University of Oslo Fall, 2023
Department of Mathematics Weekly Exercises: Week 38

Author: Jonas Semprini Næss

FYS-STK4155: Applied Data Analysis and Machine Learning

Analytical Exercises:

(I.) *The parameters $\boldsymbol{\beta}$ are in turn found by optimizing the mean squared error via the so-called cost function*

$$C(\boldsymbol{X},\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=0}^{n-1}(y_i - \tilde{y}_i)^2 = \mathbb{E}\left[(\boldsymbol{y}-\tilde{\boldsymbol{y}})^2\right].$$

Here the expected value $\mathbb{E}$ is the sample value. Show that you can rewrite this in terms of a term which contains the variance of the model itself (the so-called variance term), a term which measures the deviation from the true data and the mean value of the model (the bias term) and finally the variance of the noise. That is, show that

$$\mathbb{E}\left[(\boldsymbol{y}-\tilde{\boldsymbol{y}})^2\right] = \mathrm{Bias}[\tilde{y}] + \mathrm{var}[\tilde{y}] + \sigma^2,$$

Solution:

We start of by rewriting $\mathbb{E}\left[(\boldsymbol{y}-\tilde{\boldsymbol{y}})^2\right]$ to $\mathbb{E}\left[(\boldsymbol{f}+\boldsymbol{\epsilon}-\tilde{\boldsymbol{y}})^2\right]$. From this one gets that

$$\begin{aligned}
\mathbb{E}[(\boldsymbol{y}-\tilde{\boldsymbol{y}})^2] &= \mathbb{E}[(\boldsymbol{f}+\boldsymbol{\varepsilon}-\tilde{\boldsymbol{y}})^2] \\
&= \mathbb{E}[(\boldsymbol{f}-\tilde{\boldsymbol{y}})+\boldsymbol{\varepsilon})^2] \\
&= \mathbb{E}[(\boldsymbol{f}-\tilde{\boldsymbol{y}})^2 + 2(\boldsymbol{f}-\tilde{\boldsymbol{y}})\boldsymbol{\varepsilon} + \boldsymbol{\varepsilon}^2] \\
&= \mathbb{E}[(\boldsymbol{f}-\tilde{\boldsymbol{y}})^2] + 2\mathbb{E}[(\boldsymbol{f}-\tilde{\boldsymbol{y}})\underbrace{\boldsymbol{\varepsilon}}_{=0}] + \underbrace{\mathbb{E}[\boldsymbol{\varepsilon}^2]}_{=\sigma^2} \\
&= \mathbb{E}[(\boldsymbol{f}-\tilde{\boldsymbol{y}})^2] + \sigma^2.
\end{aligned}$$

By adding and subtracting $\mathbb{E}[\tilde{\boldsymbol{y}}]$ we obtain

$$\begin{aligned}
\mathbb{E}[(\boldsymbol{y}-\tilde{\boldsymbol{y}})^2] &= \mathbb{E}[(\boldsymbol{f}-\tilde{\boldsymbol{y}}+\mathbb{E}[\tilde{\boldsymbol{y}}]-\mathbb{E}[\tilde{\boldsymbol{y}}])^2] + \sigma^2 \\
&= \mathbb{E}[((\boldsymbol{f}-\mathbb{E}[\tilde{\boldsymbol{y}}]) + (\mathbb{E}[\tilde{\boldsymbol{y}}]-\tilde{\boldsymbol{y}}))^2] + \sigma^2 \\
&= \mathbb{E}[(\boldsymbol{f}-\mathbb{E}[\tilde{\boldsymbol{y}}])^2 + 2(\boldsymbol{f}-\mathbb{E}[\tilde{\boldsymbol{y}}])(\mathbb{E}[\tilde{\boldsymbol{y}}]-\tilde{\boldsymbol{y}}) + (\mathbb{E}[\tilde{\boldsymbol{y}}]-\tilde{\boldsymbol{y}})^2] + \sigma^2 \\
&= \mathbb{E}[(\boldsymbol{f}-\mathbb{E}[\tilde{\boldsymbol{y}}])^2 + (\mathbb{E}[\tilde{\boldsymbol{y}}]-\tilde{\boldsymbol{y}})^2] + 2\mathbb{E}[(\boldsymbol{f}-\mathbb{E}[\tilde{\boldsymbol{y}}])(\mathbb{E}[\tilde{\boldsymbol{y}}]-\tilde{\boldsymbol{y}})] + \sigma^2
\end{aligned}$$

where

$$\begin{aligned}
\mathbb{E}[(\boldsymbol{f}-\mathbb{E}[\tilde{\boldsymbol{y}}])(\mathbb{E}[\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{y}})] &= \mathbb{E}[\boldsymbol{f}\mathbb{E}[\tilde{\boldsymbol{y}}] - \boldsymbol{f}\tilde{\boldsymbol{y}} - \mathbb{E}[\tilde{\boldsymbol{y}}]^2 + \tilde{\boldsymbol{y}}\mathbb{E}[\tilde{\boldsymbol{y}}]] \\
&= \boldsymbol{f}\mathbb{E}[\tilde{\boldsymbol{y}}] - \boldsymbol{f}\mathbb{E}[\tilde{\boldsymbol{y}}] - \mathbb{E}[\tilde{\boldsymbol{y}}]^2 + \mathbb{E}[\tilde{\boldsymbol{y}}]^2 \qquad (1) \\
&= 0
\end{aligned}$$

Thus, are we left with

$$
\begin{aligned}
\mathbb{E}[(\boldsymbol{y}-\tilde{\boldsymbol{y}})^2] &= \mathbb{E}[(\boldsymbol{f}-\mathbb{E}[\tilde{\boldsymbol{y}}])^2 + (\mathbb{E}[\tilde{\boldsymbol{y}}]-\tilde{\boldsymbol{y}})^2] + \sigma^2 \\
&= \underbrace{\mathbb{E}[(\boldsymbol{f}-\mathbb{E}[\tilde{\boldsymbol{y}}])^2]}_{\mathrm{Bias}[\tilde{y}]} + \underbrace{\mathbb{E}[(\tilde{\boldsymbol{y}}-\mathbb{E}[\tilde{\boldsymbol{y}}])^2]}_{\mathrm{Var}[\tilde{y}]} + \sigma^2 \\
&= \mathrm{Bias}[\tilde{y}] + \mathrm{Var}[\tilde{y}] + \sigma^2.
\end{aligned}
$$

Which is what we wanted to show. ■