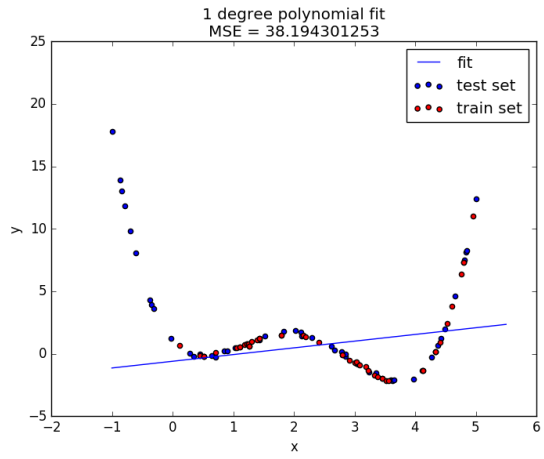
Antonio Norelli

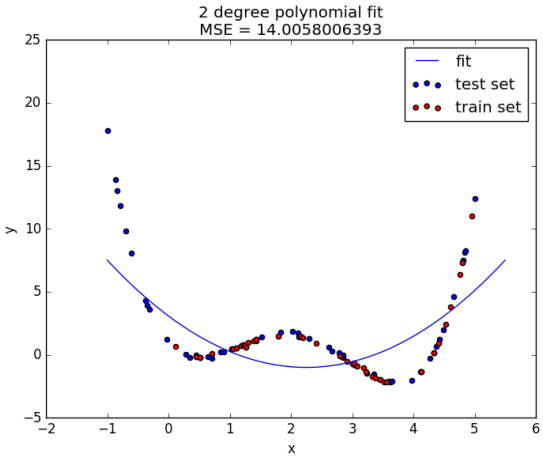
Machine learning *a.a 2016/2017* *prof. Barbara Caputo*

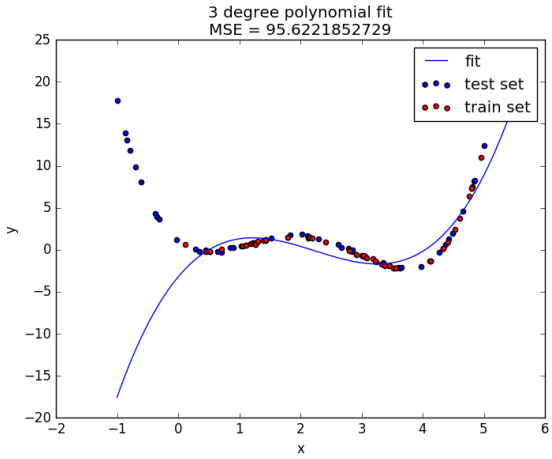
Homework 2: Regression

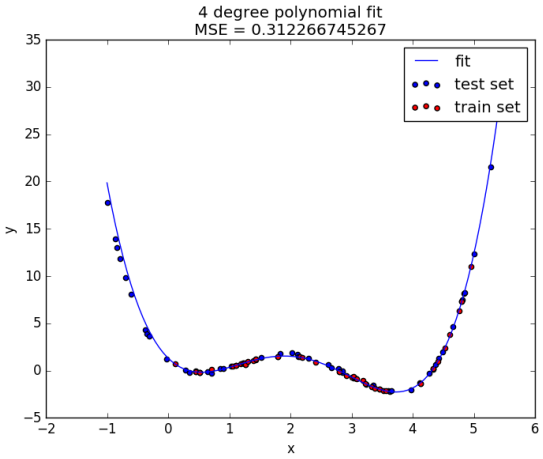
Polynomial fits of degree n, n=1:10

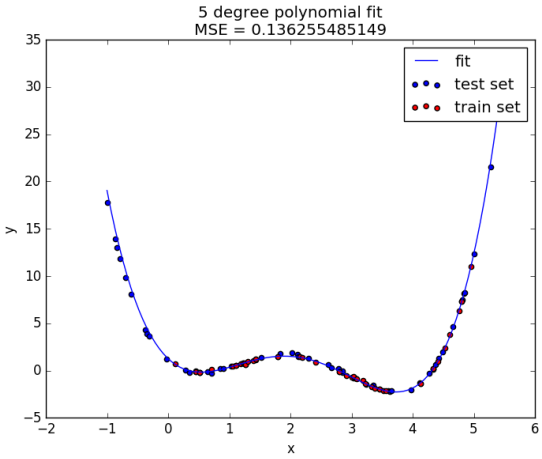
In the following plots are shown the train set, the test set and the polynomial fit. In every title is reported the mean squared error (MSE).

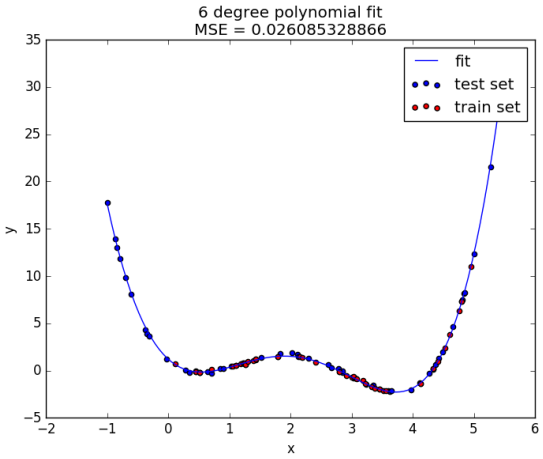


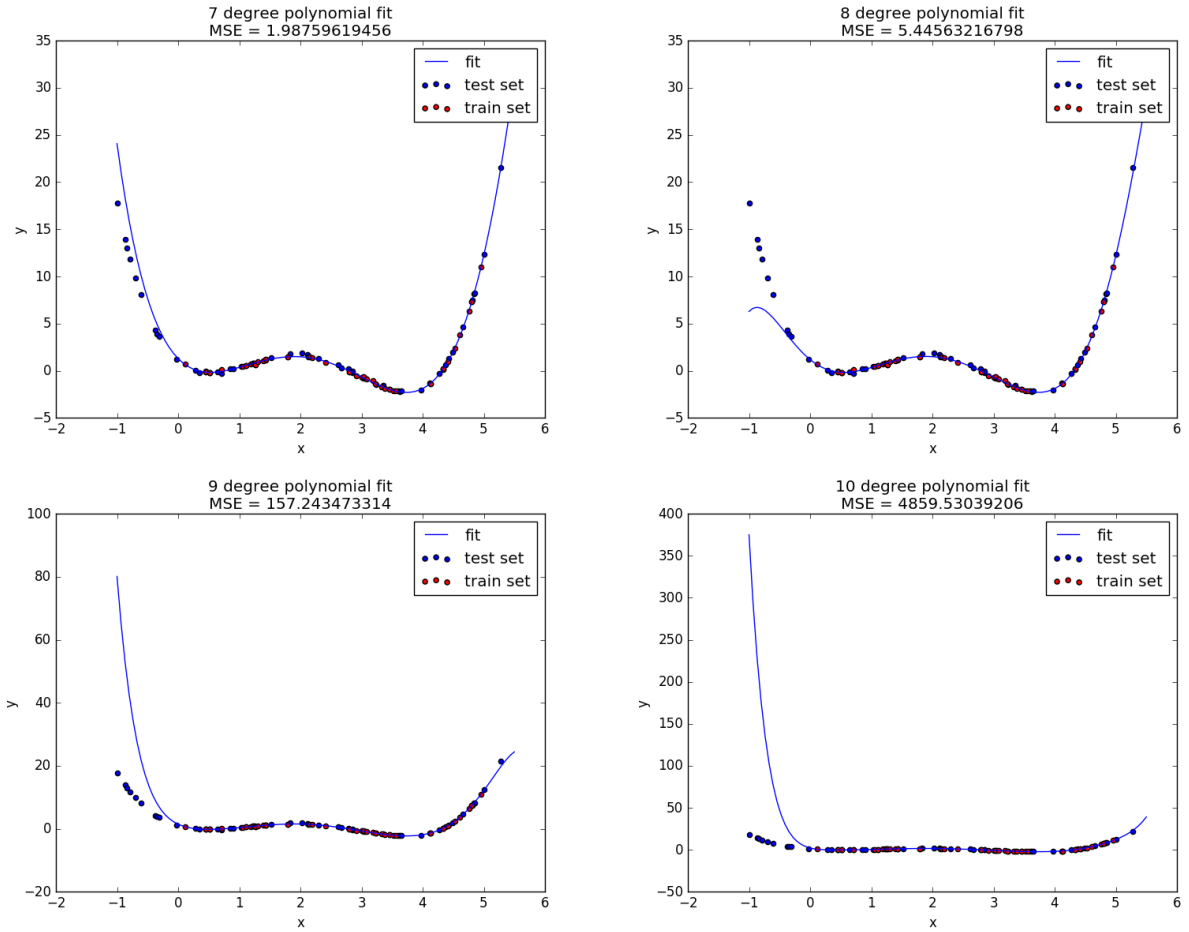






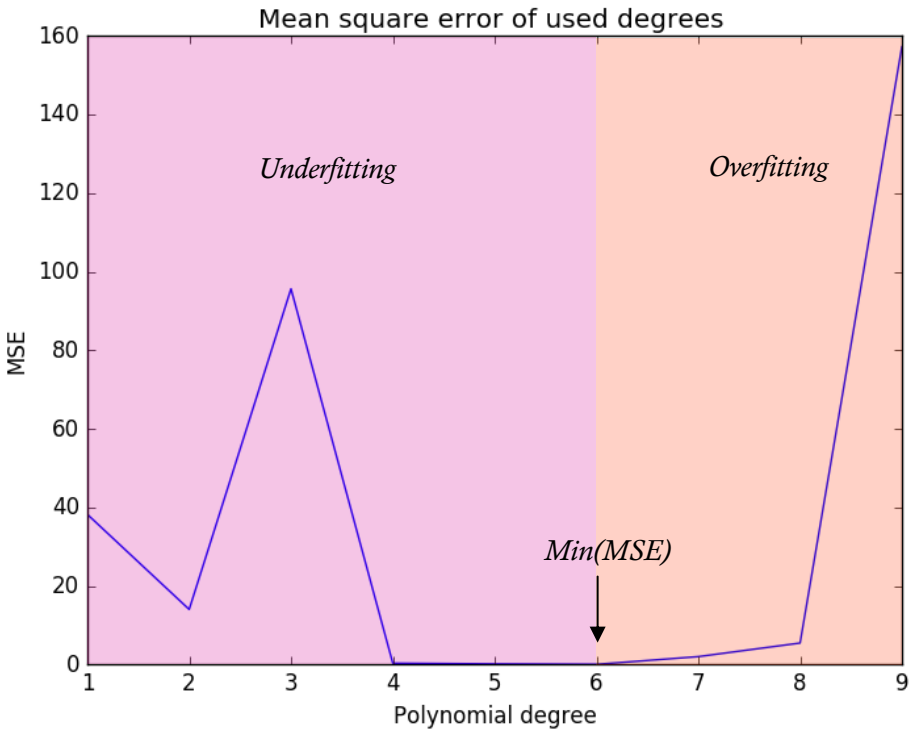






The model with the lowest error is the 6-degree polynomial (MSE=0.026).

In the plot below is reported the trend of the MSE with the degree of the considered polynomials. In order to have a better view of the trend, the MSE of the 10-degree polynomial is omitted (MSE_{10}=4860).



Why do high degree polynomial have a high error?

The fit operation is accomplished choosing the coefficients of the polynomial that minimize the sum of the squared errors (MinSE) of the data in the train set. While a n degree polynomial can fit data in the same way of a t<n degree polynomial (choosing the first t coefficients equal to the t polynomial and the final n-t coefficients equal to zero), the opposite is generally not true (unless the last n-t coefficient of the n degree polynomial are all zero). So, the MinSE of the n degree polynomial is less than or equal to the one of the t degree polynomial. This fact implies for instance that even data that we know linear have lower MinSE with polynomial of degree n>1, because higher degree polynomials are able to learn noise and statistical fluctuations of the data. We don't want to learn these aspects (overfitting), that changes between different sets of data (for instance the test set), we want to learn only the model, that do not change between different sets of data. To achieve this result, we use the test set and we choose among the model trained on the train set the one that minimize the MinSE of the test set (or equivalently the MSE of the test set), in our case the 6-degree polynomial. In this way it is possible to check when the model begins to learn noise and statistical fluctuations, the MinSE (or MSE) of the test set grows, because noise and statistical fluctuations of the train set are different from the ones of the test set.