Methods

Ordinary Least Square (OLS) on the Franke's function.

In this exercise we are generating our own data set for the two-dimensional Franke's function defined for $x, y \in [0, 1]$. We will perform the Ordinary Least Square (OLS) regression analysis (to minimize the cost function, find the β parameters and fit the Franke's function) and use polynomials in x and y up to fifth order to verify how effective this OLS model is in fitting our input data.
We will use both versions of Franke's function, with and without added noise.

Scaling the data

Before we can create a machine learning model we need to scale the data. Scaling of date is considered one of the key aspects of a good machine learning model. We will be using the Standard scaler. Standardization transforms the data to have zero mean and a variance of 1, they make our data unit-less. We need to scale the data if there is a large difference in the range of our data. If we for example have a data set that ranges from values in the hundreds to values in tens of thousands the larger numbers will fully dominate our prediction model.

Train-Test Split

For this exercise we will be utilizing the Train-Test Split method. The purpose of this method is to estimate how well the model preforms on new data. As inferred in the name of the method we split our data into two different sets, the training set and the testing set. In order for this to be usefull we need a large enough amount of data to begin with or else the sets will be to small and we will not be able to accurately mirror the original set of data. This will in turn lead us to be unable to evaluate the model. There are other methods that are more suitable for smaller sets of data like K-folding which we will be using in a later exercise. So assuming we have enough data we will need to decide how much of the data will be used for training and how much that will be used for testing. There is not any particular optimal ratio to split the data, but its common to use about 80% for training and 20% for testing.

Bias-variance trade-off and resampling techniques

In this part we will study the bias-variance trade-off by implementing the bootstrap resampling technique.