جواب سوالات تئوری تمرین دوم ماشین لرنینگ

Year. Month. Date. ( )

پاسخ سؤالات HW2

سؤال ۱)

$$Var(\hat{p}(x)) = var\left(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sigma_n}\phi\left(\frac{x_i-x}{\sigma_n}\right)\right)$$

$$\overset{*}{=} \frac{1}{N^2}\sum_{i=1}^{N} var\left(\frac{1}{\sigma_n}\phi\left(\frac{\overset{i.i.d}{\widetilde{x_i}}-x}{\sigma}\right)\right)$$

*Note: $Var(x_1+x_2) \overset{x_i \text{ is iid}}{=\!=} var(x_1)+var(x_2)$

x_i are i.i.d

$$= \frac{1}{N} var\left(\frac{1}{\sigma_n}\phi\left(\frac{x_i-x}{\sigma}\right)\right) =$$

$$\overset{**}{=} \frac{1}{N} E\left[\frac{1}{\sigma_n^2}\phi^2\left(\frac{x_i-x}{\sigma}\right)\right] - \frac{1}{N}\underbrace{E^2\left[\frac{1}{\sigma_n}\phi\left(\frac{x_i-x}{\sigma}\right)\right]}_{\bar{p}_n^2(x)}$$

**Note: $var(x) = E[x^2] - E^2[x]$

$$= \frac{1}{N}\int \underbrace{\frac{1}{\sigma_n^2}\phi^2\left(\frac{x_i-x}{\sigma}\right)}_{sup(\phi)\phi\left(\frac{x_i-x}{\sigma}\right)} p(x_i)\,dx_i - \underbrace{\frac{1}{N}\bar{p}_n^2(x)}_{\geq 0 \text{ (always positive)}}$$

$$\leq \frac{1}{N}\frac{sup(\phi)}{\sigma}\int \underbrace{\underbrace{\frac{1}{\sigma_n}\phi\left(\frac{x-x_i}{\sigma_n}\right)}_{\delta(x-x_i)} p(x_i)\,dx_i}_{p(x)} = \frac{sup(\phi)\,E[\hat{p}(x)]}{N\sigma_n}$$

تابع، این عبارت، $n \to \infty$ به ازای ← $\delta(x-x_i)$

ضربه حول نقطه x_i میل می‌کند

$p(x)$ ← Convolution هر تابعی با تابع ضربه

همان خود اون تابع

PAPCO

Year. Month. Date. ( )

سؤال 2) ۱

میانگین تخمین parzen-window از رابطه زیر حاصل میشود:

$$\bar{p}(x) = \frac{1}{v_n} \int p(v)\, \phi\left(\frac{x-v}{h_n}\right) d_v$$

where:

$p(v) = U(0, a)$

برای محاسبه‌ی انتگرال بالا دو شرط لازم است:

- $p(v) = U(0, a)$ is zero outside of $0 < v < a$
- $\phi\left(\frac{x-v}{h_n}\right)$ is zero outside $x - v > 0 \Rightarrow v < x$

انتگرال بالا را به صورت زیر حل می‌کنیم:

- when $x < 0$, v must be 0 too since $v < x$ & the integral is zero
- when $0 < x < a$, v can range from 0 to x:

$$\bar{p}(x) = \frac{1}{v_n} \int p(v)\, \phi\left(\frac{x-v}{h_n}\right) dv = \frac{1}{h_n} \int_0^x \frac{1}{a} \exp\left(\frac{v-x}{h_n}\right) d_v$$

$$= \frac{1}{a h_n} h_n \exp\left(\frac{v-x}{h_n}\right) \Bigg|_{v=0}^{v=x} = \frac{1}{a}\left(1 - \exp\left(\frac{-x}{h_n}\right)\right)$$

- when $x > a$, v is not affected by x & ranges from 0 to a

$$\bar{p}(x) = \frac{1}{v_n} \int p(v)\, \phi\left(\frac{x-v}{h_n}\right) d_v = \frac{1}{h_n} \int_a^a \frac{1}{a} \exp\left(\frac{v-x}{h_n}\right) d_v$$

$$= \frac{1}{a h_n} h_n \exp\left(\frac{v-x}{h_n}\right) \Bigg|_{v=0}^{v=a} = \frac{1}{a}\left(\exp\left(\frac{a-x}{h_n}\right) - \exp\left(\frac{-x}{h_n}\right)\right)$$

PAPCO

$$= \frac{1}{a}\left(\exp\left(\frac{a}{h_n}\right)-1\right)\exp\left(\frac{-x}{h_n}\right)$$

2) مقدار بایاس برابر

$$E(p(x)-\hat{p}(x)) = p(x)-\bar{p}(x)$$

برای محاسبه بایاس نسبی عبارت بالا را بر $p(x)$ تقسیم می‌کنیم:

$$bias(x) = \frac{|p(x)-\bar{p}(x)|}{p(x)} = \frac{\frac{1}{a}-\bar{p}(x)}{\frac{1}{a}} = 1-ap(x) = e^{-x/h_n}$$

بایاس به ازای $0<x<a$ کاهشی می‌باشد پس برای اینکه بایاس کمتر از 0.01 باشد در 1٪ بازه داریم:

$$bias\left(\frac{a}{100}\right) = 0.01 \Leftrightarrow \exp\left(-\frac{a}{100h_n}\right) = 0.01$$

حل رابطه بالا:

$$h_n = a/_{100 \ln 100}$$

سوال 3)

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{n}(y_i - x_i\beta)^2 + \lambda\|\beta\|_2^2$$

$$\Rightarrow \sum_{i=1}^{N}(y_i - x_i\beta)^T(y_i - x_i\beta) + \lambda\beta^T\beta$$

فرم ماتریسی $= (y-X\beta)^T(y-X\beta) + \lambda\beta^T\beta$

$$= y^Ty - y^TX\beta - \beta^TX^Ty + \beta^TX^TX\beta + \lambda\beta^T\beta$$

$$\frac{\partial}{\partial\beta} = 0 \quad -\cancel{2}X^Ty + \cancel{2}X^TX\beta + \cancel{2}\lambda\beta = 0$$

$$(X^TX+\lambda I)\beta = X^Ty \Rightarrow \hat{\beta} = (X^TX+\lambda I)^{-1}X^Ty$$

Subject:
Year. Month. Date. ( )

سوال 3-2) تفاوت‌ها

1) بارزترین تفاوت نرم L1 و نرم L2 regulator است. نرم 1 مجموع اندازه وزن‌ها، نرم 2 مجموع مربع وزن‌ها است. این بدان معنی است که L_1 ضرایب بی‌اهمیت را به صفر نزدیک کرده و ستون ضرایب برابر صفر باشد. بدین معنا اگر تعداد زیادی از وزن‌ها بالا نباشد و ضرایب sparse باشد. اما در L_2 ضرایب به صفر نزدیک شده ولی صفر نمی‌شوند.

2) در نرم L_1 عددهای با واریانس کم بیشتر کوچک شده اما در نرم L_2 عدد ابعاد خارج از محدوده واریانس، آهسته‌تر کوچک می‌شوند.

3) برای L_2 سوال حل بسته داشته ولی L_1 حل بسته ندارد.

4) L_1 برای داده‌های پراکنده و sparse مناسب است اما L_2 برای داده‌های متراکم.

5) از L_1 می‌توان برای Feature selection استفاده کرد.

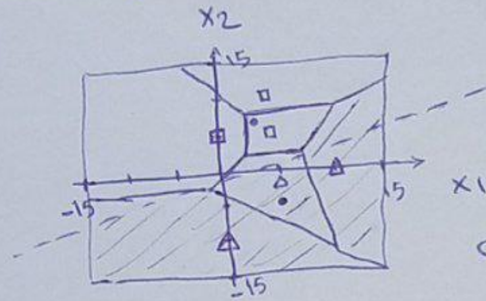
سوال ۴ :

۱) w_2 , w_1

$w_1 \to \Delta$

$w_2 \to \square$

mean $\to \bullet$



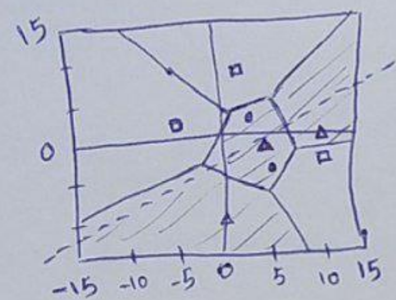
خط چین جدا کننده براساس میانگین است.

۲) w_3 , w_1

$w_1 \to \Delta$

$w_3 \to \square$

mean $\to \bullet$

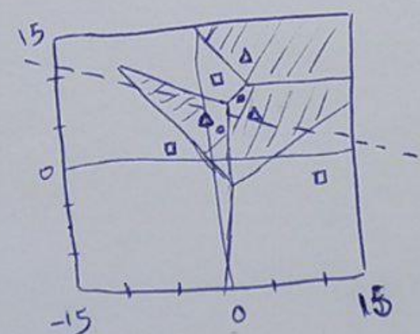


۳) w_3 , w_2

$w_2 \to \Delta$

$w_3 \to \square$

mean $\to \bullet$



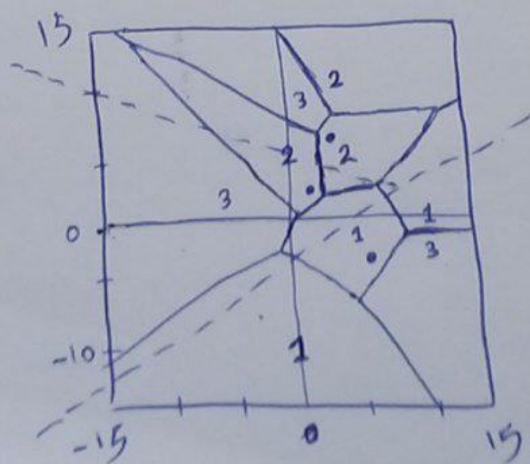
۴) w_3 , w_2 , w_1

$w_1 \to 1$

$w_2 \to 2$

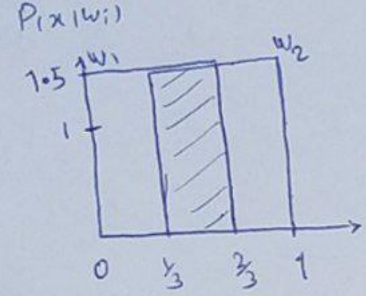
$w_3 \to 3$

mean $\to \bullet$



(الف)

سوال ۵) فرض می‌کنیم $P(w_1)=P(w_2)=\frac{1}{2}$ و توزیع ها به شکل زیر هستند



$$P=\int_0^1 \min\left[P(w_1)p(x|w_1), P(w_2)p(x|w_2)\right]dx$$

$$=\int_{1/3}^{2/3} P(w_1)p(x|w_1)\,dx$$

برای بازه‌ی $\frac{1}{3}$ تا $\frac{2}{3}$ فرض شده که مرز تصمیم در $x^*=\frac{1}{3}$ قرار می‌گیرد.

$$=0.5\left(\frac{2}{3}-\frac{1}{3}\right)\frac{3}{2}=0.25$$

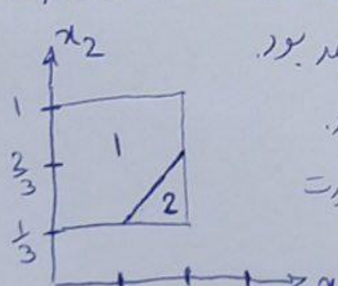
ب) با توجه به محل x_1 و x_2 انتخاب شده دو حالت $x_1\geq x_2$ و $x_2\geq x_1$ را باید بررسی کنیم

حالت اول $x_2\geq x_1$ و $\frac{1}{3}<\frac{x_1+x_2}{2}<\frac{2}{3}$: در این حالت x^* بین $\frac{1}{3}$ تا $\frac{2}{3}$ می‌افتد

که این دقیقا شبیه نرخ خطای حالت Bayes در قسمت اول سوال است پس مقدار آن 0.25 است.

حالت دوم $x_1\geq x_2$ ، $\frac{1}{3}<\frac{x_1+x_2}{2}<\frac{2}{3}$: در این حالت x^* بین $\frac{1}{3}$ تا $\frac{2}{3}$ است

اما با کمی بررسی به این نتیجه می‌رسیم که خطا در این حالت کاملا متضاد حالت Bayes در قسمت قبل است و نرخ خطا $1-0.25=0.75$ خواهد بود.



همچنین باید میزان احتمال رخداد حالت اول و دوم را هم بدست آورد.
شکل روبرو حالت هایی که $x_1>x_2$ و برعکس را نشان می‌دهد که در اینصورت
حالت اول با احتمال $\frac{7}{8}$، حالت دوم با احتمال $\frac{1}{8}$ رخ می‌دهد

پس خطای کل به صورت زیر است:

$$P_1=\frac{7}{8}\cdot 0.25+\frac{1}{8}\cdot 0.75=0.3125$$