Exercises Week 37

Oskar Idland

Exercise 1

As we are working with a normally distributed error $\boldsymbol{\varepsilon} \sim N(0, \sigma^2)$ we know the expectation value of the i'th element of $\boldsymbol{\varepsilon}$ is 0 with a variance of σ^2. With our approximation of $f(\mathbf{x})$ as $\tilde{\mathbf{y}} = \mathbf{X}\boldsymbol{\beta}$, we get an expectation value of the following:

$$\mathbb{E}(\mathbf{y}) = \mathbb{E}(\mathbf{X}\boldsymbol{\beta}) + \mathbb{E}(\boldsymbol{\varepsilon}) = \mathbb{E}(\mathbf{X}\boldsymbol{\beta})$$

The i'th element of $\mathbf{y}$ has an expectation value of the following:

$$\mathbb{E}(y_i) = \mathbb{E}\left(\sum_j X_{ij}\boldsymbol{\beta}_j\right) = \sum_j X_{ij}\boldsymbol{\beta}_j = \mathbf{X}_{i,*}\boldsymbol{\beta}$$

The variation becomes:

$$\mathrm{Var}(y_i) = \mathrm{Var}\left(\mathbf{X}_{i*}\boldsymbol{\beta}\right) + \mathrm{Var}(\boldsymbol{\varepsilon}_i) = \sigma^2$$

With this we know that:

$$y_i \sim N(\mathbf{X}_{i,*\boldsymbol{\beta},\sigma^2})$$

In the OLS, we know the optimal parameters to be:

$$\boldsymbol{\beta}_{\mathrm{OLS}} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}$$

with an expectation value of:

$$\mathbb{E}\left(\boldsymbol{\beta}_{\mathrm{OLS}}\right) = \mathbb{E}\left(\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}\right) = \mathbb{E}\left(\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\right)\mathbb{E}(\mathbf{y}) = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{X}\boldsymbol{\beta}$$

If x and y are independent variables, we know the variance of their product to be:

$$\begin{aligned}
\mathrm{Var}(xy) &= \mathbb{E}\left(x^2y^2\right) - \left(\mathbb{E}(xy)\right)^2 \\
&= \mathbb{E}(x^2)\mathbb{E}(y^2) - \left(\mathbb{E}(x)\right)^2\left(\mathbb{E}(y)\right)^2, \\
&= \left[\mathbb{E}(x^2) - \left(\mathbb{E}(x)\right)^2 + \left(\mathbb{E}(x)\right)^2\right]\left[\mathbb{E}(y^2) - \left(\mathbb{E}(y)\right)^2 + \left(\mathbb{E}(y)\right)^2\right] - \mathbb{E}(x^2)\mathbb{E}(y^2), \\
&= \left[\mathrm{Var}(x) + \left(\mathbb{E}(x)\right)^2\right]\left[\mathrm{Var}(y) + \left(\mathbb{E}(y)\right)^2\right] - \mathbb{E}(x^2)\mathbb{E}(y^2), \\
&= \mathrm{Var}(x)\mathrm{Var}(y) + \mathrm{Var}(x)(\mathbb{E}(y))^2 + \mathrm{Var}(y)(\mathbb{E}(x))^2 + \mathbb{E}(x^2)\mathbb{E}(y^2) - \mathbb{E}(x^2)\mathbb{E}(y^2), \\
&= \mathrm{Var}(x)\mathrm{Var}(y) + \mathrm{Var}(x)(\mathbb{E}(y))^2 + \mathrm{Var}(y)(\mathbb{E}(x))^2
\end{aligned}$$

Inserting $x = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T$ and $y = \mathbf{y}$, we can find the variance of $\boldsymbol{\beta}_{\mathrm{OLS}}$. We use the fact that $\left(\mathbf{X}^T\mathbf{X}\right)^{-1}$ is itself as the matrix is square and symmetric.

$$\begin{aligned}
\mathrm{Var}(\beta_{\mathrm{OLS}}) &= \underbrace{\mathrm{Var}\left[\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\right]\mathrm{Var}(\mathbf{y})}_{0} + \underbrace{\mathrm{Var}\left[\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\right](\mathbb{E}(\mathbf{y}))^2}_{0} + \mathrm{Var}(\mathbf{y})\left(\mathbb{E}\left[\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\right]\right)^2 \\
&= \sigma^2\left[\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\right]^2 \\
&= \sigma^2\left[\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\right]\left[\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\right]^{\mathrm{T}} \\
&= \sigma^2\left[\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\right]\left[\mathbf{X}\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\right] \\
&= \sigma^2\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}
\end{aligned}$$

Exercise 2

By using the same approach as in the previous exercise, we can show that

$$\mathbb{E}(\mathbf{y}) = \mathbb{E}(\mathbf{X}\boldsymbol{\beta}) = \mathbf{X}\boldsymbol{\beta}$$

and that

$$\operatorname{Var}(\mathbf{y}) = \operatorname{Var}(\boldsymbol{\varepsilon}) = \sigma^2$$

for the ridge regression as well as the the OLS. We use the same method again:

$$\mathbb{E}(\boldsymbol{\beta}_{\text{Ridge}}) = \mathbb{E}\left(\left(\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}_{pp}\right)^{-1}\mathbf{X}^T\mathbf{y}\right) = \mathbb{E}\left(\left(\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}_{pp}\right)^{-1}\mathbf{X}^T\right)\mathbb{E}(\mathbf{y}) = \left(\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}_{pp}\right)^{-1}\mathbf{X}^T\mathbf{X}\boldsymbol{\beta}$$

The variance then becomes:

$$\begin{aligned}
\operatorname{Var}(\beta_{\text{Ridge}}) &= \underbrace{\operatorname{Var}\left(\left(\mathbf{X}^{\mathrm{T}}\mathbf{X} + \lambda\mathbf{I}_{p\times p}\right)^{-1}\mathbf{X}^{\mathrm{T}}\right)\operatorname{Var}(\mathbf{y})}_{0} + \underbrace{\operatorname{Var}\left(\left(\mathbf{X}^{\mathrm{T}}\mathbf{X} + \lambda\mathbf{I}_{p\times p}\right)^{-1}\mathbf{X}^{\mathrm{T}}\right)(\mathbb{E}(\mathbf{y}))^2}_{0} + \operatorname{Var}(\mathbf{y})\left(\mathbb{E}\left(\left(\mathbf{X}^{\mathrm{T}}\mathbf{X} + \lambda\mathbf{I}_{p\text{:}}\right.\right.\right. \\
&= \sigma^2\left(\left(\mathbf{X}^{\mathrm{T}}\mathbf{X} + \lambda\mathbf{I}_{p\times p}\right)^{-1}\mathbf{X}^{\mathrm{T}}\right)^2 \\
&= \sigma^2\left(\left(\mathbf{X}^{\mathrm{T}}\mathbf{X} + \lambda\mathbf{I}_{p\times p}\right)^{-1}\mathbf{X}^{\mathrm{T}}\right)\left(\left(\mathbf{X}^{\mathrm{T}}\mathbf{X} + \lambda\mathbf{I}_{p\times p}\right)^{-1}\mathbf{X}^{\mathrm{T}}\right)^{\mathrm{T}} \\
&= \sigma^2\left(\left(\mathbf{X}^{\mathrm{T}}\mathbf{X} + \lambda\mathbf{I}_{p\times p}\right)^{-1}\mathbf{X}^{\mathrm{T}}\right)\left(\mathbf{X}\left\{\left(\mathbf{X}^{\mathrm{T}}\mathbf{X} + \lambda\mathbf{I}_{p\times p}\right)^{-1}\right\}^{\mathrm{T}}\right) \\
&= \sigma^2\left(\mathbf{X}^{\mathrm{T}}\mathbf{X} + \lambda\mathbf{I}_{p\times p}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{X}\left(\left(\mathbf{X}^{\mathrm{T}}\mathbf{X} + \lambda\mathbf{I}_{p\times p}\right)^{-1}\right)^{\mathrm{T}}
\end{aligned}$$

$$\boldsymbol{\varepsilon}$$