hw2 - پویا شجاع الاسلامی ۸۶۱۰۰۲۹۴

①

$$f(x_i;\theta_0,\theta_1,\dots\theta_D)=\sum_{d=1}^{D}\theta_d\cdot x_i(d)+\theta_0$$

$$\left[y^Ty-(X\theta)^Ty-y^TX\theta+(X\theta)^TX\theta\right]$$

$$R(\theta)=\frac{1}{2N}\|y-X\theta\|_2^2+\theta^TH\theta+\theta^T\theta+\alpha^T\theta=\frac{1}{2N}(y-X\theta)^T(y-X\theta)+\theta^TH\theta+\theta^T\theta+\alpha^T\theta \quad \frac{\partial}{\partial\theta}=0 \Rightarrow$$

a: بردار ; H: ماتریس , $H=H^T$

$$\longrightarrow \frac{\partial R(\theta)}{\partial\theta}=\left[\frac{-2y^TX+2X^TX\theta}{2N}\right]+2H\theta+2\theta+\alpha^T=0 \Rightarrow$$

$$\left[2I+2H+\frac{X^T\cdot X}{N}\right]\theta-\frac{y^TX}{N}+\alpha^T=0 \Rightarrow \theta=\left(2I+2H+\frac{X^T\cdot X}{N}\right)^{-1}\cdot\left(\frac{y^TX}{N}-\alpha^T\right)$$

② الف)

تعداد نمونه (N) — تعداد features (K)

تابع هزینه در L_1:

$$\longrightarrow L(x,y)=\sum_{i=1}^{N}(y_i-\hat{y}_i)^2+\lambda\sum_{1}^{K}|\theta_i|$$

تابع هزینه در L_2:

$$\longrightarrow L(x,y)=\sum_{i=1}^{N}(y_i-\hat{y}_i)^2+\lambda\sum_{i=1}^{K}\theta_i^2$$

① در هر دو تابع باید بخش error را حداقل کنیم اما در L_1 علاوه بر این باید مقدار $|\theta_i|$ را نیز min کنیم و در برخی موارد برای اینکه از overfit جلوگیری کنیم باید ضرایب θ_i را صفر کنیم حال آنکه در L_2 باید مربع ضرایب θ_i $[\theta_i^2]$ را min کنیم در L_2 همچنین باید این ضرایب min می‌شوند اما صفر نه. به صفر نزدیک می‌شوند. به طور کلی از L_2 بیشتر استفاده می‌شود. [به طور کلی L_1 ضرایب را صفر می‌کند و L_2 ضرایب را کاهش می‌دهد] ② L_1 نسبت به L_2 (feature) کمتری دارد (چون ضرایب را صفر می‌کند.) ③ L_2 می‌تواند منجر به overfit و L_1 می‌تواند منجر به underfit گردد. ④ L_2 برای مسائل ... نتیجه بهتری مناسب است و L_1 برای مسائل ...

ب)

$$\hat{\beta}=\arg\min_{\beta}\sum_{1}^{N}(Y_i-X_i\beta)^2+\lambda\|\beta\|_2^2$$

$$\longrightarrow F=\sum(Y_i-X_i\beta)^T(Y_i-X_i\beta)+\lambda\beta^T\beta \Rightarrow$$

$$y_i^T\cdot y_i-\beta^Tx_i^T y_i-y_i^Tx_i\beta+(x_i\beta)^Tx_i\beta$$

$\lambda\geq 0$

$$\frac{\partial}{\partial\beta}=0$$

$X_i=[X_i^{(1)},\dots X_i^{(P)}]$

$Y=[Y_1,\dots Y_n]$ $\quad \overset{?}{\longrightarrow}\ \hat{\beta}=(A^TA+\lambda I)^{-1}\cdot A^T\cdot Y$

$A=[X_1,\dots X_n]$

$$\frac{\partial F}{\partial\beta}=\sum(2y_i^T\cdot x_i+2x_i^Tx_i\beta)+2\lambda\beta=0 \Rightarrow$$

$$\left(\sum x_i^T\cdot x_i+\lambda I\right)\beta=\sum y_i^Tx_i \Rightarrow (A^T\cdot A+\lambda I)\beta=A^TY \Rightarrow \beta=(A^T\cdot A+\lambda I)^{-1}A^TY$$

③

$$p(Y=y_k|X)\propto\exp\left(w_{k0}+\sum_{1}^{d}w_{ki}x_i\right)\quad(I)$$

for $k=1,\dots,K-1$

$$\longrightarrow p(y=y_k|X)=\frac{\exp(w_{k0}+\sum_{i=1}^{d}w_{ki}x_i)}{1+\sum_{1}^{K-1}(w_0+\sum_{1}^{d}w_{ki}x_i)}$$

الف) $p(Y=Y_K|X)$

$$(II)\quad p(Y=y_K|X)=\frac{1}{1+\sum_{k=1}^{K-1}\exp(w_0+\sum_{i=1}^{d}w_{ki}x_i)}$$

ب)

با اعمال $\sum$ روی I

$$p(Y=y_K|X)+\sum_{k=1}^{K-1}p(y=y_k|X)=1\quad(III)$$

(II) $\quad$ I

$$p(Y=Y_K|X)=1-\sum_{k=1}^{K-1}p(Y=y_k|X)$$

ب) $\hat{y}_k = \underset{k}{\arg\max}\, P(Y = y_k | X)$

رابطه‌ی طبقه‌بند

④ الف) $\beta_0, \beta_1, \sigma^2$?

ب) واریانس ?

ج) Correlation ?

$$\hat{y}_i = \beta_1 x_i + \beta_0 + e_i \longrightarrow \begin{cases} \beta_1 = \dfrac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \\ \beta_0 = \bar{y} - \beta_1 \bar{x} \end{cases}$$

$$\bar{y} = \frac{\sum y_i}{n} = \boxed{56.4}$$

$$\bar{x} = \frac{\sum x_i}{n} = \boxed{10}$$

برای بدست آوردن β_1, β_0 ، $\xrightarrow{\text{صورت } \beta_1} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = (4-10)(31-56.4) + \dots$

$= \boxed{1305}$

$\xrightarrow{\text{مخرج } \beta_1} \sum (x_i - \bar{x})^2 = 36 + 1 + 0 + \dots = \boxed{376} = \sigma_x^2$

$$\Longrightarrow \begin{cases} \boxed{\beta_1 \simeq 3.471} \\ \boxed{\beta_0 \simeq 21.69} \end{cases}$$

$$\sigma^2 = \frac{\sum (y_i - \hat{y}_i)^2}{n-2} = \frac{179.387}{8} \simeq 22.42$$

$$\beta_1 \longrightarrow \begin{cases} Var(\beta_1) = \dfrac{\sigma^2}{\sum (x_i - \bar{x})^2} \\ Var(\beta_1) = \dfrac{22.42}{376} \simeq \boxed{0.06} \end{cases}$$

$$\beta_0 \longrightarrow \begin{cases} Var(\beta_0) = \dfrac{\sigma^2}{n} + \dfrac{(\bar{x})^2 \sigma^2}{n \sigma_x^2} = \\ \dfrac{22.42}{10} + \dfrac{100 \times 22.42}{10 \times 376} \Longrightarrow \end{cases}$$

$$\boxed{Var(\beta_0) = 2.838}$$

$$Cov(\beta_0, \beta_1) = \frac{-\bar{x}\, \sigma^2}{n \sigma_x^2} = -\frac{22.42 \times 10}{10 \times 376} \simeq \boxed{-0.06}$$

$$\rho(\beta_0, \beta_1) = \frac{-0.06}{\sqrt{2.838 \times 0.06}} = \boxed{-0.15}$$