统计方法与机器学习 理论作业3 参考答案

1

对任意 $j \in \{1, 2, \cdots, p\}$，设

$$
\boldsymbol{T}_j = \begin{pmatrix}
 & 1 & & & & & \\
 & & \ddots & & & & \\
 & & & 1 & & & \\
1 & & & & & & \\
 & & & & 1 & & \\
 & & & & & \ddots & \\
 & & & & & & 1
\end{pmatrix} \tag{1}
$$

（即将第 j 列与前 $j-1$ 列交换），则

$$
\boldsymbol{T}_j^{-1} = \begin{pmatrix}
 & & & 1 & & & \\
1 & & & & & & \\
 & \ddots & & & & & \\
 & & 1 & & & & \\
 & & & & 1 & & \\
 & & & & & \ddots & \\
 & & & & & & 1
\end{pmatrix} = \boldsymbol{T}_j^T \tag{2}
$$

若令 $\boldsymbol{X}_t = (\boldsymbol{x}_j, \boldsymbol{X}_0)$，其中 $\boldsymbol{X}_0 = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_{j-1}, \boldsymbol{x}_{j+1}, \cdots, \boldsymbol{x}_p)$

则根据题意，有 $\boldsymbol{X}_t = \boldsymbol{X}_s \boldsymbol{T}_j$

因此 $\boldsymbol{X}_s = \boldsymbol{X}_t \boldsymbol{T}_j^T$

于是

$$
\begin{aligned}
\boldsymbol{C} &= \left(\boldsymbol{X}_s^T \boldsymbol{X}_s\right)^{-1} \\
&= \left(\left(\boldsymbol{X}_t \boldsymbol{T}_j^T\right)^T \left(\boldsymbol{X}_t \boldsymbol{T}_j^T\right)\right)^{-1} \\
&= \left(\boldsymbol{T}_j \boldsymbol{X}_t^T \boldsymbol{X}_t \boldsymbol{T}_j^T\right)^{-1} \\
&= \boldsymbol{T}_j \left(\boldsymbol{X}_t^T \boldsymbol{X}_t\right)^{-1} \boldsymbol{T}_j^T
\end{aligned} \tag{3}
$$

而由于

$$
\begin{aligned}
\boldsymbol{X}_t^T \boldsymbol{X}_t &= \begin{pmatrix} \boldsymbol{x}_j^T \\ \boldsymbol{X}_0^T \end{pmatrix} \begin{pmatrix} \boldsymbol{x}_j & \boldsymbol{X}_0 \end{pmatrix} \\
&= \begin{pmatrix} \boldsymbol{x}_j^T \boldsymbol{x}_j & \boldsymbol{x}_j^T \boldsymbol{X}_0 \\ \boldsymbol{X}_0^T \boldsymbol{x}_j & \boldsymbol{X}_0^T \boldsymbol{X}_0 \end{pmatrix}
\end{aligned} \tag{4}
$$

故由分块矩阵求逆公式可知

$$
\left(\boldsymbol{X}_t^T\boldsymbol{X}_t\right)^{-1}=\begin{pmatrix}\boldsymbol{X}_{11}^* & \boldsymbol{X}_{12}^*\\ \boldsymbol{X}_{21}^* & \boldsymbol{X}_{22}^*\end{pmatrix} \tag{5}
$$

其中

$$
\begin{aligned}
\boldsymbol{X}_{11}^* &= \left(\boldsymbol{x}_j^T\boldsymbol{x}_j-\boldsymbol{x}_j^T\boldsymbol{X}_0\left(\boldsymbol{X}_0^T\boldsymbol{X}_0\right)^{-1}\boldsymbol{X}_0^T\boldsymbol{x}_j\right)^{-1}\\
\boldsymbol{X}_{12}^* &= -\left(\boldsymbol{x}_j^T\boldsymbol{x}_j-\boldsymbol{x}_j^T\boldsymbol{X}_0\left(\boldsymbol{X}_0^T\boldsymbol{X}_0\right)^{-1}\boldsymbol{X}_0^T\boldsymbol{x}_j\right)^{-1}\boldsymbol{x}_j^T\boldsymbol{X}_0\left(\boldsymbol{X}_0^T\boldsymbol{X}_0\right)^{-1}\\
\boldsymbol{X}_{21}^* &= -\left(\boldsymbol{X}_0^T\boldsymbol{X}_0\right)^{-1}\boldsymbol{X}_0^T\boldsymbol{x}_j\left(\boldsymbol{x}_j^T\boldsymbol{x}_j-\boldsymbol{x}_j^T\boldsymbol{X}_0\left(\boldsymbol{X}_0^T\boldsymbol{X}_0\right)^{-1}\boldsymbol{X}_0^T\boldsymbol{x}_j\right)^{-1}\\
\boldsymbol{X}_{22}^* &= \left(\boldsymbol{X}_0^T\boldsymbol{X}_0\right)^{-1}+\left(\boldsymbol{X}_0^T\boldsymbol{X}_0\right)^{-1}\boldsymbol{X}_0^T\boldsymbol{x}_j\left(\boldsymbol{x}_j^T\boldsymbol{x}_j-\boldsymbol{x}_j^T\boldsymbol{X}_0\left(\boldsymbol{X}_0^T\boldsymbol{X}_0\right)^{-1}\boldsymbol{X}_0^T\boldsymbol{x}_j\right)^{-1}\boldsymbol{x}_j^T\boldsymbol{X}_0\left(\boldsymbol{X}_0^T\boldsymbol{X}_0\right)^{-1}
\end{aligned} \tag{6}
$$

而由于 $\boldsymbol{X}_s$ 为标准化后的特征，因此 $\boldsymbol{x}_j^T\boldsymbol{x}_j=1$

于是由 $\boldsymbol{T}_j$ 的定义可知

$$
\begin{aligned}
c_{jj} &= \boldsymbol{X}_{11}^*\\
&= \left(\boldsymbol{x}_j^T\boldsymbol{x}_j-\boldsymbol{x}_j^T\boldsymbol{X}_0\left(\boldsymbol{X}_0^T\boldsymbol{X}_0\right)^{-1}\boldsymbol{X}_0^T\boldsymbol{x}_j\right)^{-1}\\
&= \left(1-\boldsymbol{x}_j^T\boldsymbol{X}_0\left(\boldsymbol{X}_0^T\boldsymbol{X}_0\right)^{-1}\boldsymbol{X}_0^T\boldsymbol{x}_j\right)^{-1}
\end{aligned} \tag{7}
$$

另一方面，由于

$$
\bar{x}_j=\frac{1}{n}\sum_{k=1}^n x_{jk}=0 \tag{8}
$$

因此

$$
\begin{aligned}
SS_R^j &= \sum_{k=1}^n\left(\hat{x}_{jk}-\bar{x}_j\right)^2\\
&= \sum_{k=1}^n \hat{x}_{jk}^2\\
&= \hat{\boldsymbol{x}}_j^T\hat{\boldsymbol{x}}_j\\
&= \left(\boldsymbol{X}_0\hat{\boldsymbol{\beta}}\right)^T\boldsymbol{X}_0\hat{\boldsymbol{\beta}}\\
&= \boldsymbol{x}_j^T\boldsymbol{X}_0\left(\boldsymbol{X}_0^T\boldsymbol{X}_0\right)^{-1}\boldsymbol{X}_0^T\boldsymbol{X}_0\left(\boldsymbol{X}_0^T\boldsymbol{X}_0\right)^{-1}\boldsymbol{X}_0^T\boldsymbol{x}_j\\
&= \boldsymbol{x}_j^T\boldsymbol{X}_0\left(\boldsymbol{X}_0^T\boldsymbol{X}_0\right)^{-1}\boldsymbol{X}_0^T\boldsymbol{x}_j
\end{aligned} \tag{9}
$$

而由标准化条件可知

$$
SS_T^j=\sum_{k=1}^n(x_{jk}-\bar{x}_j)^2=\sum_{k=1}^n x_{jk}^2=1 \tag{10}
$$

故该回归模型的复决定系数

$$
R^2=\frac{SS_R^j}{SS_T^j}=\boldsymbol{x}_j^T\boldsymbol{X}_0\left(\boldsymbol{X}_0^T\boldsymbol{X}_0\right)^{-1}\boldsymbol{X}_0^T\boldsymbol{x}_j \tag{11}
$$

由此可得

$$
c_{jj} = \left(1 - \boldsymbol{x}_j^T \boldsymbol{X}_0 \left(\boldsymbol{X}_0^T \boldsymbol{X}_0\right)^{-1} \boldsymbol{X}_0^T \boldsymbol{x}_j\right)^{-1} = \frac{1}{1 - R^2} \tag{12}
$$

2

由线性回归最小二乘估计的性质可知

$$
\begin{aligned}
E\left[\hat{\boldsymbol{\beta}}\right] &= \boldsymbol{\beta} \\
\mathbf{Cov}\left(\hat{\boldsymbol{\beta}}\right) &= \sigma^2 \left(\boldsymbol{X}^T \boldsymbol{X}\right)^{-1}
\end{aligned} \tag{13}
$$

因此

$$
\begin{aligned}
MSE\left(\hat{\boldsymbol{\beta}}\right) &= E\left[\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)^T \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)\right] \\
&= E\left[\left(\hat{\boldsymbol{\beta}} - E[\hat{\boldsymbol{\beta}}] + E[\hat{\boldsymbol{\beta}}] - \boldsymbol{\beta}\right)^T \left(\hat{\boldsymbol{\beta}} - E[\hat{\boldsymbol{\beta}}] + E[\hat{\boldsymbol{\beta}}] - \boldsymbol{\beta}\right)\right] \\
&= E\left[\left(\hat{\boldsymbol{\beta}} - E[\hat{\boldsymbol{\beta}}]\right)^T \left(\hat{\boldsymbol{\beta}} - E[\hat{\boldsymbol{\beta}}]\right)\right] \\
&= \mathrm{tr}\left[\mathbf{Cov}\left(\hat{\boldsymbol{\beta}}\right)\right] \\
&= \sigma^2 \mathrm{tr}\left[\left(\boldsymbol{X}^T \boldsymbol{X}\right)^{-1}\right]
\end{aligned} \tag{14}
$$

进一步的，若 $\boldsymbol{X} \in \mathbb{M}_{n \times (p+1)}(\mathbb{R})$，设 $\boldsymbol{X}^T \boldsymbol{X}$ 的特征值为 $\lambda_1, \lambda_2, \cdots, \lambda_{p+1}$

则由 $\boldsymbol{X}^T \boldsymbol{X}$ 为对称矩阵可知 $\lambda_1, \lambda_2, \cdots, \lambda_{p+1} \in \mathbb{R}$

因此 $\left(\boldsymbol{X}^T \boldsymbol{X}\right)^{-1}$ 的特征值为 $\frac{1}{\lambda_1}, \frac{1}{\lambda_2}, \cdots, \frac{1}{\lambda_{p+1}}$

此时

$$
\begin{aligned}
MSE\left(\hat{\boldsymbol{\beta}}\right) &= \sigma^2 \mathrm{tr}\left[\left(\boldsymbol{X}^T \boldsymbol{X}\right)^{-1}\right] \\
&= \sigma^2 \sum_{i=1}^{n} \frac{1}{\lambda_i}
\end{aligned} \tag{15}
$$