بنام خدا

مریم رضوانی
شماره دانشجویی: ۹۹۲۱۱۶۰۰۱۹

تمرین سری پنجم:درس یادگیری ماشین

1)

الف) در مورد ماشین بردار پشتیبان خطای کلی به چه معناست؟

a) میزان فاصله خط از بردارهای پشتیبان

b) میزان دقت svm برای پیشبینی نتایج دادههای تست

c) مقدار آستانه خطا در svm

خطای کلی به معنای مقدار آستانه خطا در SVM است. این مقدار نشان می دهد که چه میزان از داده ها را می توان در طرف اشتباه از مرز جداسازی قرار داد. اگر خطای کلی صفر باشد، یعنی ماشین بردار پشتیبان سخت یا SVM with hard margin است که هیچ خطایی را تحمل نمی کند. اگر خطای کلی مثبت باشد، یعنی ماشین بردار پشتیبان نرم است که برخی از خطاها را تحمل می کند

بنابراین، پاسخ صحیح به سوال شما گزینه c) مقدار آستانه خطا در SVM است.

ب) وقتی پارامتر c روی بی‌نهایت تنظیم می‌شود کدام یک از موارد زیر صحیح می‌باشد؟

a) ابر صفحه بهینه در صورت وجود صفحه‌ای خواهد بود که داده‌ها را بطور کامل جدا می کند

b) طبقه بندی کننده soft margin داده‌ها را جدا می‌کند

c) هیچ یک از موارد فوق

گزینه a) صحیح است؛ با ضریب c به خطا اهمیت می‌دهیم؛ c برابر بی‌نهایت سعی می‌کند خطا نداشته باشد در این حالت، ما یک طبقه بندی کننده hard margin داریم که سعی می‌کند یک ابر صفحه بهینه پیدا کند که داده ها را بطور کامل جدا کند.

ج) اثر بخشی یک svm به کدام یک از موارد زیر بستگی دارد؟

a) انتخاب هسته

b) پارامترهای هسته

c) پارامتر soft margin

d) همه موارد

- انتخاب هسته: تابع هسته برای تبدیل داده های ورودی از فضای ویژگی اصلی به فضای ویژگی با ابعاد بالاتر استفاده می‌شود. انواع مختلفی از توابع هسته وجود دارد، مانند خطی، چندجمله ای، تابع پایه شعاعی (RBF) و سیگموئید. هر تابع هسته ویژگی های خاص خود را دارد و برای انواع مختلف داده‌ها مناسب است

- پارامترهای هسته: پارامترهای هسته مقادیری هستند که تأثیر تابع هسته را بر روی داده‌ها تعیین می‌کنند. برای مثال، در تابع هسته چندجمله‌ای، پارامترهای هسته شامل درجه و ثابت هستند. این پارامترها باید به گونه ای انتخاب شوند که بتوانند داده ها را به خوبی تمایز دهند

- پارامتر soft margin: پارامتر soft margin یا C یک عامل تنظیم است که میزان جریمه را برای داده‌های نادرست طبقه بندی شده تعیین می‌کند. اگر C بزرگ باشد، SVM سعی می‌کند هیچ داده ای را اشتباه طبقه بندی نکند، اما ممکن است به خطر بیفتد. اگر C کوچک باشد، SVM می‌تواند برخی از داده‌ها را اشتباه طبقه‌بندی کند، اما ممکن است حاشیه بزرگتری داشته باشد

د) در خصوص روش soft margin و hard margin با هسته گوسی کدام یک از موارد زیر صحیح است؟

a) با کاهش پارامتر امکان رخداد بیش‌برازش افزایش می‌یابد.

b) با افزایش هسته گاوسی نرخ خطا در داده‌های آزمون افزایش نمی‌یابد.

c) با کاهش ضریب جریمه، در الگوریتم soft margin svm دسته بندی دچار بیش برازش می شود.

d) با افزایش ضریب جریمه، در الگوریتم soft margin svm میزان داده‌های آموزش نادرست دسته بندی شده کاهش، اما داده‌های آزمون درست دسته بندی شده افزایش می‌یابد.

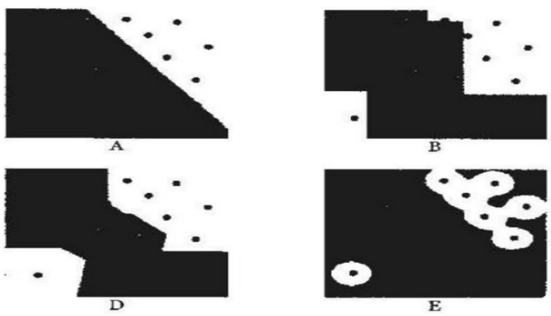
گزینه c صحیح است. زیرا با کاهش ضریب جریمه یا پارامتر C، SVM کمتر به داده های نادرست طبقه بندی شده جریمه می کند و ممکن است حاشیه را بیش از حد بزرگ کند و داده ها را بیش برازش کند. گزینه های a و b غلط هستند چون با کاهش پارامتر، احتمال بیش برازش کاهش می یابد و با افزایش هسته گاوسی، نرخ خطا در داده های آزمون ممکن است افزایش یابد. گزینه d نیز غلط است چون با افزایش ضریب جریمه، SVM سخت گیرتر می شود و ممکن است داده های آموزش را کمتر اشتباه طبقه بندی کند، اما داده های آزمون را بیشتر اشتباه طبقه بندی کند.

2)

یک ماشین بردار پشتیبان خطی سخت، (SVM Linear Hard) یا SVM خطی بدون در نظرگرفتن پنالتی) در یک مسأله دسته بندی دو کلاسه در فضای دو بعدی با N داده آموزشی داده شده است. نتیجه حاصل K=2 بردار پشتیبان بوده است. در صورتی که یک داده برچسب دار دلخواه به مجموعه داده‌های قبلی اضافه کرده و مجدداً دسته بند را آموزش دهیم، حداکثر چند بردار پشتیبان ممکن است بدست آید؟

حداکثر تعداد بردارهای پشتیبان برابر با تعداد داده های آموزشی است. زیرا اگر همه داده ها روی یک خط قرار داشته باشند، همه آن‌ها بردار پشتیبان خواهند بود. اما این حالت کم احتمال است و معمولا تعداد بردارهای پشتیبان کمتر از تعداد داده های آموزشی است. بنابراین، اگر یک داده برچسب دار دلخواه به مجموعه داده های قبلی اضافه کنیم، حداکثر تعداد بردارهای پشتیبان برابر با N+1 خواهد بود. البته این تعداد ممکن است کمتر باشد اگر داده اضافه شده دور از ابر صفحه جداکننده باشد و تأثیری بر حاشیه نداشته باشد.

۳-برای داده های نشان داده شده درشکل زیر، کدام خروجی میتواند به ترتیب حاصل ماشین بردار خطی، ماشین بردار با کرنل گاوسی $\sigma = 1$ و ماشین بردار پشتیبان با کرنل گاوسی $\sigma = 0.25$ باشد.



عرض هسته (σ): انعطاف پذیری مرز تصمیم را کنترل می‌کند. مقادیر σ بالاتر منجر به مرزهای هموارتر می‌شود و الگوهای جهانی بیشتری را ثبت می‌کند. مقادیر σ کمتر منجر به مرزهای پیچیده‌تر می‌شود و بر الگوهای محلی تمرکز می‌کند.

درک تأثیر عرض هسته: σ = 1: انعطاف پذیری متوسط، متعادل کردن الگوهای جهانی و محلی.

σ = 0.25: مرزهای پیچیده‌تر، با تأکید بر الگوهای محلی.

تجزیه و تحلیل داده ها و خروجی ها: نقاط داده را بررسی کنید: آیا آن‌ها به صورت خطی در فضای اصلی خود قابل تفکیک هستند؟ آیا آن‌ها الگوهای غیر خطی را نشان می‌دهند که ممکن است از هسته گاوسی بهره‌مند شوند؟ مرزهای تصمیم گیری بالقوه را در نظر بگیرید: برای σ = 1، مرزی را تجسم کنید که الگوهای جهانی و محلی را متعادل می‌کند. برای σ = 0.25، مرزی را تصور کنید که بر جزئیات محلی تمرکز دارد و به طور بالقوه مناطق تصمیم گیری پیچیده‌تری را ایجاد می‌کند. خروجی‌ها را با مرزهای تصمیم مورد انتظار مطابقت دهید: به نظر می‌رسد کدام خروجی بیشتر از یک مرز هموارتر و جهانی تر (σ = 1) ناشی می شود؟ کدام خروجی با یک مرز پیچیده‌تر و متمرکز به صورت محلی (σ = 0.25) تراز بهتری دارد؟

ملاحظات اضافی: پارامترهای SVM: سایر پارامترها مانند منظم سازی نیز می توانند بر مرز تصمیم گیری تأثیر بگذارند.

پیچیدگی داده ها: پیچیدگی ذاتی داده ها در خود نقش دارد.

طبق مفاهیم بالا شکل A برای کرنل گوسی، سیگما برابر با 1 مطابقت دارد و شکل E برای کرنل گوسی، سیگما برابر با 0.25مطابقت می‌کند.اندازه سیگما را اگر درست انتخاب نکنیم و اگر خیلی کوچک در نظر بگیریم داده‌ها را حفظ می‌کند و مشکل اورفیت داریم.

۴- فرض کنید که داده های جدول ۱ را بخواهیم توسط یک ماشین بردار پشتیبان جدا کنیم.

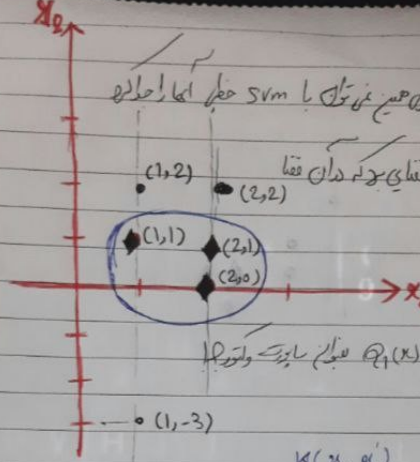
الف) آیا این کار امکان‌پذیر است؟

ب) در صورت استفاده از تابع تبدیل غیرخطی $\varphi_1(x) = (x_1{}^2 + x_2{}^2, x_1 - x_2)$ و $\varphi_2(x) =$ $(x_2, x_1 - x_2)$چطور؟

ج) آیا دو نقطه $A = \begin{bmatrix}4\\0\end{bmatrix}$ و $B = \begin{bmatrix}9\\-4\end{bmatrix}$ برای تابع $\varphi_1(x)$ به عنوان ساپورت وکتورها هستند؟

Data Point	Label
(1,1)	1
(2,1)	1
(2,0)	1
(1,2)	-1
(2,2)	-1
(1.-3)	-1

Data Point	label
(1,1)	1
(2,1)	1
(2,0)	1
(1,2)	-1
(2,2)	-1
(1,-3)	-1



الف) داده‌ها جدایی پذیر خطی نیستند برای همین نمی‌توان با svm خطی آنها را جدا کرد

با استفاده از کرنل می‌توان داده‌ها را به فضایی برد که در آن فضا جدایی پذیر خطی باشند.

2) $A=\begin{bmatrix}4\\0\end{bmatrix}, B=\begin{bmatrix}9\\-4\end{bmatrix}$ برای $\varphi_1(x)$ به عنوان ساپورت وکتورها

$K(x, x')$

ب)

$$\varphi_1(x) = (x_1^2 + x_2^2, x_1 - x_2)$$

$$g(x) = \left(\sum_{i=1}^{m} \alpha^T y^T K(x^T, x)\right) + b$$

$$\varphi_2(x) = (x_2, x_1 - x_2)$$

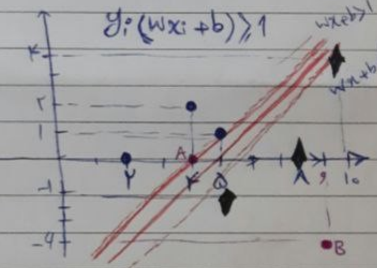
$$wx + b = 0$$

$$y_i(wx_i + b) \geq 1$$

$$W = \sum_{i=1}^{n} \alpha_i y_i x_i$$

$$b = y_i - \sum_{i=1}^{m} \lambda_i y_i x_i x_j$$

$\varphi_1(x)$



D	L
(2,0)	1
(5,1)	1
(4,2)	1
(۵,-1)	-1
(۸,۰)	-1
(۱۰,۴)	-1

تابع لاگرانژ: $L(x,\lambda,y) = f(x,y) - \lambda g(x,y)$

$$\nabla_{x,y,\lambda} L(x, y, \underset{\alpha}{\lambda}) = 0$$

مشتق لاگرانژ (گرادیان) که از x و y و λ برابر صفر، ضرایب را بدست می‌آید.

$$\frac{\partial L}{\partial x} = 0, \quad \frac{\partial L}{\partial y} = 0, \quad \frac{\partial L}{\partial \lambda} = 0$$

$$\frac{d}{dx}\left(f(x) + \alpha\, g(x)\right)\Big|_{x=x_0} = 0$$

$$g(x) = 0$$

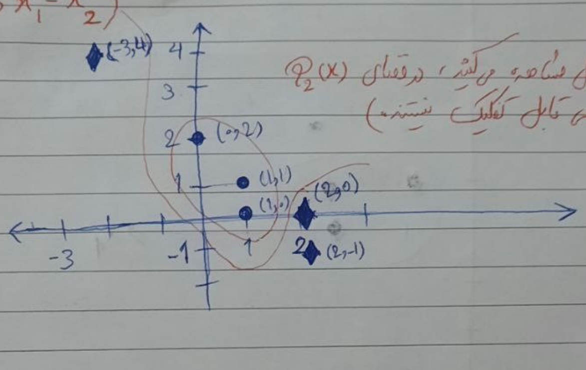
ج) نقطه A بسیار نزدیک به مرز جداکننده قرار دارد و آن را به یک بردار پشتیبان تبدیل می‌کند

اما نقطه B مستقیماً در حاشیه نیست و احتمالاً با توجه به پارامترهای svm و مدل خاص می‌تواند یک بردار پشتیبان باشد.

P4PCO STATIONERY

$$\varphi_2(x) = (x_2, x_1 - x_2)$$

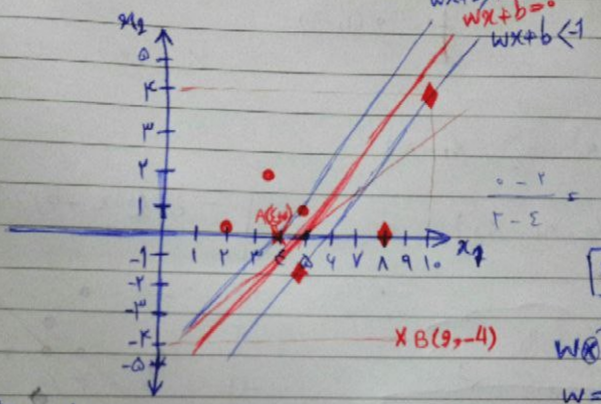
D	L
(1,0)	1
(1,1)	1
(0,2)	1
(2,-1)	-1
(2,0)	-1
(-3,4)	-1



(همانگونه که در شکل مشاهده می‌کنید، در فضای $\varphi_2(x)$ داده‌ها به صورت خطی قابل تفکیک نیستند.)

$(-5,4)\,|\,-1$

$(x_1, x_2) = (5, -1)$



$W = \frac{-1-4}{5-10} = 1$

$\frac{0-2}{4-2} =$ [redacted]

$\boxed{x_2 = x_1 + b}$

$W x + b = 0 \rightarrow (5,-1)$

$w = 1$

[redacted] $\rightarrow$ $x_1 - x_2 + 5 = -1$

[redacted] $-1 - 5 + 5 = -1$

$x_1 - x_2 = (x_1^2 + x_2^2) + b$

$\xrightarrow{(5,-1)} \quad 5-(-1) = (5^2 + (-1)^2) + b \Rightarrow$ [redacted]

[redacted]