PHS596 HW3

Shuang Gao

September 2020

1 Derivation of regression on indicator matrix

Suppose there are K categories in the classification problem, the dummy variable coding for Y is,

$$Y_{n,K} = \begin{pmatrix} 0 & 0 & \cdots & 1 \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & \cdots & 0 \end{pmatrix} = \begin{pmatrix} y_1 & y_2 & \cdots & y_K \end{pmatrix}$$

design matrix $\mathbf{X}$ with dimension nxp, and parameter matrix $\mathbf{B} = (\beta_1, \cdots \beta_K)$ with dimension pxK.

The objective function for classification is,

$$min_\beta \operatorname{vec}(\mathbf{Y} - \mathbf{XB})^T \operatorname{vec}(\mathbf{Y} - \mathbf{XB})$$

Taking the first derivative with respect to β_k and get the least square solution by setting the first derivative to 0.

$$\begin{aligned} \frac{\partial \operatorname{vec}(\mathbf{Y} - \mathbf{XB})^T \operatorname{vec}(\mathbf{Y} - \mathbf{XB})}{\partial \beta_k} &= \frac{\sum_{k=1}^{K} (y_k - \mathbf{X}^T \beta_k)^T (y_k - \mathbf{X}^T \beta_k)}{\partial \beta_k} \\ &= \frac{(y_k - \mathbf{X}^T \beta_k)^T (y_k - \mathbf{X}^T \beta_k)}{\partial \beta_k} \\ &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X} y_k \end{aligned}$$

which is the same as the least square solution for regress y_k over $\mathbf{X}$. Thus, $\hat{\mathbf{B}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}(y_1 \cdots y_K) = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{XY}$