Statistics problems: personal-overall

May 2019

**1**. Consider the multiple linear regression model

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \beta + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \end{pmatrix},$$

where $Y_1 \in \mathbb{R}^{n_1}$, $Y_2 \in \mathbb{R}^{n_2}$, $X_1$ is a $(n_1 \times p)$ matrix, $X_2$ is a $(n_2 \times p)$ matrix, $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ has rank $p$, and error terms $\epsilon_1$ and $\epsilon_2$ are independent of each other with

$$\epsilon_1 \sim N_{n_1}(0, \sigma^2 I_{n_1}) \text{ and } \epsilon_2 \sim N_{n_2}(0, \rho\sigma^2 I_{n_2}) \qquad (\rho > 0).$$

The unknown parameters are $\beta \in \mathbb{R}^p$ and $\sigma^2$.

1. Treat $\rho$ as a constant and driver the maximum likelihood estimates of $\beta$ and $\sigma^2$, denoted by $\hat{\beta}_\rho$ and $\hat{\sigma}^2_\rho$.
2. Suppose that $X_1$ has full rank $p$ and $X_2$ has full rank $n_2 < p$. Prove that as $\rho$ goes to zero, $\hat{\beta}_\rho$ converges to
$$\hat{\beta} + \left(X_1'X_1\right)^{-1} X_2' \left[X_2 \left(X_1'X_1\right)^{-1} X_2'\right]^{-1} \left(Y_2 - X_2\hat{\beta}\right)$$
where $\hat{\beta} = (X_1'X_1)^{-1}X_1'Y_1$.
3. Interpret the above limit in some context of multiple linear regression with constraints on $\beta$.

**2**. Consider the simple linear regression

$$Y_i = \alpha + \beta X_i + \epsilon_i, \qquad i = 1, \ldots, n.$$

Define quadratic function

$$Q(\alpha, \beta) = \sum_{i=1}^{n} (Y_i - \alpha - \beta X_i)^2.$$

Let $\hat{\alpha}$ and $\hat{\beta}$ be the estimators of $\alpha$ and $\beta$, which minimizes $Q(\alpha, \beta)$. Let $\hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i$.
1. Find the gradient vector of $\hat{Y}_i$ with respect to the vector $Y = (Y_1 \ldots, Y_n)'$.
2. The degree of freedom, $d_{LM}$ of the fitted model is defined to be the trace of matrix $\frac{\partial \hat{Y}}{\partial Y}$, where $\hat{Y} = (\hat{Y}_1 \ldots, \hat{Y}_n)'$. Find $d_{LM}$. How is it related to the model?