Statistical Machine Learning

Homework 2 yc3356 YL CHEN

(1) We have $\begin{cases} \text{training data } (x_i, y_i),\ i=1,\dots,n \Rightarrow \hat{\beta} = (X^TX)^{-1}X^TY \\ \text{testing data } (\tilde{x}_i, \tilde{y}_i),\ i=1,\dots,m \Rightarrow \tilde{\beta} = (\tilde{X}^T\tilde{X})^{-1}\tilde{X}^T\tilde{Y} \end{cases}$

method one

$E(R_{tr}(\hat{\beta})) \leq E(R_{te}(\hat{\beta}))$

$\Leftrightarrow E[\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{\beta}^T x_i)^2] \leq E[\frac{1}{M}\sum_{i=1}^{M}(\tilde{y}_i - \hat{\beta}^T \tilde{x}_i)^2]$

△ First: As we know: the expect error rate is the same whether we have just one data or N data:

$$E(\frac{1}{M}\sum_{i=1}^{m}(y_i - \hat{\beta}^T x_i)^2) = E(y_i - \hat{\beta}^T x_i)^2$$

For the same reason

$$E(\frac{1}{M}\sum_{i=1}^{M}(\tilde{y}_i - \hat{\beta}^T \tilde{x}_i)^2) = E(\tilde{y}_i - \hat{\beta}^T \tilde{x}_i)^2$$

Thus, the numer of data do not matter. to simplify the question, we assume that ~~M=N~~ :

$$E(\frac{1}{M}\sum_{i=1}^{M}(\tilde{y}_i - \hat{\beta}^T \tilde{x}_i)^2) = E(\frac{1}{N}\sum_{i=1}^{N}(\tilde{y}_i - \hat{\beta}^T \tilde{x}_i))$$

△ Based on the regression:

$$\begin{cases} \sum_{i=1}^{N} E(y_i - \hat{\beta}^T x_i)^2 = (N-p-1)\sigma^2 \\ \sum_{i=1}^{M} E(\tilde{y}_i - \tilde{\beta}^T \tilde{x}_i)^2 = (M-p-1)\sigma^2 \end{cases}$$ (σ^2 based on the same true data and true model)

$\Rightarrow$ if we assume $m=n$ (do not matter)

$$E(\frac{1}{N}\sum_{i=1}^{N}(\tilde{y}_i - \tilde{\beta}^T \tilde{x}_i)^2) = E(\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{\beta}^T x_i)^2) = (N-p-1)\sigma^2$$

△ Based on the defination of least equare estimate:

$$E(\frac{1}{N}\sum_{i=1}^{N}(\tilde{y}_i - \hat{\beta}^T \tilde{x}_i)^2) \geq E(\frac{1}{N}\sum_{i=1}^{N}(\tilde{y}_i - \tilde{\beta}^T \tilde{x}_i)^2)$$

(Not least square estimator) (least square estimator)

△ Thus, we have:

$$E(\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{\beta}^T x_i)) = E(\frac{1}{N}\sum_{i=1}^{N}(\tilde{y}_i - \tilde{\beta}^T \tilde{x}_i)^2) \leq E(\frac{1}{M}\sum_{i=1}^{M}(\tilde{y}_i - \hat{\beta}^T \tilde{x}_i)^2)$$

$\Rightarrow E(R_{tr}(\hat{\beta})) \leq E(R_{te}(\hat{\beta}))$