

1 Radial Basis Functions (RBF)

1.1 Normalization

Set mean to 0: For each column(feature), compute the current mean m , then let each elements e equals to $e - m$.

Set standard deviation to 1: For each column(feature), compute the current standard deviation sd , then let each elements e equals to $e - sd$.

1.2 Setting the width parameter of the basis function

The experiment shows that when the width parameter of the basis functions is set to be the average of several pairwise distance is better than the distance between two randomly chosen points. That may because the distribution of points may not liner in one dimension, so that average of several pairwise distance may better than randomly chosen points.

1.3 Setting the locations of the basis function

Kmeans function `sklearn.cluster.KMeans.fit(Train_data)` is used to cluster the data. The parameter of Kmenas object `kmeans.cluster_centers` could export the cluster centre.

1.4 Training and Testing

The training results and testing results when M equals to 20, 100, 200 are shown as Figure 1, Figure 2, Figure 3, Figure 4, Figure 5 and Figure 6.



Figure 1: Training Set M=20 Figure 2: Training Set M=100 Figure 3: Training Set M=200



Figure 4: Testing Set M=20

Figure 5: Testing Set M=100

Figure 6: Testing Set M=200

Table 1 shows the experiment results. From the result, the difference between error of training and testing are small when M equals to 20 which means that the model perform good. With the increasing of M, the numbers of basis function increased and model getting more and more over-fitting.

Table 1: Experiment Results

M	Training Set Error	Testing Set Error
20	2.8×10^3	3.0×10^3
100	1.7×10^3	3.5×10^3
200	1.1×10^3	3.2×10^3

2 Cross Validation

Ten-fold cross validation is applied using function `sklearn.model_selection.KFold(folderNumbers)`. Figure 7 shows the box plot of testing error of applying 3 models. The results shows that SKlearn model is better than my RBF model and my RBF model is better than Linear Regression model.

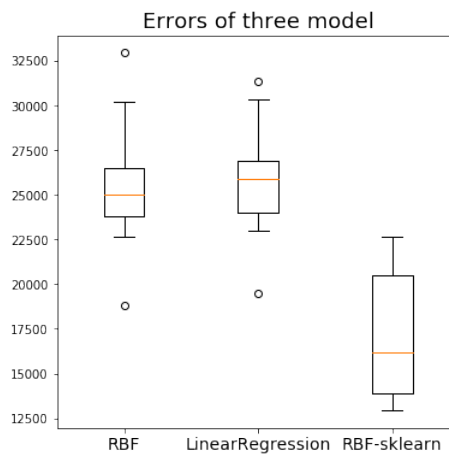


Figure 7: Testing errors of three model