

Figure 1 For the visualization a data set of 60 samples and a polynomial order of 12 is used. Red line is the output of the predicted function without any regularization, as we set lambda to 0. Other lines are prediction with regularization lambda. The blue line is the plot of the original data set

By trying out different regularization parameters we figured out that values below *-3* is a great choice. Taking smaller values for lambda didn’t make such a great change, which can be seen on [Figure 1].

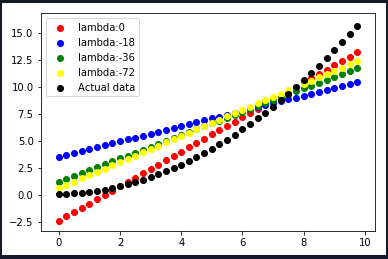


Figure 2 Fitting a 1-order polynomial function to a quadratic function

Choosing a good polynomial order is crucial for a good model. Low polynomial order can lead to under-fitting of the predicted model, when the actual data can only be generated with a higher polynomial order. [Figure 2] depicts the under-fitting where the predicted function is a 1-order polynomial and the actual data is a 2-order polynomial

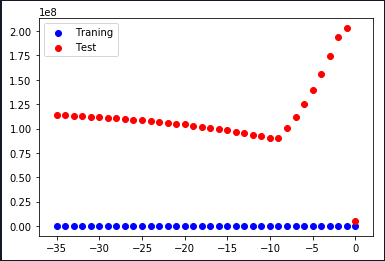


Figure 3 ERMS according to chosen lambda, also by using K-Folds. Red points are the test set and blue points are the training set. Sample data N = 100, with 3 folds and polynomial order of 9.

[Figure 3] depicts the impact of the regularization parameter to the model. The predicted model is fit the training set very well as the ERMS is nearly at 0, and for the test set the ERMS is higher, which tells us that the predicted model is over-fitted. We can see that the actual model was already an accurate model as the ERMS at lambda *0* is low. For the test set a jump can be seen on the [Figure 3] for lambda *-1*, but as lambda decreases the ERMS also decreases and at -9 there is a turning point as the ERMS increases again.

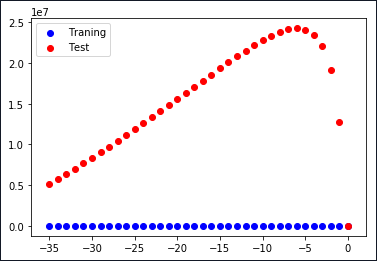


Figure 4 Impact of K-Fold to the ERMS. Sample data N = 100, with 3 folds and polynomial order of 9.

We can see by comparing [Figure 3] and [Figure 4], which have the same configuration except the K, that determining K by K-Fold does affect the predicted model, as the ERMS in [Figure 4] is greater for lambdas between -3 and –20 but then the ERMS is smaller for lower lambdas compared to a higher K in [Figure 3].