

# **Project 1**

**Course: FYS-STK4155**

**Semester: Autumn 2020**

**Name: Sander Losnedahl**

## Exercise 1.

a) The goal for this part of the exercise is trying to fit a linear model (linear in terms of regression coefficients) to the Franke function given as

$$f(x, y) = \frac{3}{4} \exp \left( -\frac{(9x-2)^2}{4} - \frac{(9y-2)^2}{4} \right) + \frac{3}{4} \exp \left( -\frac{(9x+1)^2}{49} - \frac{(9y+1)^2}{10} \right) + \frac{1}{2} \exp \left( -\frac{(9x-7)^2}{4} - \frac{(9y-3)^2}{4} \right) - \frac{1}{5} \exp \left( -(9x-4)^2 - (9y-7)^2 \right) \quad (1)$$

where polynomial combinations of  $x$  and  $y$  in the span  $x, y \in [0, 1]$  will be the explanatory variables. The linear regression equation then takes the form

$$\widehat{f(x, y)} = \hat{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (2)$$

where  $\widehat{f(x, y)}$  or  $\hat{y}$  is the least squares estimate of the Franke function,  $\mathbf{X}$  is the  $n \times p$  design matrix consisting of the aforementioned polynomial combinations of  $x$  and  $y$ ,  $\boldsymbol{\beta}$  are the  $p \times 1$  regression coefficients and  $\boldsymbol{\epsilon}$  is just random noise/unobserved random variables. In order to get the best estimate for the Franke function, we want to choose  $\boldsymbol{\beta}$ -values so that we minimize the residual sum of squares (thereby the name "least squares"). Solving equation 2 with the intent to minimize the residual sum of squares yields the estimates for the  $\boldsymbol{\beta}$ -values

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T f(x, y) \quad (3)$$

where T marks the transpose of the matrix. We can now get the estimate  $\hat{\boldsymbol{\beta}}$  with equation 3 and then compute the estimate of the Franke function as seen in equation 2. There is always uncertainty when estimating  $\hat{\boldsymbol{\beta}}$ . These uncertainties can be quantified by finding the confidence interval (95% confidence intervals in this case) around each coefficient. The interval will then be given by  $\pm$  the standard deviation of the respective coefficients. In other words, we can be 95% sure that the true coefficient will lie in the span of  $[\beta_i - \sigma, \beta_i, \beta_i + \sigma]$ . We can investigate this further by plotting each coefficient calculated from the ordinary least squares (equation 3) with their respective confidence intervals as seen in figure 1

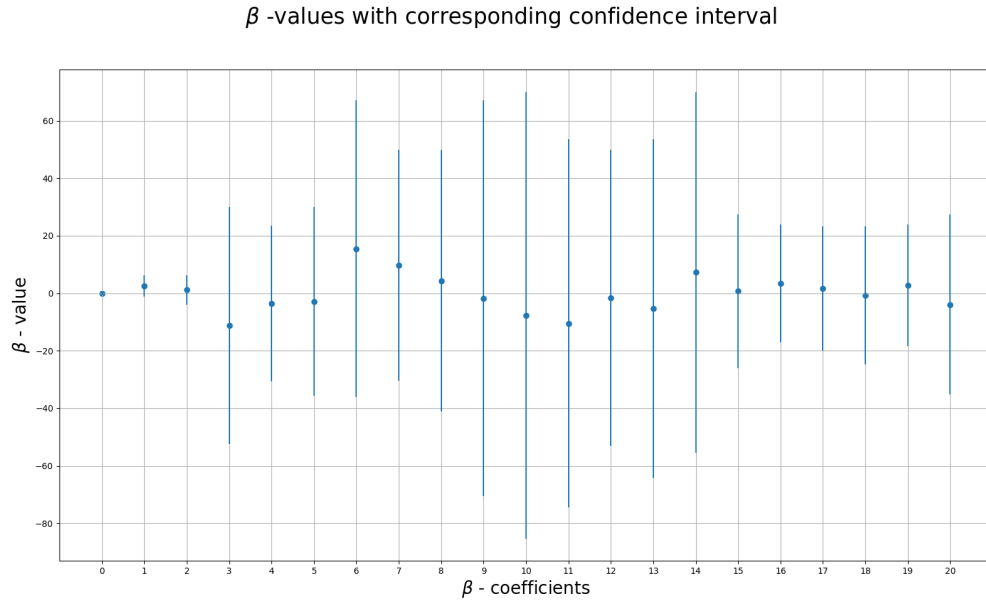


Figure 1:  $\hat{\beta}$ -coefficients from performing ordinary least squares regression Dots indicate the actual  $\hat{\beta}$ -coefficients value while bars around indicate the 95% confidence interval ( $\pm\sigma$ ).

From figure 1, it can be observed that some coefficients have relatively narrow confidence intervals, while others have relatively large. The larger the interval, the less certain we are of the actual estimate.

b) When we fit a linear model like in equation 2, we always want to minimize the mean square error (MSE). The MSE is often a good reference of the performance of our model and where lower MSE indicates a better model, and vice versa. The MSE is simply the expected value of the difference between the actual response and the predicted response using regression, which in math term is written as  $E[(y - \hat{y})^2]$  where  $y$  is the actual response while  $\hat{y}$  is the predicted response (as we already know). By adding and subtracting the term  $E[\hat{y}]$  to the inner bracket we can expand the MSE like in equation qqqqqq

$$\begin{aligned} MSE &= E[(y - \hat{y})^2] = E[(y - \hat{y} + E[\hat{y}] - E[\hat{y}])^2] \\ &= E[(\hat{y} - E[\hat{y}])^2 + 2(\hat{y} - E[\hat{y}])(E[\hat{y}] - y) + (E[\hat{y}] - y)^2] \\ &= E[(\hat{y} - E[\hat{y}])^2] + E[2(\hat{y} - E[\hat{y}])(E[\hat{y}] - y)] + E[(E[\hat{y}] - y)^2] \end{aligned} \quad (4)$$

Since the expected value of  $\hat{y}$  equals  $y$ , we can write  $E[\hat{y}] - y = \text{constant}$  and thus

$$\begin{aligned} &E[(\hat{y} - E[\hat{y}])^2] + 2(E[\hat{y}] - y)E[\hat{y} - E[\hat{y}]] + (E[\hat{y}] - y)^2 \\ &= E[(\hat{y} - E[\hat{y}])^2] + 2(E[\hat{y}] - y)(E[\hat{y}] - E[\hat{y}]) + (E[\hat{y}] - y)^2 \\ &= E[(\hat{y} - E[\hat{y}])^2] + (E[\hat{y}] - y)^2 \end{aligned} \quad (5)$$

where the constant  $E[\hat{y}] - E[\hat{y}]$  equals zero, making the whole term equal zero. We recognize the term  $E[(\hat{y} - E[\hat{y}])^2]$  as the variance of estimator  $\hat{y}$  and the term  $(E[\hat{y}] - y)^2$  as the difference between the bias, but squared. The bias quantifies how well the model fits the data points and variance is how well the model would translate to other data the model is not trained on. Increasing one tends to decrease the other, but not linearly since the bias is squared. Therefore, one can decrease the bias to a certain extent, but at some point, the loss of bias is not worth the gain in variance. This is called the bias variance trade-off. We can analyse the bias variance trade-off by plotting the MSE for different polynomial degrees (complexity of model) versus their respective MSE-values

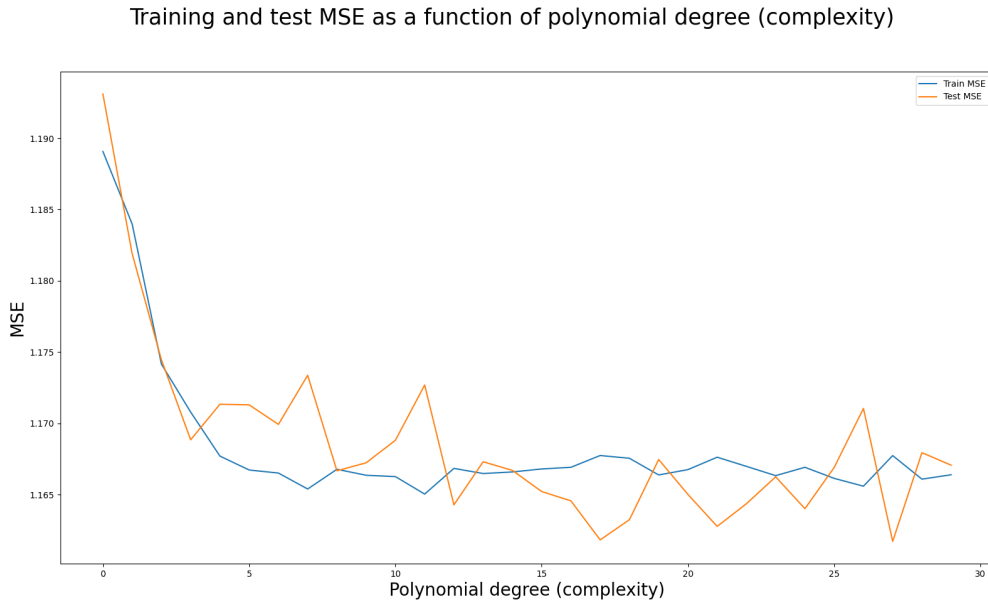


Figure 2: MSE (variance + bias<sup>2</sup> as a function of polynomial degree (complexity of model).

## References

- Reference