2.drawio.png

ب)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| class | x | 1 |  |  |
| + | 0 | 1 | 0 | 0 |
| - | -1 | 1 |  | 1 |
| - | +1 | 1 |  | 1 |

در این حالت می‌توان داده‌ها را به صورت خطی از هم جدا کرد. تنها بعدی که باعث جداشدن داده‌های مثبت و منفی شده‌است بعد است. در این بعد مقدار 0.5 بهترین حاشیه را برای جداسازی فراهم می‌کند. پس یک ابرصفحه جداکننده در فضای جدید خواهد بود. لذا می‌توان معادله زیر را برای خط جداکننده پیشنهاد کرد:

سوال ۴

|  |  |  |
| --- | --- | --- |
| ردیف | تنظیمات | جواب |
| الف | SVM خطی، soft-margin، C=0.1 | ۴ |
| ب | SVM خطی،‌ soft-margin، C=10 | ۳ |
| ج | SVM، hard-margin، | ۲ |
| د | SVM، hard-margin، | ۱ |
| ه | SVM، hard-margin، | 6 |

ابتدا الف و ب را بررسی می‌کنیم. طبیعی است که جواب این دو تصویر ۳ و ۴ خواهد بود. در تصویر ۳ خط در جایی قرار دارد که داده‌ها را به طور کامل از هم جدا کند ولی در تصویر ۴، بیشینه‌کردن حاشیه نیز اهمیت پیدا کرده است به گونه‌ای که دو داده در حوالی مرز جداکننده قرار گرفته‌اند. باتوجه به اینکه C ابرپارامتری است که زیادبودن آن اهمیت خطای slack را بیشتر می‌کند پس تنظیم ب با C بیشتر مربوط به ۳ و تنظیم الف مربوط به ۴ است.

تنظیم ج دارای یک کرنل چندجمله‌ای است؛‌ در نتیجه نمی‌تواند خروجی که در تصاویر ۱ و ۵ و ۶ نشان داده شده است را داشته باشد. پس تصویر ۲ بهترین تطبیق را با آن خواهد داشت.

نهایتا نوبت به دو تنظیم د و ه می‌رسد. این دو تنظیم مانند یکدیگر هستند با این تفاوت که تنظیم ه دارای مقدار گامای بیشتری است. در میان تصاویر هم تنها تصویر ۱ و ۶ از تعدادی توزیع گاوسین تشکیل شده است. وقتی مقدار گاما خیلی زیاد باشد شعاع ناحیه یک کلاس تنها شامل خود بردار پشتیبان می‌شود و وقتی این مقدار خیلی کم باشد، شعاع ناحیه بخش زیادی از داده‌های آموزشی یا همه‌ی آن را در می‌گیرد. با این توضیح به نظر می‌رسد تصویر ۶ متعلق به یک تنظیم با گامای بالا یعنی تنظیم ه و تصویر ۱ متعلق به یک تصویر با گامای پایین یعنی تنظیم د خواهد بود.

سوال ۵

از هشت داده موجود یک داده به اشتباه دسته‌بندی شده است پس:

بخش اول: پیاده‌سازی

سوال ۱

الف) اگر

سوال ۲

الف) اگر