统计方法与机器学习理论三

Q1:

$\forall j \in \{1, 2, \cdots p\}$，将第j列与前j-1列交换，则有

$$T_j = \begin{pmatrix} & 1 & & & \\ & & \ddots & & \\ & & & 1 & \\ 1 & & & & \\ & & & & 1 & \\ & & & & & \ddots \\ & & & & & & 1 \end{pmatrix} \quad T_j^{-1} = T_j^T = \begin{pmatrix} & & & 1 & & & \\ 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}$$

令 $X_t = (X_j, \underbrace{X_1, \cdots, X_{j-1}, X_{j+1}, \cdots X_p}_{X_0})$

$$X_t = X_s T_j \Rightarrow X_s = X_t T_j^{-1} = X_t T_j^T$$

$$\therefore \quad C = (X_s^T X_s)^{-1} = \{c_{ij}\} = ((X_t T_j^T)^T (X_t T_j^T))^{-1}$$

$$= (T_j X_t^T X_t T_j^T)^{-1} = T_j (X_t^T X_t)^{-1} T_j^T$$

$$X_t = (X_j, X_0) \quad X_t^T X_t = \begin{pmatrix} X_j^T \\ X_0^T \end{pmatrix} (X_j, X_0) = \begin{pmatrix} X_j^T X_j & X_j^T X_0 \\ X_0^T X_j & X_0^T X_0 \end{pmatrix}$$

$$(X_t^T X_t)^{-1} = \begin{pmatrix} X_{11}^* & X_{12}^* \\ X_{21}^* & X_{22}^* \end{pmatrix}$$

$$X_{11}^* = (X_j^T X_j - X_j^T X_0 (X_0^T X_0)^{-1} X_0^T X_j)^{-1}$$

$$X_{12}^* = -(X_j^T X_j - X_j^T X_0 (X_0^T X_0)^{-1} X_0^T X_j)^{-1} X_j^T X_0 (X_0^T X_0)^{-1}$$

$$X_{21}^* = -(X_0^T X_0)^{-1} X_0^T X_j (X_j^T X_j - X_j^T X_0 (X_0^T X_0)^{-1} X_0^T X_j)^{-1}$$

$$X_{22}^* = (X_0^T X_0)^{-1} + (X_0^T X_0)^{-1} X_0^T X_j (X_j^T X_j - X_j^T X_0 (X_0^T X_0)^{-1} X_0^T X_j)^{-1} X_j^T X_0 (X_0^T X_0)^{-1}$$

X_S 是经标准化后的

$\therefore \sum_{i=1}^{n} x_{ij}^2 = 1 \Rightarrow x_j^T x_j = 1$

$$C_{jj} = x_{11}^* = (\underbrace{x_j^T x_j}_{1} - x_j^T X_0 (X_0^T X_0)^{-1} X_0^T x_j)^{-1}$$

$$= (1 - x_j^T X_0 (X_0^T X_0)^{-1} X_0^T x_j)^{-1}$$

$\because \bar{x}_j = \frac{1}{n} \sum_{k=1}^{n} x_{jk} = 0$

$$\therefore SS_R^j = \sum_{k=1}^{n} (\hat{x}_{jk} - \bar{x}_j)^2 = \sum_{k=1}^{n} \hat{x}_{jk}^2 = \hat{x}_j^T \hat{x}_j = (X_0 \hat{\beta})^T X_0 \hat{\beta}$$

$$= x_j^T X_0 (X_0^T X_0)^{-1} X_0^T X_0 (X_0^T X_0)^{-1} X_0^T x_j$$

$$= x_j^T X_0 (X_0^T X_0)^{-1} X_0^T x_j$$

$$SS_T^j = \sum_{k=1}^{n} (x_{jk} - \bar{x}_j)^2 = \sum_{k=1}^{n} x_{jk}^2 = 1$$

$$R_j^2 = \frac{SS_R^j}{SS_T^j} = x_j^T X_0 (X_0^T X_0)^{-1} X_0^T x_j$$

$$\therefore C_{jj} = (1 - x_j^T X_0 (X_0^T X_0)^{-1} X_0^T x_j)^{-1} = \frac{1}{1 - R_j^2}$$

Q_2:

$$\hat{\beta} = (X'X)^{-1}X'y \qquad E(\hat{\beta}) = \beta \qquad Var(\hat{\beta}) = \sigma^2 (X'X)^{-1}$$

$$MSE(\hat{\beta}) = E(\hat{\beta}-\beta)'(\hat{\beta}-\beta) = E[\underbrace{(\hat{\beta}-E(\hat{\beta}))'(\hat{\beta}-E(\hat{\beta}))}_{一个二次型}]$$

$$= tr(E[(\hat{\beta}-E(\hat{\beta}))'(\hat{\beta}-E(\hat{\beta}))])$$

$$= tr(Var(\hat{\beta})) = tr(\sigma^2(X'X)^{-1}) = \sigma^2 tr[(X'X)^{-1}]$$

假定 $X'X$ 的特征值为 $\lambda_1, \lambda_2 \cdots\cdots \lambda_p$.

则 $(X'X)^{-1}$ 的特征值有 $\frac{1}{\lambda_1}, \frac{1}{\lambda_2}, \cdots \frac{1}{\lambda_p}$

$$tr[(X'X)^{-1}] = \sum_{i=1}^{p} \frac{1}{\lambda_i}$$

$$\therefore MSE(\hat{\beta}) = \sigma^2 tr[(X'X)^{-1}] = \sigma^2 \sum_{i=1}^{p} \frac{1}{\lambda_i}$$

从贝叶斯统计的角度解释岭回归

最大化后验估计

$$J(w) = \underbrace{L(w)}_{\text{loss}} + k\underbrace{P(w)}_{\text{penalty}}$$ 正则化框架

$$\hat{w} = \arg\min J(w), \quad k>0$$

岭回归：令 $P(w) = w^T w$

$$J(w) = \sum_{i=1}^{N} (w^T x_i - y_i)^2 + k w^T w$$

$$\hat{w} = \arg\min_w J(w) = (X^T X + kI)^{-1} X^T Y$$

贝叶斯角度

$$\varepsilon \sim N(0, \sigma^2)$$

$$Y = w^T X + \varepsilon = f(w) + \varepsilon \qquad Y \sim N(w^T X, \sigma^2)$$

假设参数w也符合高斯分布

$$w \sim N(0, \sigma_1^2)$$

$$P(w) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left\{-\frac{w^T w}{2\sigma_1^2}\right\}$$

$$P(Y|w;X) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(Y-w^TX)^T(Y-w^TX)}{2\sigma^2}\right\}$$

最大后验概率估计

$$P(w|Y) = \frac{P(w)P(Y|w)}{P(Y)}$$

$$\hat{w} = \arg\max_w P(w|Y) = \arg\max_w \frac{P(w)P(Y|w)}{P(Y)} = \arg\max_w P(w)P(Y|w) = \arg\max_w \log(P(w)P(Y|w))$$

$$= \arg\max_w \log\frac{1}{\sqrt{2\pi}\sigma_1} - \frac{w^Tw}{2\sigma_1^2} + \log\frac{1}{\sqrt{2\pi}\sigma_0} - \frac{(Y-w^TX)^T(Y-w^TX)}{2\sigma_0^2}$$

$$= \arg\min_w (Y-w^TX)^2 + \frac{\sigma^2}{\sigma_1^2} w^T w = \arg\min_w \sum_{i=1}^{N}(y_i - w^T x_i) + k w^T w$$

$$\therefore \hat{w} = \arg\min_w \underbrace{\sum_{i=1}^{N}(y_i - w^T x_i)^2}_{\text{loss}} + \underbrace{\frac{\sigma^2}{\sigma_1^2}}_{k} w^T w \rightarrow \text{penalty}$$

(最大后验)

一回事情