FYS-STK4155
week37 exercises

Yevhenii Volkov

September 17, 2023

Exercise 1

In this exercise it is needed to find expectations value and variance of the output y_i and of the optimal parameters $\hat{\beta}$.
Let's bring up all expressions:

$$y = f(x) + \varepsilon, \tag{1}$$

$$f(x) \approx \tilde{y}, \tag{2}$$

$$\tilde{y} = X\beta. \tag{3}$$

We can summarize equations (1), (2), (3) in one equation

$$y = X\beta + \varepsilon. \tag{4}$$

Here $X\beta$ is non stochastic value, so $\mathbb{E}[X\beta] = X\beta$.
In equation (1) ε is the noise, that follows Normal distribution with expectation value 0 and variance σ: $\varepsilon \sim N(0, \sigma)$.

Firstly, let's find analytical expression for the expectation value of output y:

$$\begin{aligned} \mathbb{E}[y_i] &= \mathbb{E}[X_{*,i}\beta + \varepsilon_i] = \mathbb{E}[X_{*,i}\beta] + \mathbb{E}[\varepsilon_i] \\ &= X_{*,i}\beta + 0 = X_{*,i}\beta. \end{aligned}$$

Second equality implies from the fact that $\mathbb{E}[a+b] = \mathbb{E}[a] + \mathbb{E}[b]$, third equality implies from the fact, that expectation values of noise equals 0.

Secondly, find the variance of y_i:

$$\begin{aligned} \mathrm{var}(y_i) &= \mathbb{E}\left[(y_i - \mathbb{E}[y_i])^2\right] = \mathbb{E}\left[\left(y_i^2 - 2y_i\mathbb{E}[y_i] + (\mathbb{E}[y_i])^2\right)\right] \\ &= \mathbb{E}[y_i^2] - 2\mathbb{E}[y_i] \cdot \mathbb{E}[y_i] + (\mathbb{E}[y_i])^2 = \mathbb{E}[y_i^2] - (\mathbb{E}[y_i])^2 \\ &= \mathbb{E}[(X_{*,i}\beta + \varepsilon_i)^2] - (X_{*,i}\beta)^2 \\ &= \mathbb{E}[(X_{*,i}\beta)^2 + 2X_{*,i}\beta\varepsilon_i + \varepsilon_i^2] - (X_{*,i}\beta)^2 \\ &= (X_{*,i}\beta)^2 + 2X_{*,i}\beta\mathbb{E}[\varepsilon_i] + \mathbb{E}[\varepsilon_i^2] - (X_{*,i}\beta)^2 \\ &= \mathbb{E}[\varepsilon_i^2] = \sigma^2. \end{aligned}$$

First equality $-$ rewriting $(...)^2$,
Second equality $-$ rewriting $\mathbb{E}[...]$,
Fourth equality $-$rewriting $(...)^2$,
Fifth equality $-$ rewriting $\mathbb{E}[...]$,
Fifth equality $-$ implies from the facts that $\mathbb{E}[\varepsilon_i] = 0,\ \mathbb{E}[\varepsilon_i^2] = \sigma$.

Now, compute expected value of $\hat{\beta}$, where $\hat{\beta}$ is given by expression

$$\hat{\beta} = (X^T X)^{-1} X^T y. \tag{5}$$

$$\begin{aligned}\mathbb{E}[\hat{\beta}] &= \mathbb{E}[(X^T X)^{-1} X^T y] = (X^T X)^{-1} X^T \mathbb{E}[y] \\ &= (X^T X)^{-1} X^T X \beta = \beta.\end{aligned}$$

Third equality implies from the expression for $\mathbb{E}[y]$, which we computed earlier.

Finally, variance of $\hat{\beta}$

$$\begin{aligned}\mathrm{var}(\hat{\beta}) &= \mathbb{E}[(\beta - \mathbb{E}[\beta])(\beta - \mathbb{E}[\beta])^T] \\ &= \mathbb{E}[((X^T X)^{-1} X^T y - \mathbb{E}[\beta])((X^T X)^{-1} X^T y - \mathbb{E}[\beta])^T] \\ &= (X^T X)^{-1} X^T \mathbb{E}[y y^T] X (X^T X)^{-1} - \beta\beta^T \\ &= (X^T X)^{-1} X^T (X \beta \beta^T X^T + \sigma^2) X (X^T X)^{-1} - \beta\beta^T \\ &= \beta\beta^T + \sigma^2 (X^T X)^{-1} - \beta\beta^T \\ &= \sigma^2 (X^T X)^{-1}.\end{aligned}$$

Second equality $-$ replaced beta using expression (5),
Third equality is obtained using the facts that $\mathbb{E}[\beta] = \beta$ and two terms cancel each other,
Fourth equality $-$ expanding $\mathbb{E}[y y^T]$.

1 Exercise 2

Now we want to find similar expression for the variance and expectation value for the optimal parameters β for Ridge regression.
The steps are very similar, but now β is given by expression

$$\hat{\beta} = (X^T X + \lambda I)^{-1} X^T y. \tag{6}$$

Expectation value of β

$$\begin{aligned}\mathbb{E}[\hat{\beta}] &= \mathbb{E}[(X^T X + \lambda I)^{-1} X^T y] = (X^T X + \lambda I)^{-1} X^T \mathbb{E}[y] \\ &= (X^T X + \lambda I)^{-1} X^T X \beta.\end{aligned}$$

We can see that $\mathbb{E}[\hat{\beta}^{OLS}] \neq \mathbb{E}[\hat{\beta}^{Ridge}]\ \forall\ \lambda$.

In the end, find the variance

$$
\begin{aligned}
\mathrm{var}(\hat{\beta}) &= \mathbb{E}[(\beta - \mathbb{E}[\beta])(\beta - \mathbb{E}[\beta])^T] \\
&= \mathbb{E}[((X^TX + \lambda I)^{-1}X^Ty - \mathbb{E}[\beta])((X^TX + \lambda I)^{-1}X^Ty - \mathbb{E}[\beta])^T] \\
&= (X^TX + \lambda I)^{-1}X^T\mathbb{E}[yy^T]X(X^TX + \lambda I)^{-1} - \beta\beta^T \\
&= (X^TX + \lambda I)^{-1}X^T(X\beta\beta^TX^T + \sigma^2)X(X^TX + \lambda I)^{-1} - \beta\beta^T \\
&= \sigma^2(X^TX + \lambda I)^{-1}X^TX((X^TX + \lambda I)^{-1})^T.
\end{aligned}
$$

Because of the $(... + \lambda I)$ term we can not simplify the expression, but at least we have an analytical expression.