Subject: 810198358 امیرمهدی افشاری پور

Year. Month. Date. ( )

1) الف:
$$\varepsilon_1^2 = \sum_i (y_i - \beta_0)^2 \xrightarrow{\partial/\partial\beta = 0} \sum_i -2(y_i - \beta_0) = 0$$

$$\Rightarrow \sum_i y_i = n\beta_0 \Rightarrow \boxed{\beta_0 = \frac{\sum_i y_i}{n} = \bar{y}} = 59.6$$

1) ب:
$$\varepsilon_1^2 = \sum_i (y_i - \beta_1 x_i)^2 \xrightarrow{\partial/\partial\beta_1 = 0} \sum_i -2x_i(y_i - \beta_1 x_i) = 0$$

$$\Rightarrow \sum_i x_i y_i = \sum_i \beta_1 x_i \Rightarrow \overline{xy} = \beta_1 \bar{x} \Rightarrow \boxed{\beta_1 = \frac{\overline{xy}}{\bar{x}}} = \frac{9774}{146} = 66.94$$

1) ج: معادله اول بیانگر رگرسیون خطی است که رابطه متغیر X و میانگین متغیر Y را نشان میدهد.
اما معادله دوم بیانگر رابطه میانگین متغیر X و متغیر Y است. تفاوتی که در معادله اول صدق میکند
لزوما در معادله دوم صدق نمی‌کند (بدلیل نبود error term). در معادله اول، پارامترهای β_0 و β_1
بصورت مستقیم از داده‌ها بدست می‌آیند، اما در معادله دوم β_0 و β_1 ناشناخته اند

1) د:
$$\sigma^2 = \frac{1}{n-2} \sum_i (\hat{y} - y_i)^2$$

$$y = \beta_0 + \beta_1 x + \varepsilon_1 \Rightarrow \varepsilon_1^2 = \sum_i (y_i - \beta_0 - \beta_1 x_i)^2$$

$$\frac{\partial}{\partial\beta_0} = 0 \Rightarrow \sum_i -2(y_i - \beta_0 - \beta_1 x_i) = 0 \Rightarrow \sum_i y_i - n\beta_0 - \beta_1 \sum_i x_i = 0$$

$$\Rightarrow n\bar{y} - n\beta_0 - n\beta_1\bar{x} = 0 \Rightarrow \bar{y} = \beta_0 + \beta_1 \bar{x} \quad \text{I}$$

$$\frac{\partial}{\partial\beta_1} = 0 \Rightarrow \sum_i -\cancel{2} x_i (y_i - \beta_0 - \beta_1 x_i) = 0 \Rightarrow$$

$$\sum_i x_i y_i - \beta_0 \sum_i x_i - \beta_1 \sum_i x_i^2 = 0 \Rightarrow \overline{xy} - \beta_0 \bar{x} - \beta_1 \overline{x^2} = 0 \quad \text{II}$$

Subject:
Year. Month. Date. ( )

با استفاده از (I) و (II) داریم: $(\bar{y} = \beta_0 + \beta_1 \bar{x}) \times \bar{x} = \beta_0 \bar{x} = \bar{x}\bar{y} - \beta_1 (\bar{x})^2$

$$\overline{xy} - \beta_0 \bar{x} - \beta_1 \overline{x^2} = 0$$

$$\Rightarrow \overline{xy} - \bar{x}\bar{y} + \beta_1 (\bar{x})^2 - \beta_1 \overline{x^2} = 0 \Rightarrow \beta_1 = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - (\bar{x})^2}$$

$$\Rightarrow \beta_1 = \frac{977.4 - 870.16}{249.8 - 213.16} = \frac{107.24}{36.64} = 2.92$$

$$\Rightarrow \beta_0 = \bar{y} - \beta_1 \bar{x} = 59.6 - 2.92(14.6) = 16.96$$

$$\Rightarrow \hat{y} = 2.92x + 16.96$$

x:	5	11	18	15	21	6	17
$\hat{y}$:	31.56	49.8	69.52	60.76	78.28	34.48	66.6
$\lvert\hat{y}-y\rvert$:	2.44	2.8	3.52	8.76	1.72	0.52	0.6

x:	10	24	19
$\hat{y}$:	46.16	87.04	72.44
$\lvert\hat{y}-y\rvert$:	1.84	0.04	8.56

$$\Rightarrow \sigma^2 = \frac{1}{8}(183.1712) = 22.89$$

1) د: بطور دقیق نمی‌توان پیش بینی کرد زیرا ورودی جدید در واقع یک توزیع نرمال با میانگین $\hat{y}$ خواهد بود.

PAPCO

۲) الف: هر دوی L_1_regulization و L_2_regulization برای جریمه کردن وزن ها در گرادیان به کار می روند. به عبارت دیگر loss_function در این حالت برابر خواهد بود با:

$$\sum_{i=1}^{n} (y_i - w_i x_i)^2 + \lambda \sum_i |w_i| \quad (\text{for } L1)$$

$$\sum_{i=1}^{n} (y_i - w_i x_i)^2 + \lambda \sum_i w_i^2 \quad (\text{for } L2)$$

همانطور که دیده می شود، در L1، مجموع قدر مطلق وزن ها لحاظ می شود، در حالیکه L2 مجموع مربعات را شامل می شود، در نتیجه L1 منتهی به مدل sparse می شود.

۲) ب:

$$\hat{\beta} = \frac{\partial}{\partial \beta}\left[\|A\beta - Y\|_2^2 + \lambda \|\beta\|_2^2 \right]$$

where : $A = \begin{bmatrix} \longleftarrow a_1^T \longrightarrow \\ \vdots \\ \longleftarrow a_n^T \longrightarrow \end{bmatrix}_{n\times d}$ && $\beta = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_d \end{bmatrix}_{d\times 1}$

$$\Rightarrow \hat{\beta} = \frac{\partial}{\partial \beta}\left[(A\beta - Y)^T (A\beta - Y) + \lambda \beta^T \beta \right] =$$

$$\frac{\partial}{\partial \beta}\left[(\beta^T A^T - Y^T)(A\beta - Y) + \lambda \beta^T \beta \right] =$$

$$\frac{\partial}{\partial \beta}\left[\beta^T A^T A \beta - 2Y^T A \beta + Y^T Y + \lambda \beta^T \beta \right] =$$

$$2A^T A \beta - 2A^T Y + 2\lambda \beta = 0$$

$$\Rightarrow (A^T A + \lambda I)\beta = A^T Y \Rightarrow \beta = (A^T A + \lambda I)^{-1} A^T Y$$

۳) الف: اگر w_{i0} را هم داخل بردار وزن ببریم:

$$X = \begin{bmatrix} 1 \\ x_1 \\ \vdots \\ x_d \end{bmatrix} \quad w_i = \begin{bmatrix} w_{i0} \\ \vdots \\ w_{id} \end{bmatrix}$$

$$P(Y=y_i|X) \propto \exp(w_i^T X) \Rightarrow P(Y=y_i|X) = \beta \exp(w_i^T X)$$

هدف یافتن β است. از آنجا که جمع تمامی احتمال‌ها 1 میشود:

$$\sum_{i=1}^{K} P(Y=y_i|X) = 1 \Rightarrow \beta \left[\sum_{i=1}^{K-1} \exp(w_i^T X) \right] + P(Y=y_K|X) = 1$$

$$\Rightarrow \beta = \frac{1 - P(Y=y_K|X)}{\sum_{i=1}^{K-1} \exp(w_i^T X)}$$

برای یافتن $P(Y=y_K|X)$، میتوانیم از logistic regression استفاده کنیم بنحوی که

$K-1$ کلاس را + و کلاس Kام را - در نظر بگیریم:

$$\underbrace{\sum_{i=1}^{K-1} P(Y=y_i|X)}_{+\text{کلاس}} = \frac{1}{1 + \sum_{i=1}^{K-1} \exp(w_i^T X)} \rightarrow$$

$$\underbrace{P(Y=y_K|X)}_{-\text{کلاس}} = \frac{\sum_{i=1}^{K-1} \exp(w_i^T X)}{1 + \sum_{i=1}^{K-1} \exp(w_i^T X)}$$

$$\Rightarrow \beta = \frac{1 - P(Y=y_K|X)}{\sum_{i=1}^{K-1} \exp(w_i^T X)} = \frac{1}{1 + \sum_{i=1}^{K-1} \exp(w_i^T X)}$$

حال که β بدست آمد، می‌توان نوشت:

$$P(Y=y_i|X) = \frac{\exp(w_i^T X)}{1 + \sum_{i=1}^{K-1} \exp(w_i^T X)} \quad 1 \leq i \leq K-1$$

Subject:
Year. Month. Date. ( )

3) ب: با محاسبه $P(Y=y_i \mid X)$ به ازای تمام iها، آن کلاسی که مقدار احتمال را بیشینه می‌کند انتخاب می‌کنیم:

$$\hat{i} = \underset{i}{\operatorname{argmax}}\ P(Y=y_i \mid X)$$

5) ویژگی های Distance metric:

I) $D(a,b) \geq 0$

II) $D(a,b) = 0 \iff a=b$

III) $D(a,b) = D(b,a)$

IV) $D(a,b) + D(b,c) \geq D(a,c)$

با ضرب شدن تمامی feature ها در عدد β داریم:

$$D_{new}(a,b) = \sqrt{\beta^2 \sum_{k=1}^{d} (a_k - b_k)^2} = |\beta| \sqrt{\sum_{k=1}^{d} (a_k - b_k)^2}$$

حال داریم:

$D_{new}(a,b) \geq 0$ I

$D_{new}(a,b) = 0 \Rightarrow |\beta| \sqrt{\sum_{k=1}^{d} (a_k - b_k)^2}$ && $\beta \neq 0 \Rightarrow$

$\sum_{k=1}^{d} (a_k - b_k)^2 = 0 \Rightarrow a_k = b_k \ \forall k \Rightarrow a=b$ طرف اول

$a=b \Rightarrow D_{new}(a,b) = 0$ طرف دوم $\Rightarrow$ ویژگی II اثبات شد

$D_{new}(a,b) = D_{new}(b,a)$ طبق فرمول D_{new}، این ویژگی به وضوح درست است، در نتیجه III نیز اثبات شد.

Subject:
Year. Month. Date. ( )

برای ویژگی 4 نیز داریم : در هر یک فاصله اقلیدسی ، این ویژگی برقرار است.

بعبارت دیگر :

$$\sqrt{\sum_1 (a_k-b_k)^2} + \sqrt{\sum_1 (b_k-c_k)^2} \geq \sqrt{\sum_1 (a_k-c_k)^2}$$

حال اگر طرفین رابطه را در $|\beta|$ ضرب کنیم نیز همچنان رابطه برقرار است.

$$\underbrace{|\beta|\sqrt{\sum_1 (a_k-b_k)^2}}_{D_{new}(a,b)} + \underbrace{|\beta|\sqrt{\sum_1 (b_k-c_k)^2}}_{D_{new}(b,c)} \geq \underbrace{|\beta|\sqrt{\sum_1 (a_k-c_k)^2}}_{D_{new}(a,c)}$$

در نتیجه ویژگی IV نیز اثبات شد.

Subject:
Year. Month. Date. ( )

4) الف:

$$E[P_n(x)] = \frac{1}{nh_n}\left[\sum_{i=1}^{n} E\left(\phi\left(\frac{x-x_i}{h_n}\right)\right)\right] =$$

$$\frac{1}{h_n} E_{x_i}\left(\phi\left(\frac{x-x_i}{h_n}\right)\right) = \int_{-\infty}^{\infty} \frac{1}{h_n}\phi\left(\frac{x-x_i}{h}\right) P(x_i)\, dx_i$$

می‌دانیم:

$$P(x_i) = \frac{1}{\sqrt{2\pi}\,\sigma'} \exp\left\{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma'}\right)^2\right\}$$

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(x-0)^2\right\} \Rightarrow$$

$$\phi\left(\frac{x-x_i}{h}\right) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-x_i}{h}\right)^2\right\}$$

$$\rightarrow E[P_n(x)] = \int \frac{1}{2\pi\sigma' h_n} \exp\left\{-\frac{1}{2}\left(\left(\frac{x_i-\mu}{\sigma'}\right)^2 + \left(\frac{x-x_i}{h}\right)^2\right)\right\} dx_i$$

حال باید انتگرال را حساب کنیم:

$$e^{-\frac{1}{2}\left(\frac{x_i^2}{\sigma^2}\right)} \times e^{-\frac{1}{2}\cancel{\left(\frac{\mu^2}{\sigma^2}\right)}} \times e^{\frac{x_i\mu}{\sigma'^2}} \qquad e^{x_i\left(\frac{\mu}{\sigma}+\right)}$$

$$e^{-\cancel{\frac{1}{2}}\left(\frac{x^2}{h^2}\right)} \times e^{-\frac{1}{2}\left(\frac{x_i}{h}\right)^2} \times e^{\frac{x x_i}{h^2}}$$

$$\rightarrow E[P_n(x)] = \frac{1}{2\pi\sigma' h_n} \exp\left\{-\frac{1}{2}\left(\frac{\mu^2}{\sigma^2}\right) - \frac{1}{2}\left(\frac{x^2}{h_n^2}\right)\right\}$$

$$\times \int \exp\left\{-\frac{1}{2}x_i^2\left(\frac{1}{\sigma^2}+\frac{1}{h^2}\right) + x_i\left(\frac{x}{h^2}+\frac{\mu}{\sigma^2}\right)\right\} dx_i$$

برای آنکه انتگرال بالا بصورت مربع کامل دربیاید باید تغییر متغیر انجام دهیم:

$\upsilon^2 . \frac{\sigma^2 h^2}{\sigma^2+h^2}$ $\quad \frac{\sigma h}{\sqrt{\sigma^2+h^2}}$

$$\left.\begin{aligned} \upsilon^2 &= \frac{1}{\frac{1}{h_n^2}+\frac{1}{\sigma^2}} \\ u &= \upsilon^2\left(\frac{x}{h_n^2}+\frac{\mu}{\sigma^2}\right) \end{aligned}\right\} \Rightarrow E[P_n(x)] =$$

$$\frac{1}{2\pi\sigma h_n} * \exp\left\{-\frac{1}{2}\left(\frac{\mu^2}{\sigma^2}+\frac{x^2}{h_n^2}-\frac{u^2}{\upsilon^2}\right)\right\} \cdot \underbrace{\int \exp\left(-\frac{1}{2}\left(\frac{x_i-u}{\upsilon}\right)^2\right) dx_i}_{\text{فرمت توزیع گوسی}}$$

$$m = \frac{x_i-u}{\upsilon} \Rightarrow dx_i = \upsilon\, dm = \upsilon \underbrace{\int \exp\left(-\frac{m^2}{2}\right) dm}_{\sqrt{2\pi}}$$

$$\Rightarrow E[P_n(x)] = \frac{\sqrt{2\pi}\,\cancel{\upsilon}}{2\cancel{\pi}\,\sigma h_n} \exp\left\{-\frac{1}{2}\left(\frac{\mu^2}{\sigma^2}+\frac{x^2}{h_n^2}-\frac{u^2}{\upsilon^2}\right)\right\}$$

$$\frac{1}{\sqrt{2\pi}\times\sqrt{\sigma^2+h^2}}$$

$$\frac{-u^2}{\upsilon^2} = -\upsilon^2\left(\frac{x}{h_n^2}+\frac{\mu}{\sigma^2}\right)^2 = \frac{-\sigma^2 h_n^2}{\sigma^2+h_n^2}\left(\frac{x\sigma^2+\mu h_n^2}{h_n^2\sigma^2}\right)^2$$

$$= \frac{-(x\sigma^2+\mu h_n^2)^2}{h_n^2\sigma^2(\sigma^2+h_n^2)}$$

$$\Rightarrow \exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma^2 h_n^2}\left(x^2(h_n^2+\sigma^2) - \frac{(x\sigma^2+\mu h^2)^2}{\sigma^2+h_n^2}\right)\right)\right\}$$

$$= \exp\left\{-\frac{1}{2}\left(\frac{1}{\cancel{\sigma^2 h_n^2}}\left(\frac{x\cancel{h_n\sigma}-\mu\cancel{h_n\sigma}}{h_n^2+\sigma^2}\right)^2\right)\right\} =$$

$$\exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{h_n^2+\sigma^2}\right)^2\right\}$$

Subject:
Year. Month. Date. ( )

پس در آخر داریم:

$$E[P_n(x)] = \overline{P_n(x)} = \frac{1}{\sqrt{2\pi}\sqrt{h_n^2+\sigma^2}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{h_n^2+\sigma^2}\right)^2\right\} = N(\mu, h_n^2+\sigma^2)$$

4) ب: در مسئله قبل نشان دادیم:

$$\overline{P_n(x)} = \frac{1}{\sqrt{2\pi}\sqrt{\sigma^2+h_n^2}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{h_n^2+\sigma^2}\right)^2\right\}$$

از طرفی $P(x)$ نیز برابر است با:

$$P(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}$$

حال با فاکتور گیری از $P(x)$ داریم:

$$P(x) - \overline{P_n(x)} = P(x)\left[1 - \frac{\sigma}{\sqrt{\sigma^2+h_n^2}} \exp\left\{-\frac{1}{2}\frac{(x-\mu)^2 h^2}{\sigma^2(h^2+\sigma^2)}\right\}\right]$$

با در نظر گرفتن اینکه h_n مقدار کوچکی است:

$$P(x) - \overline{P_n(x)} \simeq P(x) \cdot \frac{h_n}{2\sigma^2}\left[1 - \left(\frac{x-\mu}{\sigma}\right)^2\right]$$