COMP6245 : Lab 5 Report

Thanakorn Panyapiang(tp2n19@soton.ac.uk)

The performance of the original RBF implementation with 200 basis functions is show in Figure 1. The graph indicates that the model is significantly overffited with the training data and perform poorly on the test set.
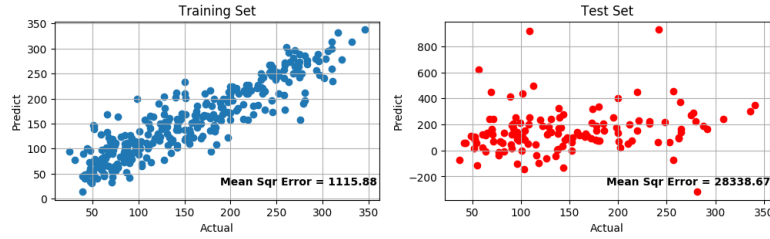

Figure 1

To improve the model, the following changes have been added:

1. Normalizing each feature so its value is between 0 to 1.

2. Computing the width of basis functions(σ) using average pairwise distance of 50 points which are randomly sampled from the training set.

3. Computing locations of the basis functions using K-mean clustering where K is set to be 30.

After above-mentioned improvements, although the error on the training set increases, the model accuracy on the test set is significantly better than the original version as can be seen from Figure 2. However, if the value of M is set to be very high(f.e. 100), the model is still overfiffed with the training data similar the original implementaion.
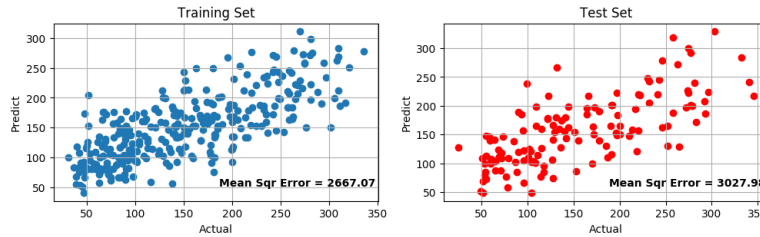

Figure 2

To compare RBF with Linear Regression(with L2 regularization), the data has been splitted into 10 parts to run ten-fold cross validation. The sample distributions of predicted values of both models are shown in the figure below. One

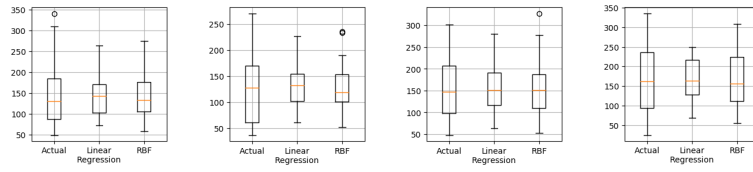thing which can be noticed is that RBF model generaly has a wider range of predicted values.

Figure 3