Exercises 1,4

Group 2

2023-10-18

Exercise 1

In this exercise we want to quantify the expected in-sample error, the expected training error and the difference between the two. Firstly, as we derived from **exercise 3**, we have that

$$E[Err_{in} - \overline{err}] = E[op] = \frac{2}{N}\sum_{i=1}^{N} Cov(\hat{y}_i, y_i)$$

In relation to $E[op]$, in **exercise 4** we have a very similar point to make. For a linear model with the assumptions given in the exercise, we have that the sum over the diagonal elements is the trace of the covariance matrix. Recalling $\hat{y}_i = X(X^TX)^{-1}X^Ty$:

$$\begin{aligned}
&E[op] = \frac{2}{N}\sum_{i=1}^{N} Cov(\hat{y}_i, y_i) = \\
&\frac{2}{N}\, trace(X(X^TX)^{-1}X^T)\ \sigma_\epsilon^2 = \\
&\frac{2}{N}\, trace((X^TX)(X^TX)^{-1})\ \sigma_\epsilon^2 = \\
&\frac{2}{N}\, trace(I_d)\ \sigma_\epsilon^2 = \frac{2d\sigma_\epsilon^2}{N}
\end{aligned}$$

So, this is the actual difference between the Expected in-sample error and the Expected training error. We now turn to deriving the Expected in-sample error. For simplicity in notation we define $H = X(X^TX)^{-1}X^T$ as the hat matrix.

Now we turn to deriving to the task of deriving the expected in-sample error. Note y_0 denote new observations which are independent from the training dataset observations. Then $\hat{y}_0$ is just the estimated value conditional to the training dataset that we have.

$$\begin{aligned}
&E[Err_{in}] = \frac{1}{N}\left(y_0 - \hat{y}_0\right)^T\left(y_0 - \hat{y}_0\right) = \\
&\frac{1}{N}(\epsilon^T(H-I)^T(H-I)\epsilon)
\end{aligned}$$

Now use that $(H - I)$ is idempotent, since these are the matrices that give the orthogonal projections. Furthermore, since H is symmetric $I - H$ is also symmetric.

$$\begin{aligned}
&E[Err_{in}] = \frac{1}{N}(\epsilon^T(H-I)^T(H-I)\epsilon) = \\
&\frac{1}{N}(\epsilon^T(I-H)\epsilon) = \frac{1}{N}\epsilon^T\epsilon - \frac{1}{N}\epsilon^T H\epsilon
\end{aligned}$$

This is the in-sample error given a training set we have. Now we take the average over all training sets:

$$E_\tau[E[Err_{in}]] = \frac{1}{N}\left(E_\tau[\epsilon^T\epsilon] - E_\tau[\epsilon^T H\epsilon]\right) = \sigma_\epsilon^2 - \frac{\sigma_\epsilon^2\ (d+1))}{N}$$

This is because:

$$E_\tau[\epsilon^T\epsilon] = E_\tau[e_1^2 + e_2^2 + \cdots + e_N^2] = N\sigma_\epsilon^2 \quad (1)$$
$$E_\tau[\epsilon^T H\epsilon] = E_\tau\left[\sum_{i=1}^N H_{ii}e_i^2 + \sum_{i\neq j}^N H_{ij}\epsilon_i\epsilon_j\right] = trace(H)\ \sigma^2 + 0 \quad (2)$$

For (2) we can say the errors are assumed to be independent. We now derived the expected in-sample error to be:

$$E_\tau[E[Err_{in}]] = \sigma_\epsilon^2\left(1 - \frac{(d+1)}{N}\right)$$

The Expected training error is:

$$E[\overline{err}] = E_\tau[Err_{in}] - E[op] = \sigma_\epsilon^2\left(1 - \frac{(d+1)}{N}\right) - \frac{2d\sigma_\epsilon^2}{N} = \sigma_\epsilon^2\left(1 - \frac{(1-d)}{N}\right)$$

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:

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

Solution:

The term $\sum_{i=1}^N cov(\hat{y}_i, y_i)$ represents the sum of the diagonal values of the covariance matrix, which is therefore the trace of the covariance matrix. Furthermore, let's consider $cov(\hat{y}, y)$ for vectors y and $\hat{y}$. We have that:

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

Adding the two observations together we have that:

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