

Exercise 1.4 Week 3

Want to show that

$$\mathbb{E}[R_{\text{train}}(\hat{\beta})] \leq \mathbb{E}[R_{\text{test}}(\hat{\beta})]$$

First, note that

$$\begin{aligned} \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n (y_i - x_i^T \hat{\beta})^2\right] &= \\ \frac{1}{n} \sum_{i=1}^n \mathbb{E}\left[(y_i - x_i^T \hat{\beta})^2\right] &= \mathbb{E}\left[(y_i - x_i^T \hat{\beta})^2\right] \end{aligned}$$

(*) For every ~~Thus, it is independent of the sample size n (or m).~~

We do however know that in any training set of size n ,

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} R_{\text{train}}(\beta)$$

such that

$$\mathbb{E}[R_{\text{train}}(\hat{\beta})] \leq \mathbb{E}[R_{\text{train}}(\beta)] \quad , \quad \beta \in \mathbb{R}^p$$

where the expectation is over all training sets.

Similarly, there is a vector $\tilde{\beta}$ that minimizes $R_{\text{test}}(\beta)$,

$$\mathbb{E}[R_{\text{test}}(\tilde{\beta})] \leq \mathbb{E}[R_{\text{test}}(\beta)]$$

and in particular

$$\mathbb{E}[R_{\text{test}}(\tilde{\beta})] \leq \mathbb{E}[R_{\text{test}}(\hat{\beta})]$$

Suppose the two sets are of equal size, $n=m$. Since we randomly assign observations to train and test sets,

$$\begin{aligned} (*) \quad \mathbb{E}[R_{\text{train}}(\hat{\beta})] &= \mathbb{E}[R_{\text{test}}(\tilde{\beta})] \\ &\leq \mathbb{E}[R_{\text{train}}(\tilde{\beta})] \\ &= \mathbb{E}[(y_i - x_i^T \tilde{\beta})^2 | i > n] \end{aligned}$$

which proves the result.

(*) "Independent" is not correct here. The expected average test loss is independent of m :

$$\mathbb{E}[R_{\text{test}}(\tilde{\beta})] = \mathbb{E}[(y_i - x_i^T \tilde{\beta})^2 | i > n]$$

The expected training loss, can in fact be shown equal to:

$$\mathbb{E}[R_{\text{train}}(\hat{\beta})] = \mathbb{E}[(y_i - x_i^T \hat{\beta})^2 | i \leq n] = \sigma^2 \left(1 - \frac{p}{n}\right)$$

thus dependent of the sample size n .

What is meant is that the expected average loss is the same as the expected loss for a single observation, both in the test and training set

(**) The equality holds for $n=m$.

The inequality holds also for $n \neq m$, exactly because $\mathbb{E}[R_{\text{test}}(\tilde{\beta})]$ is independent of the test sample size m .

In particular, the expected value is the same in a test set of size n and m . We can therefore set $m=n$ to obtain the equality we need here.

We can write out this argument as follows:

$$\begin{aligned} \mathbb{E}[R_{\text{train}}(\hat{\beta})] &= \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n (y_i - x_i^T \hat{\beta})^2\right] \leq \\ &\mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n (y_i - x_i^T \tilde{\beta})^2\right] = \\ &\mathbb{E}[(y_i - x_i^T \tilde{\beta})^2 | i > n] = \mathbb{E}[R_{\text{test}}(\tilde{\beta})] \end{aligned}$$

← minimized RSS in a test set of size n

Exercise 2.2: Biased coefficients *

Suppose

$$y_i = \beta x_i + e_i$$

where $\beta = 2$ and $e_i \sim N(0,1)$.

You were not asked to show this, but the question was raised in class.

Then the least square estimate is given by

$$\hat{\beta} = \frac{x^T y}{x^T x} = \frac{\sum_i x_i y_i}{\sum_i x_i^2} = \frac{\sum_i x_i (2x_i + e_i)}{\sum_i x_i^2} = 2 + \frac{\sum_i x_i e_i}{\sum_i x_i^2}$$

where $\sum_i x_i e_i \approx 0$ since x_i and e_i is independent.

Now, consider regressing the same x_i 's on the same y_i 's

$$x_i = \alpha y_i + u_i, \quad \alpha = \frac{1}{\beta}, \quad u_i = -\frac{e_i}{\beta}$$

Then

$$\hat{\alpha} = \frac{\sum_i x_i y_i}{\sum_i y_i^2} = \frac{\sum_i (\alpha y_i + u_i) y_i}{\sum_i y_i^2}$$

$$= \alpha + \frac{\sum_i u_i y_i}{\sum_i y_i^2}$$

$$= \alpha - \frac{1}{\beta} \frac{\sum_i e_i (2x_i + e_i)}{\sum_i y_i^2}$$

$$= \alpha - \frac{1}{\beta} \left[\underbrace{\frac{2 \sum_i x_i e_i}{\sum_i y_i^2}}_{\approx 0} + \underbrace{\frac{\sum_i e_i^2}{\sum_i y_i^2}}_{> 0} \right]$$

This explains the downward bias in our α estimates, which stem from the regressor y_i being correlated with the residual x_i .