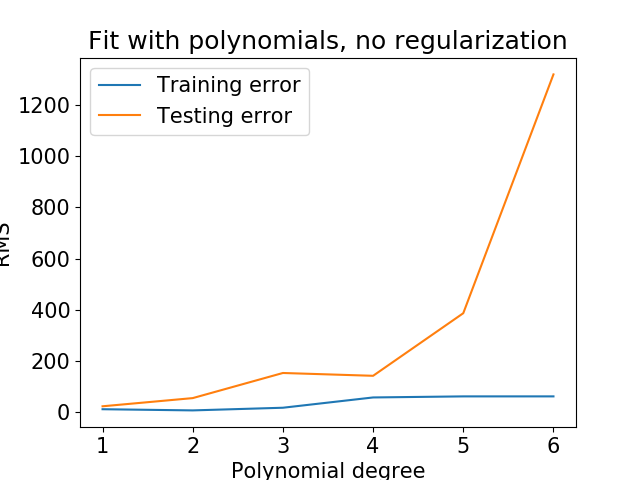
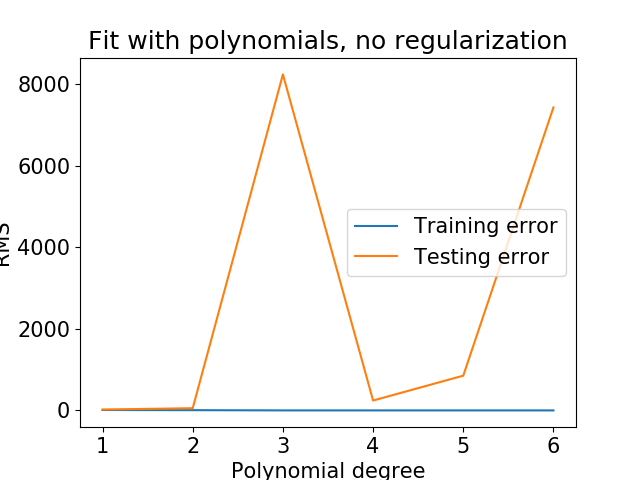
5.2

1)

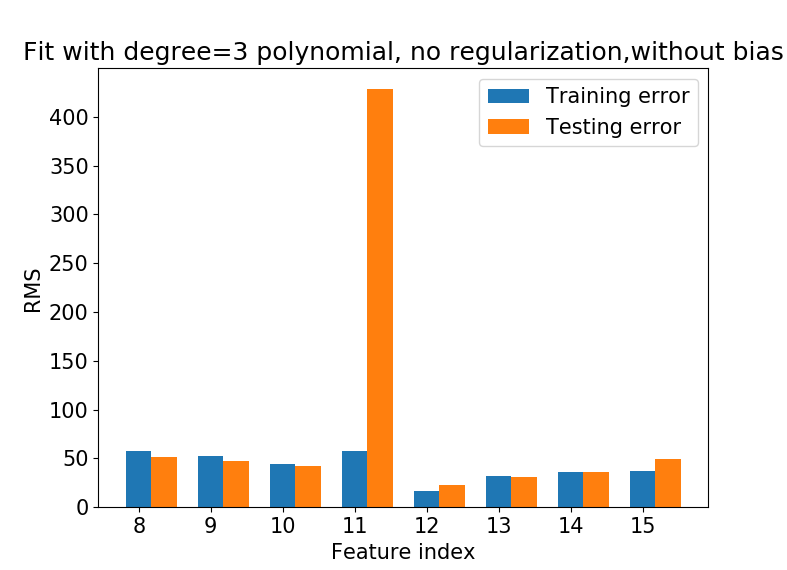


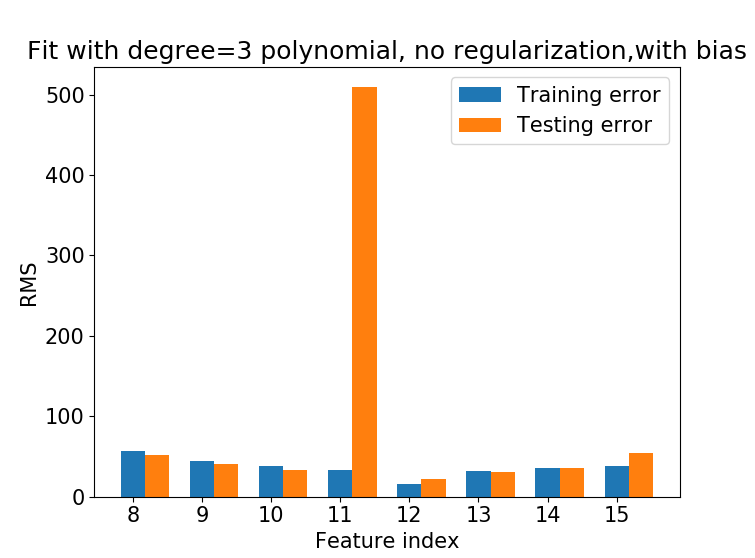
There are two things wrong here. The first one is usually as polynomial degree increases the training error will decrease since higher degree usually fits the data better. After we apply the normalization on input features this problem is solved. The reason is that we may have some features whose orders of magnitudes are larger than others, and they may dominate the objective function and make the estimator unable to learn from other features correctly as expected.

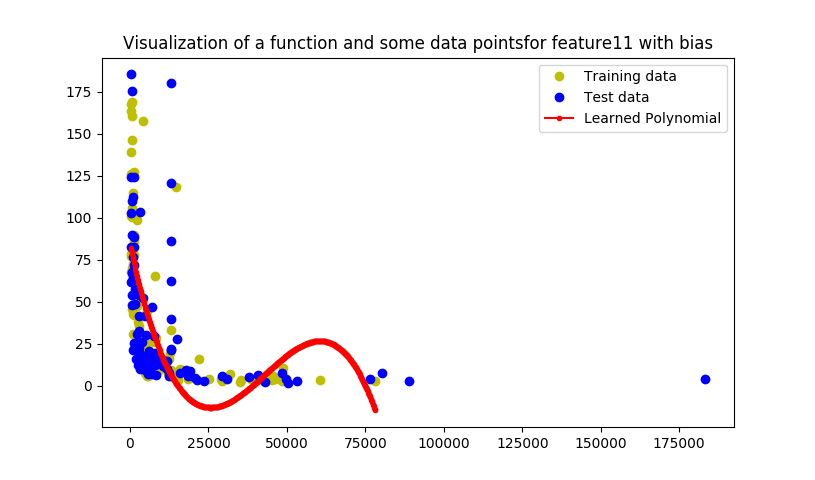
Another problem is as orders of degree increase, the model is being more over-fitted. This problem can be solved with regularized regression which we will do in 5.4.



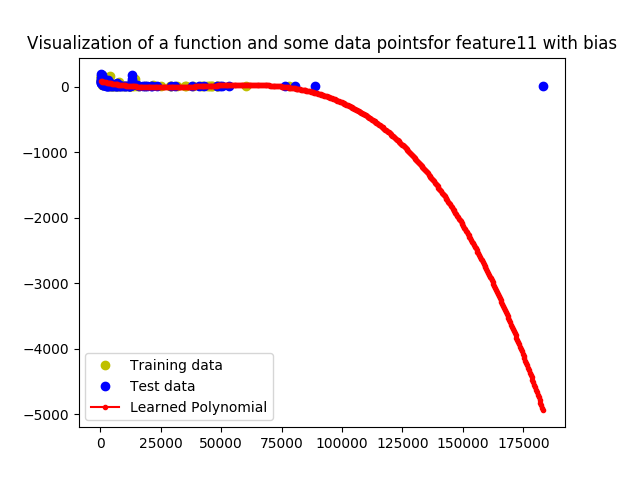
2)

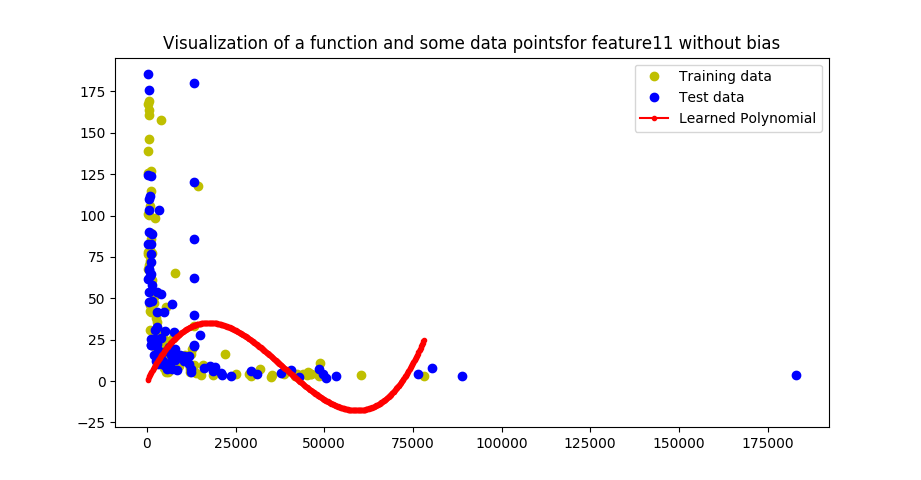


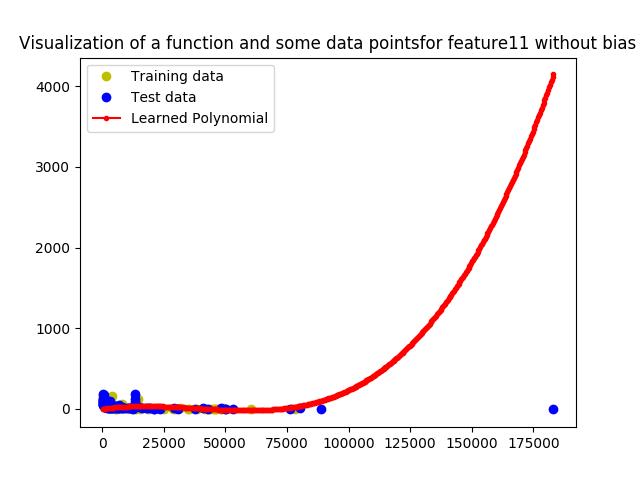


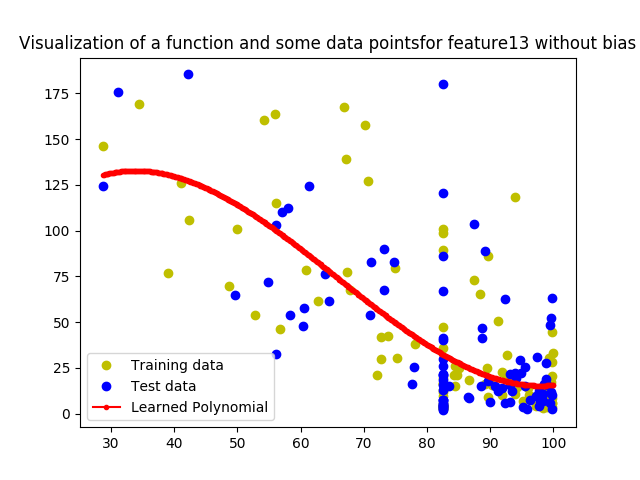
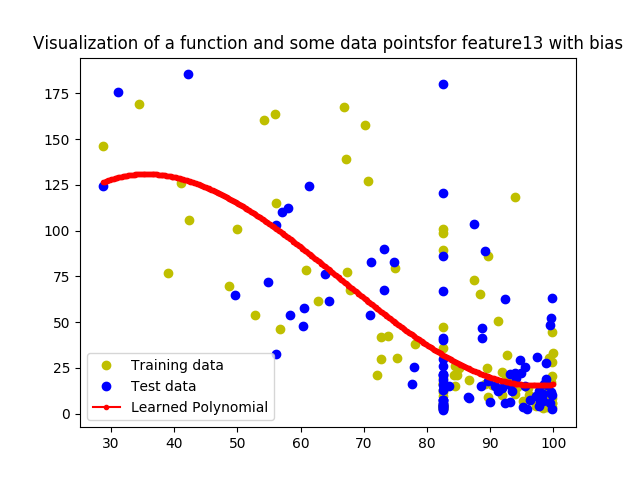
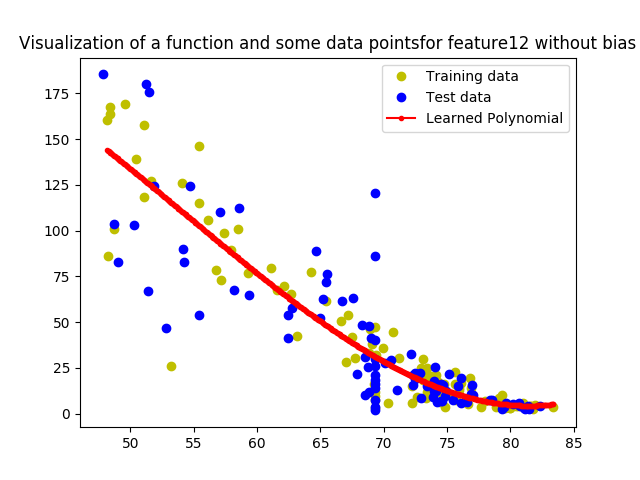
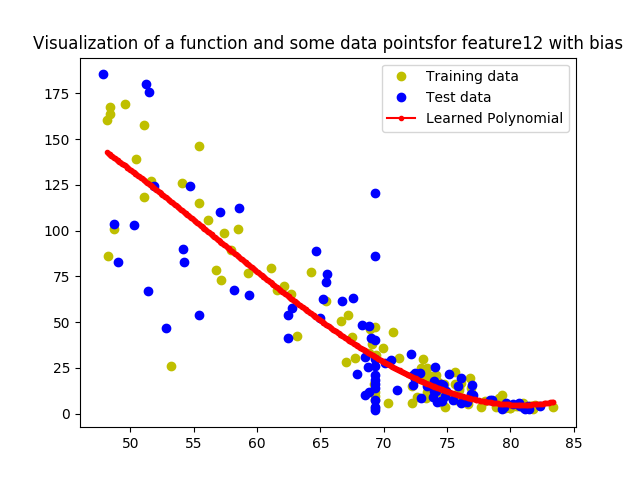


Since there is an outlier so we need to use a larger scale.

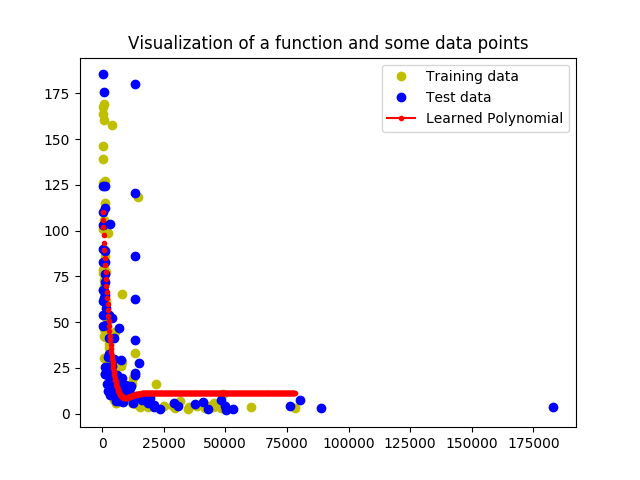








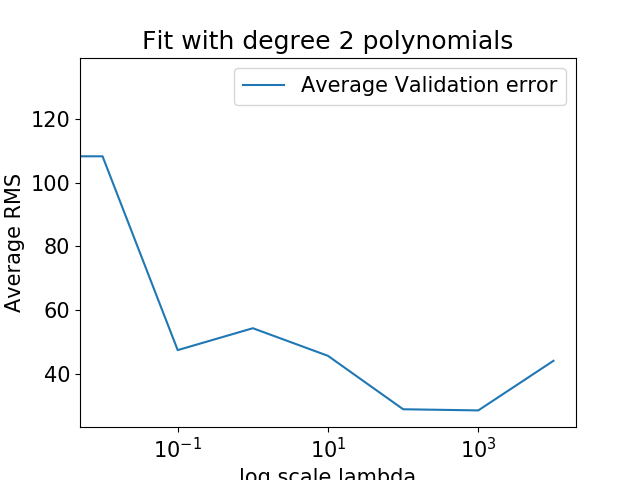
5.3



training error: [[28.45793776]]

testing error: [[33.8067249]]

5.4



the errors for lambda 0 to 1000 are :

lambda=0 134.08724800120225

lambda=0.01 108.33938137863713

lambda=0.1 47.42014676680511

lambda=1 54.30153018685256

lambda=10 45.61747891270658

lambda=100 28.827211650792474

lambda=1000 28.46922327978877

lambda=10000 44.06076704806573

Lambda=1000 yields the lowest average validation error so we will choose lambda=1000