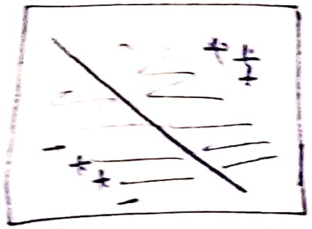
سوال ۱

الف ۱

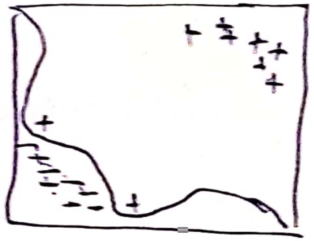
اگر $C \to 0$ باشد اندازه مارجین برای ما خیلی مهم است

اگر $C \to \infty$ باشد تفکیک پذیری مهم‌تر است

$C=0$



$C=\infty$

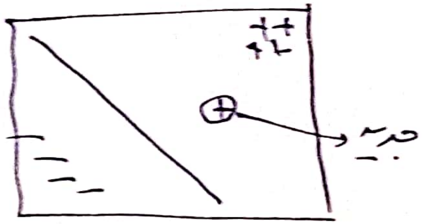


$C=0$ بهتر است زیرا منعطف است

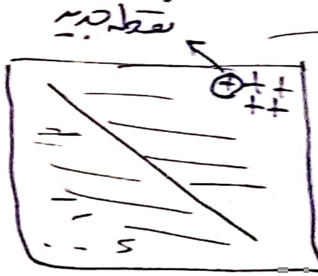
دو نقطه به صورت اشتباه دسته بندی شده اند و نباید generalization را خراب کند.

الف ۲

با تاثیر



بدون تاثیر



اگر داده در خارج Margin باشد تاثیری ندارد و اگر در داخل مارجین باشد تاثیر دارد.
(مرز تصمیم تغییر)

ب

$$K_1(a,b) = \left([a_1\ a_2]\begin{bmatrix}b_1\\ b_2\end{bmatrix}\right)^2 = (a_1b_1+a_2b_2)^2 = (a_1b_1)^2+(a_2b_2)^2+2a_1b_1a_2b_2 \quad (1)$$

$$= [a_1^2\ a_2^2\ \sqrt{2}a_1a_2]\begin{bmatrix}b_1^2\\ b_2^2\\ \sqrt{2}b_1b_2\end{bmatrix} \to \varphi = \begin{pmatrix}x_1^2\\ \sqrt{2}x_1x_2\\ x_2^2\end{pmatrix}$$

$$= \varphi(a)\,\varphi(B)$$

سؤال ۱ ب)

(۲)

$$K(x,y) = (a_1 b_1 + a_2 b_2 + 1)^2 = a_1^2 b_1^2 + a_2^2 b_2^2 + 2a_1 b_1 a_2 b_2 +$$

$$2a_1 b_1 1 + 2a_2 b_2 1 + 1^2$$

$$\Rightarrow \varphi(x) = (a_2^2, a_1^2, \sqrt{2}\, a_2 a_1, \sqrt{2}\, a_1, \sqrt{2}\, a_2, 1)$$

۲) الف) ~~محاسبه~~ استدلال توابع گوسی راحت است.

مشتق توابع گوسی راحت است.

توزیع شرطی و جوینت و بقیه توزیع های مربوطه همگی به سادگی گوسی می شوند

~~اگر توزیع جوینت را داشته باشیم با جایگزینی...~~

روابط ریاضی برای مثال: اسلاید ۲۲ جلسه ۲۲ مربوط به GP ، اسلاید ۱۲ مربوط به GP ، اسلاید ۱۱

تمامی توزیع ها نرمال هستند و فقط میانگین و واریانس تغییر می کند.

(۲ (ب ~~الگو~~ طبق تعریف مسئله داریم

$$X \in R^{n\times d} \Rightarrow \begin{cases} X^TX \in R^{n\times n} \\ XX^T \in R^{D\times D} \end{cases} \qquad \underline{D \gg n}$$

محاسبه X^TX خیلی منطقی‌تر خواهد بود!

ماتریس کوواریانس X به صورت های زیر خواهد بود

$$C = \frac{X^TX}{n-1}, \quad C' = \frac{XX^T}{n-1}$$

حال اگر PCA به روش SVD محاسبه شود می‌توانیم بنویسیم:

تجزیه SVD

$$X = ASB^T$$

ماتریس یکانی

ماتریس قطری

پس ماتریس کوواریانس می‌شود

$\boxed{S = S^T}$ می‌دانیم

$$C = \frac{X^TX}{n-1} = \frac{ASB^TBS^TA^T}{n-1} = A\frac{S^2}{n-1}A^T$$

پس مولفه‌های اصلی ماتریس ما، ستون‌های ماتریس AS می‌شود و مرتبه زمانی هم، چون خیلی مقدار ابعاد بیشتر است

از مرتبه $O(n^3)$ خواهد بود.

درست

۲) ب-۲) ۱) بله: ~~[illegible]~~ بردار ویژه تنها به همان اندازه rotate می‌شود هر بردار ویژه نیز همان محورهای اصلی است
که تنها با همان می‌چرخد.

۲) ب-۲) ۳) : درست: چون ~~[illegible]~~ واریانس نسبت به جمع تقسیم نمی‌کند ثابت است.

ثابت

۲) ب-۲) ۲) درست زیرا در هر بار یک سری از ویژگی‌ها را می‌توان کم کرد و مشکلی ندارد.

$\boxed{\text{Leave one out}}$

نقاط ما $\boxed{1}\ (-2,0)$ ، $\boxed{2}\ (2,0)$ ، $\boxed{3}\ (v,1)$ است.

ب) ۲

تابع ثابت : $\boxed{h(x)=b}$

$\boxed{1,2}$: $h(x)=\frac{0+0}{2}=0 \xrightarrow{\text{خطا}} (0-1)^2=1$

$\boxed{1,3}$: $h(x)=\frac{0+1}{2}=\frac{1}{2} \xrightarrow{\text{خطا}} (\frac{1}{2}-1)^2=\frac{1}{4}$

$\boxed{2,3}$: $h(x)=\frac{0+1}{2}=\frac{1}{2} \xrightarrow{\text{خطا}} (\frac{1}{2}-1)^2=\frac{1}{4}$

$$\frac{1+\frac{1}{4}+\frac{1}{4}}{3}=\frac{1.5}{3}=\frac{3}{6}=\%50$$

حال خطای ما در حالت دوم

$h(x)=a^Tx+b \Rightarrow h(x)=ax+b \xrightarrow{\text{چون ابعادی است ۱}}$

$\boxed{1},\boxed{2}$: $a=\text{شیب}=\frac{0-0}{2-(-2)}=0,\ b=0 \Rightarrow h(x)=0 \Rightarrow^{\text{خطا}} (0-1)^2=1$

h_1 $\boxed{1},\boxed{3}$: $a=\frac{1-0}{v-(-2)}=\frac{1}{v+2},\ y-0=\frac{1}{v+2}(x-(-2)) \Rightarrow y=\frac{x}{v+2}+\frac{2}{v+2} \Rightarrow \boxed{b=\frac{2}{v+2}}$

h_2 $\boxed{2,3}$: $\text{شیب}=\frac{1-0}{v-2} \Rightarrow y-0=\frac{1}{v-2}(x-2) \Rightarrow y=\frac{x}{v-2}-\boxed{\frac{2}{v-2}} \to b$

$h_1(x)=h_1(+2)=\frac{4}{v+2}$, $h_2(x)=h_2(-2)=\frac{-4}{v-2}$

خطا ↓ $\left(\frac{4}{v+2}\right)^2$

خطا ↓ $\left(\frac{-4}{\cancel{v+2}}\right)^2$

ت 2

$$\frac{۱ + \frac{4^2}{(v+2)^2} + \frac{(+4)^2}{(v-2)^2}}{3} \Rightarrow \frac{1}{2} \Rightarrow ۱ + \frac{16}{v^2+4v+4} + \frac{16}{v^2-4v+4} = \frac{3}{2}$$

تساوی
حل

$$\Longrightarrow v = \pm ۸/۴۷$$

۳) می‌دانیم

$$L_D(h_S) = \underset{(x,y)\sim D}{E}\left[\mathbb{I}(h_S(x)\neq y)\right]$$

الف) می‌توان بدست آورد که $E_S[L_D(h_S)]$ احتمال نمونه‌برداری از داده آموزشی S و یک نمونه (x,y) است به طوری که برچسب $\pi_1(x)$ متفاوت از y باشد.

می‌توانیم ابتدا از m نمونه آموزشی که برچسب ندارند نمونه بگیریم ($S_x = \{x_1, \dots, x_m\}$)

سپس می‌توانیم از داده بدون برچسب $x \sim D_x$ استفاده کنیم و $\pi_1(x)$ نزدیک‌ترین همسایه x در S_x خواهد بود و نهایتاً می‌توانیم از y به صورت $y \sim \eta(x)$ و از $y_{\pi_1(x)}$ به صورت $y_{\pi_1(x)} \sim \eta(\pi_1(x))$ نمونه‌برداری کنیم.

پس داریم

$$E_S[L_D(h_S)] = \underset{S_x\sim D_x^m,\, y\sim\eta(x),\, x\sim D_x,\, y'\sim\eta(\pi_1(x))}{E}\left[\mathbb{I}(y\neq y')\right] =$$

$$\underset{S_x\sim D_x^m,\, x\sim D_x}{E}\left[\underset{y\sim\eta(x),\, y'\sim\eta(\pi_1(x))}{\mathbb{P}}[y\neq y']\right]$$

ب)

$$\underset{y\sim\eta(x),\, y'\sim\eta(\pi_1(x))}{P}[y\neq y'] = \eta(x')(1-\eta(x)) + (1-\eta(x'))\eta(x)$$

$$= \eta(x')(1-\eta(x)) + (1-\eta(x)+\eta(x)-\eta(x'))\eta(x)$$

$$= (\eta(x)-\eta(x)+\eta(x'))(1-\eta(x)) + \dots$$

$$= 2\eta(x)(1-\eta(x)) + (\eta(x)-\eta(x'))(2\eta(x)-1)$$

۳ ج) با داشتن $|2\eta(x)-1| \leqslant 1$ و فرض این که η در واقع c-lipschitz است می‌توان فهمید

$$\underset{y\sim\eta(x),\,y'\sim\eta(x')}{\mathbb{P}}[y \neq y'] \leqslant 2\eta(x)(1-\eta(x)) + c\|x-x'\|$$

با جایگذاری در رابطه ب داریم

$$E_S[L_D(h_S)] \leqslant E_x[2\eta(x)(1-\eta(x))] + c\underset{S,x}{E}[\|x-x_{\pi_1(x)}\|]$$

سپس خطای دسته‌بند بهینه بیز برابر خواهد بود با:

$$L_D(h^*) = E_x[\min(\eta(x), 1-\eta(x))] \geqslant E_x[\eta(x)(1-\eta(x))]$$

با ترکیب دو رابطه آخر رابطه نهایی اثبات می‌شود.

ت) این رابطه می‌گوید که اگر اول توزیع تولیدکننده داده را ثابت درنظر بگیریم و $m \to \infty$ برود می‌توان فهمید که خطای 1NN (نزدیک‌ترین ۱-همسایه) به دو برابر خطای بیز همگرا خواهد شد.

4 (الف) علت این آن است که $M(r, x_*, \hat{\theta})$ مقدار کمینه خود را در نقطه $r=0$ می‌گیرد و فرض این که همیشه مشتق‌پذیر باشد باعث می‌شود که $\nabla_r M(r, x, \hat{\theta})$ برابر صفر باشد بنابراین بسط تیلور برابر خواهد بود با

$$D(r, x, \hat{\theta}) \approx \frac{1}{2} r^T H(x, \hat{\theta}) r$$

4 (ب) با استفاده از تقریب که در بالا بدست آمد r_* در واقع اولین بردار ویژه $u(x, \hat{\theta})$ از ماتریس Hessian $H(x, \hat{\theta})$ خواهد شد:

$$r_* \approx \arg\max \{ r^T H(x, \hat{\theta}) r, \ \|r\|_2 \le \epsilon \} = \epsilon \overline{u(x, \hat{\theta})}$$

که $\bar{v}$ بردار واحد هست که برابر مقدار بردار افزونه v است.

4 (ج) بردار v یک بردار واحد تصادفی برداری شده است، با شرط این که v عمود بر بردار ویژه u نیست. محاسبه تکراری $\overline{Hv_{n-1}}$ باعث می‌شود که به u کانورج کند.

4 (د) می‌دانیم

$$Hv \approx \frac{\nabla_r M(r, x, \hat{\theta})|_{r=\epsilon d} - \nabla_r M(r, x, \hat{\theta})|_{r=0}}{\epsilon} =$$

$$\frac{\nabla_r M(r, x, \hat{\theta})|_{r=\epsilon d}}{\epsilon} \xrightarrow{\nabla_r M(r, x, \hat{\theta})|_{r=0}} \quad v \leftarrow \overline{\nabla_r M(r, x, \hat{\theta})|_{r=\epsilon v}}$$

سوال ۵ — برای هر خانه، بهترین حالت این است که کمترین هزینه تا خانه پایانی را پیدا کند

پس داریم

برای خانه ۷ بهترین حالت این است که بالا برود

$$V^*(7) = -1 + 0.5\,V^*(9) = 9$$

برای خانه ۵ دو حالت داریم: پرش به خانه ۹ / ماندن در ۵

$$V(5) = -2 + 0.5 \times 20 = 8$$

$$V(5) = 3 + 0.5\,V(5) \Rightarrow V(5) = 6$$

پس برای خانه ۵ بهترین حالت پرش است پس $V^*(5)=9$

برای خانه ۹ بهترین کار ماندن است

$$10 + 0.5\,V^*(9) = V^*(9) \Rightarrow V^*(9) = 20$$

این خانه ها بدیهی بودند بقیه انقدر بدیهی نیستند:

$$Q(8, right) = -1 + 0.5\,V^*(7) = 3.5$$

$$Q(8, down) = -1 + 0.5\,V^*(5) = 3$$

بنابراین برای ۶ و ۸ داریم $\boxed{3.5 = V^*(6) = V^*(8)}$

برای بقیه خانه ها داریم:

$$Q^*(4, \text{jump to } 7) = -2 + 0.5 \times V^*(7) = 2.5$$

$$Q^*(4, right) = -1 + 0.5\,V^*(5) = 3$$

$$\max V^*(4) = 3$$

بنابراین $\boxed{V^*(4) = 3}$

$$Q^*(3, \text{jump to } 8) = -2 + 0.5\,V^*(8) = -0.25$$

$$\cancel{Q^*(3,\ )} = Q(3, right) = -1 + 0.5 \times 3 = 0.5$$

$$\boxed{V^*(3) = 0.5}$$

برای خانه ۲ داریم:

$$Q^*(2, \text{jump to } 5) = -2 + 0.5V^*(5) = 2$$
$$Q^*(2, \text{down}) = 0.5V^*(3) = -0.75$$
$$\xrightarrow{\max} \boxed{V^*(2) = 2}$$

برای خانه ۱ داریم:

$$Q^*(1, \text{jump to } 4) = -2 + 0.5V^*(4) = -0.5$$
$$Q^*(1, \text{down}) = -1 + 0.5V^*(2) = 0$$
$$\boxed{V^*(1) = 0}$$

حال برای π های بهینه داریم:

$\pi^*(1) = \text{down}$, $\pi^*(2) = \text{jump to } 5$, $\pi^*(3) = \text{right}$, $\pi^*(4) = \text{right}$,
$\pi^*(5) = \text{jump to } 9$, $\pi^*(6) = \text{up}$, $\pi^*(7) = \text{up}$, $\pi^*(8) = \text{right}$, $\pi^*(9) = \text{stay}$

ب) سیاست خانه ۹ همیشه ساکن خواهد بود زیرا:

$$V^*(9) = 10 + \gamma V^*(9) \Rightarrow V^*(9) = \frac{10}{1-\gamma}$$

برای خانه ۵ داریم:

$$V^1(5) = 3 + \gamma V^1(5) \Rightarrow V^1(5) = \frac{3}{1-\gamma}$$
$$V^2(5) = -2 + \gamma V^*(9) = -2 + \frac{10\gamma}{1-\gamma}$$

بنابراین داریم

$$V^1(5) > V^2(5) \Rightarrow \frac{3}{1-\gamma} > -2 + \frac{10\gamma}{1-\gamma} \Rightarrow \frac{3-10\gamma}{1-\gamma} > -2$$
$$\Rightarrow 3 - 10\gamma > -2(1-\gamma) \Rightarrow \boxed{\gamma < \frac{5}{12}}$$