Exercises 1,4

Group 2

2023-10-18

Exercise 1

Expected Training Error:

Firstly, we want to derive the expected training error. For simplicity, we carry out the decomposition by firstly taking into consideration the training error only.

We expand the expression by adding 0 to the expression:

$$\overline{Err} = \frac{1}{N}\left(y_\tau - X\beta_\tau\right)^T\left(y_\tau - X\beta_\tau\right) =$$
$$\frac{1}{N}\left((y_\tau - X\beta) + (X\beta - E[X\beta_\tau]) + (E[X\beta_\tau] - X\beta_\tau)\right)^T\left((y_\tau - X\beta) + (X\beta - E[X\beta_\tau]) + (E[X\beta_\tau] - X\beta_\tau)\right) =$$

Now we observe that this expression gives us 6 terms. Of course, the expectation is a linear operator, so we have:

- The irreducible error (1):

$$\frac{1}{N}E[(y_\tau - X\beta)^T(y_\tau - X\beta)] = \sigma_\epsilon^2$$

Which reduces to σ_ϵ^2, because we assume having N observations in the training set.

- The bias term (2):

$$\frac{1}{N}E[(X\beta - E[X\beta_\tau])^T(X\beta - E[X\beta_\tau])]$$

- The variance term (3):

$$\frac{1}{N}E[(X\beta - E[X\beta_\tau])^T(X\beta - E[X\beta_\tau])]$$

- Cross product (4):

$$\frac{2}{N}E[(y_\tau - X\beta)^T(X\beta - E[X\beta_\tau])]$$

For this term, we note that $E[y_\tau] = X\beta$, hence the expression goes to 0.

- Cross product (5):

$$\frac{2}{N}E[(X\beta - E[X\beta_\tau])^T(E[X\beta_\tau] - X\beta_\tau)] =$$
$$\frac{2}{N}(X\beta - E[X\beta_\tau])^T E[(E[X\beta_\tau] - X\beta_\tau)]$$

This cross-product also goes to 0, since the term $(X\beta - E[X\beta_\tau])$ is a constant, and $E[(E[X\beta_\tau] - X\beta_\tau)] = E[X\beta_\tau] - E[X\beta_\tau] = 0$

- Cross product (6):

$$\frac{2}{N}E[(y_\tau - X\beta)^T(E[X\beta_\tau] - X\beta_\tau)]$$

For this expression we recognize that both of these terms are differences between what we have in the sample $y_\tau, X\beta_\tau$, and their expectations. Therefore, we can rewrite this as:

$$\frac{2}{N}\ cov(y_\tau, X\beta_\tau) =$$
$$\frac{2}{N}cov(y_\tau, X(X^TX)^{-1}X^Ty_\tau) =$$
$$\frac{2}{N}X(X^TX)^{-1}X^Tcov(y_\tau, y_\tau) =$$
$$\frac{2}{N}X(X^TX)^{-1}X^T\sigma_\epsilon^2$$

The total expression will be:

$$\sigma_\epsilon^2 + \frac{1}{N}E[(X\beta - E[X\beta_\tau])^T(X\beta - E[X\beta_\tau])]$$
$$+\frac{1}{N}E[(X\beta - E[X\beta_\tau])^T(X\beta - E[X\beta_\tau])] + \frac{2}{N}(X(X^TX)^{-1}X^T\sigma_\epsilon^2)$$

Expected In-Sample Error

The story here is the same, with the difference that we take the expectation with relation of new data points which we never observed, and therefore they are **independent** from the training observations we have. Hence, we have a similar derivation, where the only differences are observed in the terms (1), (4) and (6). Since (4) goes to 0 anyway for this expression, we focus on (1) and (6).

- For term (1*) we have:

$$\frac{1}{N}E[E_{y^0}[(y^0 - X\beta)^T(y^0 - X\beta)]$$

Where y^0 is the vector of independent test values we get.

- Term (6*):

Let's just take this without expectation over all training sets:

$$\frac{2}{N}E_{y^0}[(y^0 - X\beta)^T(E[X_0\beta_\tau] - X\beta_\tau)] = 0$$

Since y^0 is independent from $X\beta_\tau$, and $E[y_0] = X\beta$.

The final expression would be:

$$E[err_{in}] = \sigma_\epsilon^2 + \frac{1}{N}E[(X^0\beta - E[X^0\beta_\tau])^T(X^0\beta - E[X^0\beta_\tau])] + \frac{1}{N}E[(X^0\beta - E[X^0\beta_\tau])^T(X^0\beta - E[X^0\beta_\tau])]$$

The expected difference:

We see that terms (1), (2), (3), are the same. Hence, for a linear model the expected difference between the in-sample errors and the training errors is the following:

$$E[Err_{in} - \overline{err}] = \frac{2}{N}X(X^TX)^{-1}X^T\sigma_\epsilon^2$$

Exercise 4

If **y** arises from an additive-error model $Y = f(X) + \epsilon$ with $Var(\epsilon) = \sigma_\epsilon^2$ and where $\hat{y} = Sy$, then one can show that:

$$\sum_{i=1}^{N} cov(\hat{y}, y) = trace(S)\sigma_\epsilon^2$$

First we can figure out that:

$$\sum_{i=1}^{N} cov(\hat{y}, y) = trace(\Sigma)$$

Therefore:

$$trace(\Sigma) = trace(cov(\hat{y}, y))$$
$$trace(\Sigma) = trace(cov(Sy, y)) = trace(S\sigma_y^2) = S\ trace(\sigma_\epsilon^2)$$