Assignment 3 证明题

Name: Han Siyue

ID: 17307110012

October 7, 2019

1. 证明最小二乘估计在线性回归假设成立的前提下与最大似然估计得到的解一致

   证：假设数据为$(Y_1, X_1), \ldots, (Y_i, X_i), \ldots, (Y_n, X_n)$，其中$X_i = (X_{i1}, \ldots, X_{ik})$.

   则线性回归模型 $Y_i = \sum_{j=1}^{k} \beta_j X_{ij} + \epsilon_i$ 的 pdf 为 $f_{Y_i}(y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{(Y_i - \sum_{j=1}^{k} \beta_j X_{ij})^2}{2\sigma^2}\}$.

   则 Y 的 joint pdf 为 $f_Y(y) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{(Y_i - \sum_{j=1}^{k} \beta_j X_{ij})^2}{2\sigma^2}\}$

   则 Log-likelihood 为 $L(\tilde{\beta}) = -\sum_{i=1}^{n} \frac{(Y_i - \sum_{j=1}^{k} \tilde{\beta}_j X_{ij})^2}{2\sigma^2} - \frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2$

   所以 $\widehat{\beta}_{MLE} = argmax_{\tilde{\beta}} L(\tilde{\beta}) = argmin_{\tilde{\beta}} \sum_{i=1}^{n} (Y_i - \sum_{j=1}^{k} \tilde{\beta}_j X_{ij})^2 = \widehat{\beta}_{LSE}$

   得证

2. 证明Gauss-Markov 定理： the least squares estimates of the parameters β have the smallest variance among all linear unbiased estimators.

   证：由最小二乘估计的性纸可得，$\widehat{\beta}_{LSE}$ 是线性无偏估计，假设 $\tilde{\beta}$ 是任一线性无偏估计，

   下证 $Var(\tilde{\beta}|X) \succeq Var(\widehat{\beta}_{LSE}|X)$，即 $Var(\tilde{\beta}|X) - Var(\widehat{\beta}_{LSE}|X)$ 为半正定矩阵。

   因为已知 $Var(\widehat{\beta}_{LSE}|X) = \sigma^2 (X^T X)^{-1}$，即求 $Var(\tilde{\beta}|X)$ 的表达式。

   由于 $\tilde{\beta}$ 是线性估计，存在矩阵 C，使得 $\tilde{\beta} = Cy$，又 $\widehat{\beta}_{LSE} = Ay$，其中 $A = (X^T X)^{-1} X^T$。

   令 $D = C - A$，则 $\tilde{\beta} = Cy = (A + D)y = \widehat{\beta}_{LSE} + D(X\beta + \epsilon) = \widehat{\beta}_{LSE} + DX\beta + D\epsilon$，

   由于 $\tilde{\beta}$ 是无偏估计，

   $$\beta = \mathrm{E}(\tilde{\beta}|X) = \mathrm{E}(\widehat{\beta}_{LSE} + DX\beta + D\epsilon|X) = DX\beta + D\mathrm{E}(\epsilon|X) + \mathrm{E}(\widehat{\beta}_{LSE}|X) = DX\beta + \beta$$

   即 $\forall\ \beta$，$DX\beta = 0$，可得 $DX = 0$，所以 $\tilde{\beta} = \widehat{\beta}_{LSE} + D\epsilon$，继而，

   $$\begin{aligned} Var(\tilde{\beta}|X) &= Var(\tilde{\beta} - \beta|X) = Var(\widehat{\beta}_{LSE} - \beta + D\epsilon|X) = Var[(D + A)\epsilon|X] \\ &= (D + A)Var(\epsilon|X)(D + A)^T = \sigma^2 (D + A)(D + A)^T \\ &= \sigma^2 (DD^T + DA^T + AD^T + AA^T) \\ &\text{and} \because A = (X^T X)^{-1} X^T \text{ , } DX = 0 \\ Var(\tilde{\beta}|X) &= \sigma^2 (DD^T + DX(X^T X)^{-1} + (X^T X)^{-1} X^T D^T + (X^T X)^{-1} X^T X (X^T X)^{-1}) \\ &= \sigma^2 (DD^T + 0 + 0 + (X^T X)^{-1}) = \sigma^2 (DD^T + (X^T X)^{-1}) \\ &\text{and} \because Var(\widehat{\beta}_{LSE}|X) = \sigma^2 (X^T X)^{-1} \text{ , } DD^T \succeq 0 \end{aligned}$$

   $$\therefore Var(\tilde{\beta}|X) - Var(\widehat{\beta}_{LSE}|X) = \sigma^2 DD^T \succeq 0$$

   得证