温旭和 1020550143 2. 统计方法与机器学习作业3.

1. 证：令 $X_s = (X_{s-1}, x_p)$.

则 $X_s'X_s = \begin{pmatrix} X_{s-1}'X_{s-1} & X_{s-1}'x_p \\ x_p'X_{s-1} & x_p'x_p \end{pmatrix}$

$$c_{jjpp} = \left(x_p'x_p - x_p'X_{s-1}(X_{s-1}'X_{s-1})^{-1}X_{s-1}'x_p\right)^{-1}$$

$$= \frac{1}{SSE_{-s}}$$

其中：SSE_{-s} 是将多元线性回归模型

$$x_p = \beta_0 + \beta_1 x_1 + \cdots + \beta_{p-1}x_{p-1} + \varepsilon$$

中所有因变量，自变量标准化后所得的"SS_E".

设标准化前的"SS_E""SS_T"分别为 SS_E, SS_T.

则 $SS_{E-s} = \dfrac{SS_E}{SS_T}$ （由于 $SS_T = \sum_{i=1}^{n}(y_i-\bar{y})^2$）

分块矩阵求逆

故 $c_{jjpp} = \dfrac{SS_T}{SS_E}$.

SST, SSE 剩余的（最后一个

又 $\dfrac{1}{1-R_{jp}^2} = \dfrac{1}{1-\frac{SS_R}{SS_T}} = \dfrac{SS_T}{SS_E}$. 所有的

故 $c_{jjpp} = \dfrac{1}{1-R_{jp}^2}$.

由于任意自变量都可以被换到最后一个自变量上，

又，故对 $\forall j \in [1, p] \cap \mathbb{Z}$，有

$$c_{jj} = \frac{1}{1-R_j^2}$$

2. 解：$MSE(\hat\beta) = E(\hat\beta - \beta)'(\hat\beta - \beta)$

$= E((X'X)^{-1}X'y-\beta)'((X'X)^{-1}X'y-\beta)$.

扩展的统计方法

其中: $E(y'X(X'X)^{-1}(X'X)^{-1}X'y)$

$= tr(\sigma^2 X(X'X)^{-1}(X'X)^{-1}X') + \beta'X'X(X'X)^{-1}(X'X)^{-1}X'X\beta$

(由于 $E(y) = X\beta$. $Var(y) = \sigma^2 I_n$)

$= tr(\sigma^2(X'X)^{-1}) + \beta'\beta$.

$E(\beta'(X'X)^{-1}X'y) = E(y'X(X'X)^{-1}\beta)$

$= \beta'(X'X)^{-1}X'X\beta = \beta'\beta$

$E(\beta'\beta) = \beta'\beta$.

故 $MSE(\hat{\beta}) = tr(\sigma^2(X'X)^{-1}) + \beta'\beta - \beta'\beta - \beta'\beta + \beta'\beta$

$= tr(\sigma^2(X'X)^{-1})$.

由于 X 已经标准化,

进一步化简的依据?

故 $MSE(\hat{\beta}) = tr(\sigma^2(X'X)^{-1})$

$= \sigma^2 \sum_{i=1}^{n} c_{ii}$.