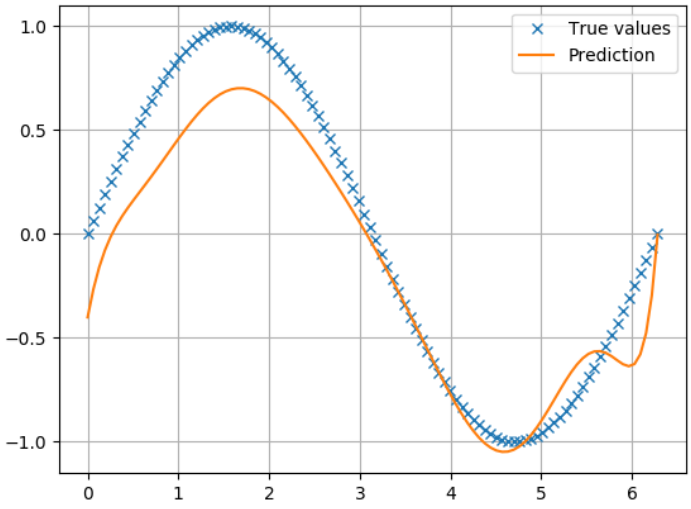
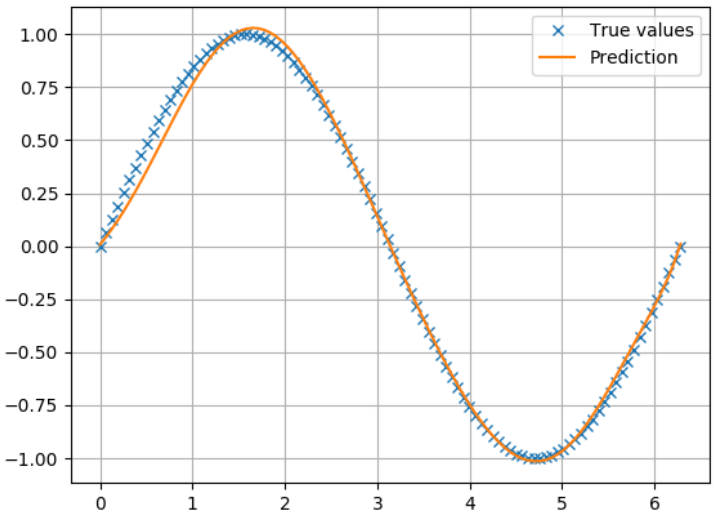
REPORT

1.) Use of regularization:

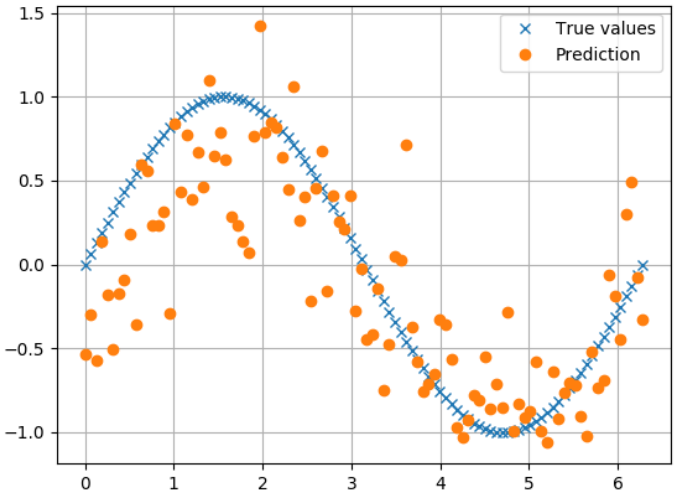
The following example clearly describes how regularization helps to avoid overfitting and improve the generalisation of the model. In this example we try to fit y=sin(x) using a polynomial of highest power 10. The model was trained on 10 equally spaced samples of x between 0 to 2π and tested on 100 samples. The left plot is normal polynomial fit whereas the right plot is ridge regression with lagrangian multiplier of 0.1.

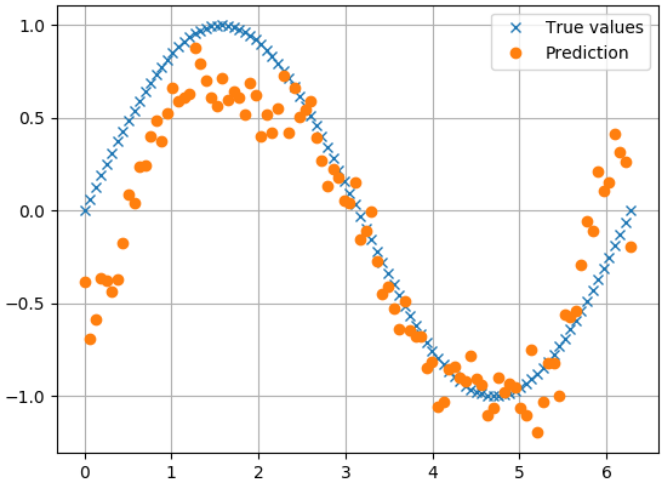




2.) Problem with high variance of predictor:

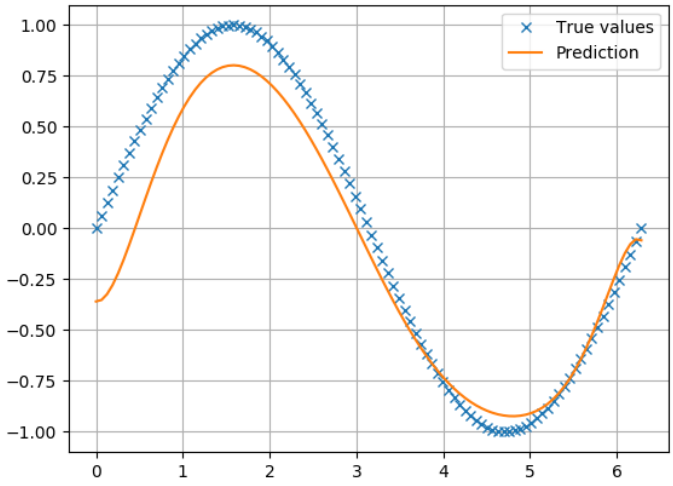
Higher the variance in predictor, the more it is vulnerable to noise. This can also be observed from the following example. The left plot has a variance(in predictor) of about 0.1 whereas the right plot has about 0.01. We can observe that higher variance in predictor leads to more randomness(both were tested with 100 samples).

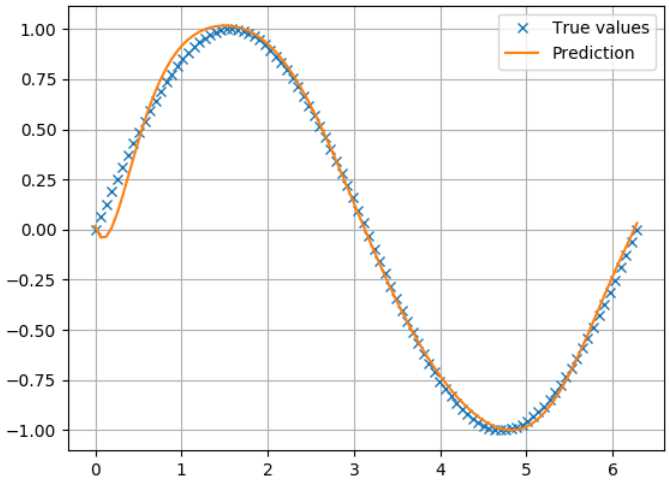




3.) More relevant training samples:

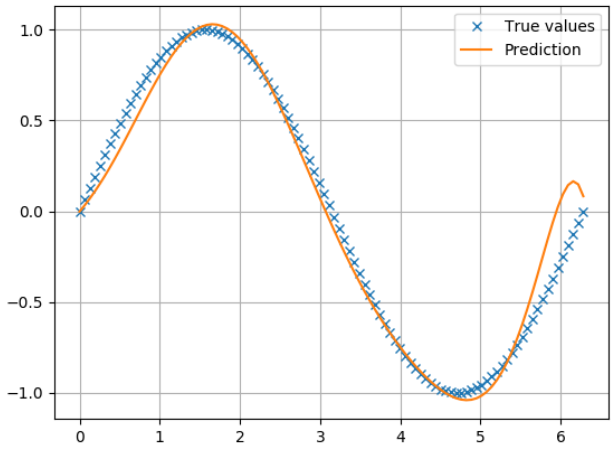
Higher the number of relevant training samples, better the performance of the model on test set. This can also be observed from the following example. The left plot was the result of training with 10 samples whereas the right plot was the result of training with 20 samples (both were tested on 100 samples).

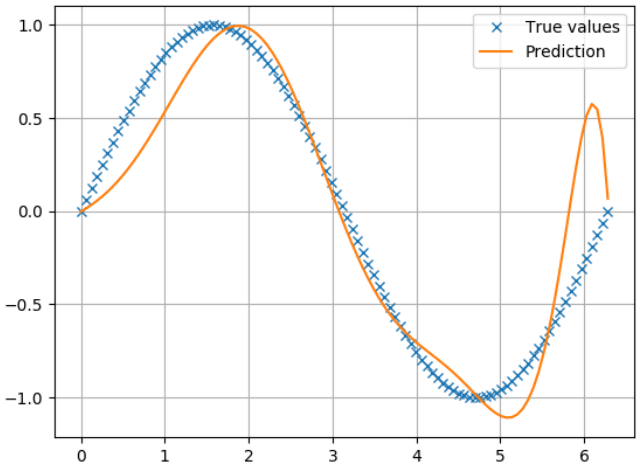




4.) Effect of higher values of Lagrangian multiplier:

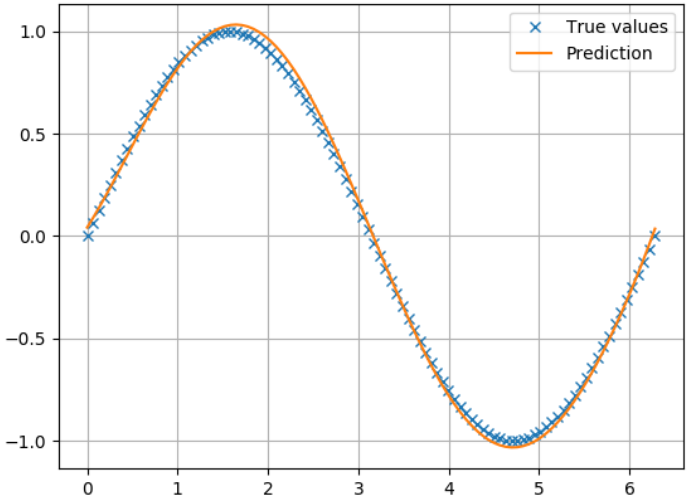
With increase in lagrangian multiplier, at least for this data, it can be observed that the performance reduces. This can be accounted for excessive shrinkage of weights. The left plot corresponds to a value of 0.1 whereas the right plot corresponds to a value of 1.

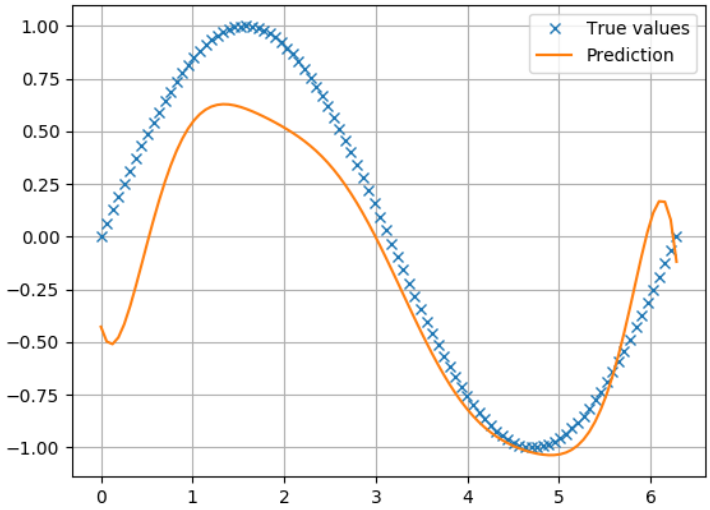




5.) Effect of the order of the polynomial used for fitting:

The best polynomial fit was found for degrees 6,7,8 (without regularization) . However if the degree is increased beyond the number of training samples, it will lead to worse fit. This can also be observed from the following plots. The left one is for degree 7 whereas the right one is for degree 10 (no regularization). Higher degree leads to more complicated model.





6.) Variance of predictor(sigma) vs Variance of weights(alpha):

This is an interesting comparision. If variance of weights is fixed, we end up with similar results as in **2. (**infact better due to added regularization term**).** However if there is a high variance, then no matter what the weights variance is, the model will not perform well even if alpha is very high(the performance improves very slowly).

Even if the ratio of sigma and alpha is the same, if sigma is high then the model does not perform well on test data. This can be observed from the following plots. The left has a higher sigma than right.