作业 **12**

1. 假设 $(y_i, \mathbf{x}_i), i = 1, \dots, n$ 满足线性模型

$$y_i = \beta_0 + \mathbf{x}_i^\top \mathbf{b} + \epsilon_i, \quad \epsilon_i \sim (0, \sigma^2), \quad i = 1, \dots, n.$$

若对某个 $1 \le k \le n$, $y_k = \bar{y} = \sum_{i=1}^n y_i/n$, $\mathbf{x}_k = \bar{\mathbf{x}} = \sum_{i=1}^n \mathbf{x}_i/n$, 证明删除 $(y_k, \mathbf{x}_k)$ 不影响最小二乘法。

2. 假设 $\mathbf{y}_1, \dots, \mathbf{y}_n$ iid $\sim N(\theta, \sigma^2 I_p)$, 其中 θ 为 $p \times 1$ 未知参数向量，假设 σ^2 已知。令 $\tilde{\theta} = \lambda \bar{\mathbf{y}}$, 其中 $0 \le \lambda \le 1$ 是常数。

   (a) 求 $\tilde{\theta}$ 的均方误差 $m(\lambda) = E\|\tilde{\theta} - \theta\|^2$.

   (b) 求 $\lambda_{opt} = arg\min_\lambda m(\lambda)$ （即求使得 $m(\lambda)$ 达到最小的 λ）.

   (c) λ_{opt} 中含有未知 $\|\theta\|^2$, 以其无偏估计代入（注意 $\|\bar{\mathbf{y}}\|^2$ 不是 $\|\theta\|^2$ 的无偏估计)，得到的 $\tilde{\theta} = \hat{\lambda}_{opt}\bar{\mathbf{y}}$ 是否等于或接近 James-Stein 估计？

3. 假设模型

$$\mathbf{y}_{n\times 1} = X_{n\times p}\boldsymbol{\beta}_{p\times 1} + \epsilon_{n\times 1}, \epsilon \sim (0, \sigma^2 I_n),$$

其中 X 的第一列为向量 $\mathbf{1}$. 回归系数的最小二乘估计记为 $\hat{\boldsymbol{\beta}} = (X^\top X)^{-1}X^\top \mathbf{y}$。假设有“新”数据 $\mathbf{x}_0, y_0$ 满足模型 $y_0 = \mathbf{x}_0^\top \boldsymbol{\beta} + \epsilon_0, \epsilon_0 \sim (0, \sigma^2)$, 其中 $\mathbf{x}_0$ 已知, 需要预测 y_0。预测统计量取为 $\hat{y}_0 = \mathbf{x}_0^\top \hat{\boldsymbol{\beta}}$。证明如下结论

   (a) $\hat{y}_0$ 的预测误差为 $pe(\hat{y}_0) = E(\hat{y}_0 - y_0)^2 = \sigma^2\left[1 + \mathbf{x}_0^\top (X^\top X)^{-1}\mathbf{x}_0\right]$.

   (b) 当 $\mathbf{x}_0 = \mathbf{x}_i$($X$ 的第 i 行), $pe(\hat{y}_0) = (1 + h_{ii})\sigma^2 \overset{\text{记作}}{=} e(\mathbf{x}_i)$, 其中 h_{ii} 为 $H = X(X^\top X)^{-1}X^\top$ 的 (i,i) 元;

   (c) 当 $\mathbf{x}_0 = \bar{\mathbf{x}}$, $\hat{y}_0 = \bar{y} = \mathbf{1}^\top \mathbf{y}/n$, 且 $pe(\hat{y}_0) = (1 + 1/n)\sigma^2 \overset{\text{记作}}{=} e(\bar{\mathbf{x}})$, 说明 $e(\bar{\mathbf{x}}) \le e(\mathbf{x}_i)$, $i = 1, \dots, n$.

   (d) 当 $\mathbf{x}_0 = X^\top \mathbf{a}$ (这里 $\mathbf{a} \in R^n$)，$pe(\hat{y}_0) \le \sigma^2(1 + \|\mathbf{a}\|^2)$.

4. 条件同上一题，对任何给定的 $\mathbf{x}_0 \in R^p$, 我们需要预测对应的 y_0, 但 y_0 的预测取为 $\tilde{y}_0 = \mathbf{x}_0^\top \tilde{\boldsymbol{\beta}}$，其中 $\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}\lambda$, $0 \le \lambda \le 1$。

   (a) 证明 $\tilde{y}_0$ 的预测误差为

$$pe(\tilde{y}_0) = E(\tilde{y}_0 - y_0)^2 = (1-\lambda)^2(\mathbf{x}_0^\top \boldsymbol{\beta})^2 + \lambda^2\sigma^2\mathbf{x}_0^\top (X^\top X)^{-1}\mathbf{x}_0 + \sigma^2.$$

   (b) 证明如果 $\|X\boldsymbol{\beta}\|^2 \le \frac{1+\lambda}{1-\lambda}\sigma^2$, 则 $pe(\tilde{y}_0) \le pe(\hat{y}_0)$，其中 $\hat{y}_0 = \mathbf{x}_0^\top \hat{\boldsymbol{\beta}}$ 为基于 LS 估计的预测。

5. 假设模型 $\mathbf{y}_{n\times 1} = X_{n\times p}\beta_{p\times 1} + \epsilon_{n\times 1}$, $\epsilon \sim (0, \sigma^2 I_n)$ 中 $X, \mathbf{y}$ 都已经中心化（因此模型中没有截距项)，对设计阵 X 进行奇异值分解:

$$X_{n\times p} = U_{n\times p}D_{p\times p}V_{p\times p}^\top,$$

其中 $U^\top U = I_p, V^\top V = I_p, D = \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_p})$, 这里 $\lambda_1 \ge \dots \ge \lambda_p > 0$ 是 $X^\top X$ 的特征根。所以 $X\beta = UDV^\top \beta = U\gamma$, 其中 $\gamma = DV^\top \beta$。由此，我们改写原模型为

$$\mathbf{y} = U\gamma + \epsilon, \epsilon \sim (0, \sigma^2 I_n),$$

此模型称为主成分回归模型。记 U 的各列为 $U = (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_p)$。

(a) 证明主成分回归模型中回归系数 γ 的最小二乘估计为 $\hat{\gamma} = U^\top \mathbf{y}$，由此求出原模型中的 β 的 LS 估计。

(b) 主成分回归模型中, 响应变量的基于最小二乘估计 $\hat{\gamma}$ 的拟合值向量为 $\hat{\mathbf{y}} = U\hat{\gamma} = \sum_{j=1}^{p} \mathbf{u}_j(\mathbf{u}_j^\top \mathbf{y})$, 求其均方误差 $m(\hat{\mathbf{y}}) = E\|\hat{\mathbf{y}} - U\gamma\|^2$.

(c) 设下标集合 $A_q = \{i_1, ..., i_q\} \subset \{1, 2, ..., p\}$, $1 \le q \le p-1$，令

$$\tilde{\mathbf{y}}^{(A_q)} = \sum_{j \in A_q} \mathbf{u}_j(\mathbf{u}_j^\top \mathbf{y}).$$

求其均方误差 $m(\tilde{\mathbf{y}}^{(A_q)}) = E\|\tilde{\mathbf{y}}^{(A_q)} - U\gamma\|^2$。假设 $\|\gamma\|^2 \le \sigma^2$, 证明

$$m(\tilde{\mathbf{y}}^{(A_q)}) \le m(\hat{\mathbf{y}}).$$

(d) 写出 C_q 准则的具体表达，给出最优子集搜索算法。