统计方法 理论作业3

1. 证明：根据题意，由数据已经被标准化，可得 $1^T x_i = 0$ 对于 $x_i^T x_i = 0$ $(i=1,2\cdots p)$

$$SS_R^j = \sum_{i=1}^{n}(\hat{x}_{ji} - \bar{x}_j)^2 = \sum_{i=1}^{n}\hat{x}_{ji}^2 = \hat{x}_j^T\hat{x}_j$$

其中 $\hat{x}_j = X_{sj}\hat{\beta} = X_{sj}(X_{sj}^T X_{sj})^{-1}X_{sj}^T \cdot x_j$

因为这里 x_j 为因变量

线性模型，估计即可

$X_{sj} = [x_1\ x_2 \cdots x_{j-1}\ x_{j+1} \cdots x_p)$

除去了第 j 列

故 $SS_R^j = \hat{x}_j^T\hat{x}_j$

$$= x_j^T X_{sj}(X_{sj}^T X_{sj})^{-1}X_{sj}^T \cdot X_{sj}(X_{sj}^T X_{sj})^{-1}X_{sj}^T \cdot x_j$$

$$= x_j^T X_{sj}(X_{sj}^T X_{sj})^{-1}X_{sj}^T x_j$$

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

该回归模型的复决定系数 $R_j^2 = \dfrac{SS_R^j}{SS_T^j} = x_j^T X_{sj}(X_{sj}^T X_{sj})^{-1}X_{sj}^T x_j$

现在只需证明 $C_{jj} = \dfrac{1}{1-R_j^2}$ $R_j^2 = \cdots$ 即可证原式

X_{sj} 除去了 x_j，我要把 x_j 抽出来，并且保证顺序不乱

而且这个 C 来源于 X_s 而不是 X_{sj}，所以要尽量找一些联系

$X_s = [x_j\ \ X_{sj}] \cdot A_j$ 由线代知识，A_j 为初等矩阵（把 I_p 的第 j 列换到第1列，第1—$j-1$ 列往后推一列）

$$A_j = \begin{bmatrix} & 1 & & & \\ & & \ddots & & \\ 1 & & & 0 & \\ & & & & \ddots \\ & & & & & 1 \end{bmatrix}$$

$A_j^{-1} = A_j^T$（A_j^T 就对行施加类似的操作即可）

$$C = (X_s^T X_s)^{-1} = \left(A_j^T \begin{bmatrix} x_j^T \\ X_{sj}^T \end{bmatrix} [x_j\ \ X_{sj}] A_j\right)^{-1}$$

$$= A_j^{-1}\begin{bmatrix} 1 & x_j^T X_{sj} \\ X_{sj}^T x_j & X_{sj}^T X_{sj} \end{bmatrix}^{-1}(A_j^T)^{-1}$$

$$= A_j^T$$

这里我搞反了!!!

$X_s \cdot A_j = [x_j\ \ X_{sj}]$

$X_s = [x_j\ \ X_{sj}] A_j^T$

$$C = (X_s^T X_s)^{-1} = \left(A_j \begin{bmatrix} x_j^T \\ X_{sj}^T \end{bmatrix} [x_j\ \ X_{sj}] A_j^T\right)^{-1}$$

$$= A_j \begin{bmatrix} 1 & x_j^T X_{sj} \\ X_{sj}^T x_j & X_{sj}^T X_{sj} \end{bmatrix}^{-1} A_j^T$$

A_j 左乘，意味着第1行放到第 j 行，然后 2—j 行往前移一行

A_j^T 右乘，意味着第1列放到第 j 列，然后 2—$j-1$ 列往前移一列

$$B = \begin{bmatrix} 1 & x_j^T X_{sj} \\ X_{sj}^T x_j & X_{sj}^T X_{sj} \end{bmatrix}^{-1}$$

那么 C_{jj} 代表的是 B 的第1行第1个元素!!!

根据逆矩阵的公式（分块）

可得 $C_{jj} = (x_j^T x_j - x_j^T X_{sj}(X_{sj}^T X_{sj})^{-1}X_{sj}^T x_j)^{-1}$

$$= \frac{1}{1 - x_j^T X_{sj}(X_{sj}^T X_{sj})^{-1}X_{sj}^T x_j}$$

再与之前求得的 R_j^2 比较，很可得原式得证

PJ

2、解：$MSE(\hat{\beta}) = E\left((\hat{\beta}-\beta)^T(\hat{\beta}-\beta)\right)$

$E(\hat{\beta})=\beta$

这都是线性回归模型的性质

$$= E\left((\hat{\beta}-E(\hat{\beta}))^T(\hat{\beta}-E(\hat{\beta}))\right)$$

$cov(\hat{\beta})=\sigma^2(X^TX)^{-1}$

$$= Tr(cov(\hat{\beta}))$$

$$= Tr(\sigma^2(X^TX)^{-1}) = \sigma^2 Tr((X^TX)^{-1})$$

3、证明：这个题主要应该去在贝叶斯估计角度 ~~证明这个解与岭回归~~ 得到一个解，然后证明它与课上学到的岭回归模型之等价性。

课上学到的岭回归模型：

$$\hat{\beta}(k) = (X^TX + kI)^{-1}(X^TX)\hat{\beta}$$

→ 从逻辑上来说，我们不能先假设模型为岭回归模型，再利用它的结论，再用贝叶斯估计来证明自己！

然后，我们从最本质的问题开始推

~~最本质：~~ 由题 $y_i \sim N(\beta_1 x_1 + \cdots + \beta_p x_p, \delta^2)$

可以看到，β_0 压根不考虑了。

→ 感觉应该是 $y_i \sim N(\beta_1 x_{i1} + \cdots + \beta_p x_{ip}, \delta^2)$

取 $X = \begin{bmatrix} x_1^T \\ \vdots \\ x_n^T \end{bmatrix}$ $\beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$ $y_i | x_i, \beta \sim N(\beta^T x_i, \delta^2)$

取 $\beta \sim N(0, \Sigma)$

$$P(y|X,\beta) = \prod_{i=1}^{n} P(y_i | x_i, \beta) = \frac{1}{(\sqrt{2\pi})^n \delta^n} \exp\{-\frac{1}{2\delta^2}\sum_{i=1}^{n}(y_i - \beta^T x_i)^2\}$$

其中转化为矩阵形式 $\sum_{i=1}^{n}(y_i - \beta^T x_i)^2 = [y_1 - \beta^T x_1 \quad y_2 - \beta^T x_2 \cdots y_n - \beta^T x_n]\begin{bmatrix} y_1 - \beta^T x_1 \\ \vdots \\ y_n - \beta^T x_n \end{bmatrix}$

$$= (y^T - \beta^T X^T)\cdot(y - X\beta)$$

$$= y^Ty - y^TX\beta - \beta^TX^Ty + \beta^TX^TX\beta$$

$$= y^Ty + \beta^TX^TX\beta - 2\beta^TX^Ty$$ （因为 y 的维数 n×1）

故 $P(y|X,\beta) = \frac{1}{(\sqrt{2\pi}\delta)^n} \exp\{-\frac{1}{2\delta^2}(y^Ty + \beta^TX^TX\beta - 2\beta^TX^Ty)\}$

$$\pi(\beta) = \frac{1}{\sqrt{2\pi}|\Sigma|^{\frac{1}{2}}} \exp(-\frac{1}{2}\beta^T\Sigma^{-1}\beta)$$

~~由贝叶斯估计~~

$$P(\beta|X,y) = \frac{P(\beta, X, y)}{P(X,y)} = \frac{P(y|X,\beta)\cdot P(X|\beta)\cdot \pi(\beta)}{P(X,y)}$$

$$\propto P(y|X,\beta)\cdot\pi(\beta) = \frac{1}{(\sqrt{2\pi})^{n+1}\delta^n|\Sigma|^{\frac{1}{2}}} \exp\{-\frac{1}{2\delta^2}(y^Ty + \beta^TX^TX\beta - 2\beta^TX^Ty) - \frac{1}{2}\beta^T\Sigma^{-1}\beta\}$$

↓ 设其为 $L(\beta)$

$$f(\beta) = \log L(\beta) = -\log((\sqrt{2\pi})^{n+1}\delta^n|\Sigma|^{\frac{1}{2}}) - \frac{1}{2\delta^2}(y^Ty + \beta^TX^TX\beta - 2\beta^TX^Ty) - \frac{1}{2}\beta^T\Sigma^{-1}\beta$$

虽然一般来说此时得到了β的后验概率分布，应该求其期望作为贝叶斯估计值

$\hat{\beta} = \arg\max_{\beta} f(\beta)$　　$f(\beta)$ 为~~凸~~凹函数　　$\nabla_\beta f(\beta) = -\frac{1}{2\delta^2}\cdot(2X^TX\beta - 2X^Ty) - \frac{1}{2}\cdot 2\Sigma^{-1}\beta$

但这里我们用的是极大值作为其点估计

令 $\nabla_\beta f(\beta) = 0$ 可得　$\hat{\beta} = (X^TX + \delta^2\Sigma^{-1})^{-1}\cancel{\delta}X^Ty$　①

课上学到的岭回归估计的具体β：$\hat{\beta} = (X^TX + kI)^{-1}X^Ty$　②

比较①和②　令 $\delta^2\Sigma^{-1} = kI$　即岭参数，δ^2，Σ^{-1} 关联了起来。

结构类似，终于，我从贝叶斯估计的角度证明了岭回归的合理性。