Weekly Exercises - FYS-STK3155

Oscar Atle Brovold

Week 38

Part i - Bias-variance decomposition of MSE

We want to show that MSE can be rewritten as follows

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

We can expand the MSE

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

We will now rewrite each term, first $\mathbb{E}[\boldsymbol{y}]$

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

Second $\mathbb{E}[\tilde{\boldsymbol{y}}^2]$

$$\mathbb{E}[\tilde{\boldsymbol{y}}^2] = \mathrm{Var}[\tilde{\boldsymbol{y}}] + \mathbb{E}[\tilde{\boldsymbol{y}}]^2$$

Last $\mathbb{E}[\boldsymbol{y}\tilde{\boldsymbol{y}}]$

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

Putting it all together we have

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

Where the last term is the bias squared, often just called bias (it will always be positive). In total we therefore have

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

Part ii - Discussion of bias and variance

Illustration of variance

In the case of high variance, our model is often overfitted. With sufficent complexity, the model can achive its goal of reducing the MSE by interpolating every data point. Consequently, the MSE for the training data converges to zero as the complexity increases. However, this can lead to issues when applying the model to test data. The model may become overly specific to the training data, struggling to make accurate prediction on new data/test data. This is illustrated in figure 1, where the same dataset is used, the distinction lies in which portions are assign to the training and test set. A sign of high variance and overfitting, is that the error varies significantly with small changes in the traning set, as demonstrated in figure 1.

Low variance, as illustrated in figure 2 is often a sign of underfitting. There is little variation in the predicted model, but the error remains significantly large.

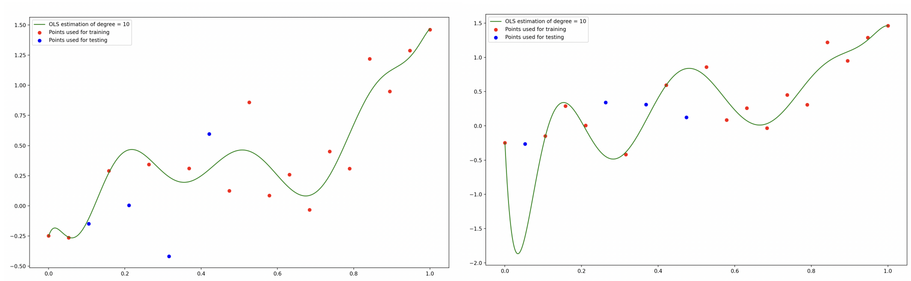

Figure 1: Overfitted model

Illustration of bias

When we look at the bias in the testing data we look at how consistently the model is making errors. A high bias would indicate that the model makes the same type of errors consistently. A good bias/low bias indicates that our model is able to caputure the patterns in the data.

Bias can also be a measure of how well the model can learn patterns in the training set. A high error in the traing set indicates a high bias, this is illustrated in figure 2. In figure 1, where the model was overfitted, the train error was low, indicating a low bias.

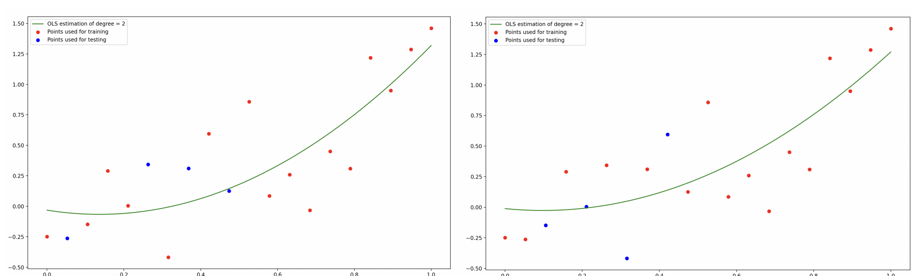

Figure 2: Underfitted model

Part iii - Bias-variance analysis of a simple one-dimensional

I will try to approximate the function

$$f(x) = e^{-x^2} + 1.5 \cdot e^{-(x-2)^2}$$

And perform a bias-variance analysis in relation to the complexity. I have also used non-parametric bootstrap, with B=100. To the function I have added a gaussian noise, that follows the distrobution

$$N \sim \mathcal{N}(0, 0.04)$$

Using OLS as the method of regression, I get the plot in figure 3.

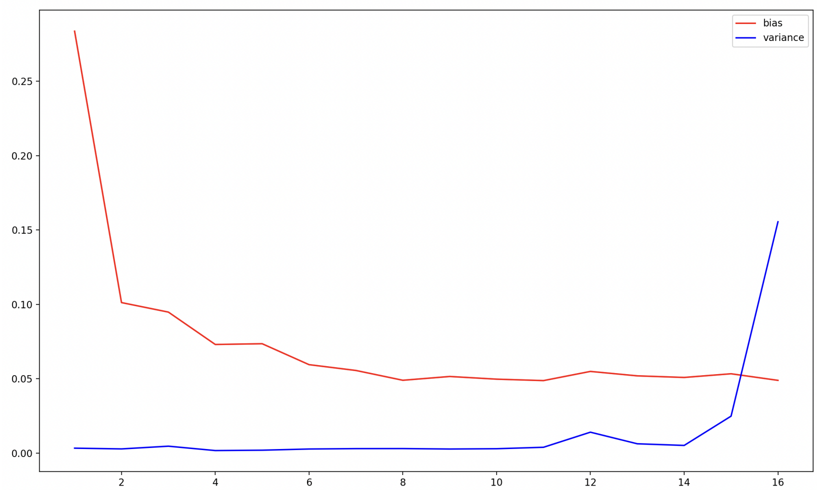

Figure 3: Bias-variance as function of the model complexity

The analysis caputures the discussion in part *ii*). When the model complexity is low, the model is underfitted. This leads to a high bias, meaning that the model stuggles to capture patterns in the data. At the same time, the variance is low because an underfitted model is very general, and changes to what data is included in the training set makes small diffrences in the model. As complexity increases the model reaches a point where both the bias and variance are relatively low. However, if complexity continues to increase, the variance starts to rise. As discussed in part *ii*, this is a sign of overfitting. Ideally we want a model with low bias and low variance. From figure 3 we can observe that the variance and bias reach a minimum around complexity level 8. This would typically be the choice for model complexity.

— End of Weekly Exercise —