

Author: Jonas Semprini Næss

FYS-STK4155: Applied Data Analysis and Machine Learning

Analytical Exercises:

(I.) The parameters β are in turn found by optimizing the mean squared error via the so-called cost function

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

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

$$\mathbb{E}[(\mathbf{y} - \tilde{\mathbf{y}})^2] = \text{Bias}[\tilde{\mathbf{y}}] + \text{var}[\tilde{\mathbf{y}}] + \sigma^2,$$

Solution:

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

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

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

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

where

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

Thus, are wq left with

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

Which is what we wanted to show. ■