**Machine learning lab5 report**

**Name: Xinyu Hu**

**Student ID: 29350379**

**Date: 17/11/2017**

**Exercise 1-6**

Aiming at use RBF to predict the Boston housing price, we should learn the formula of this model first and calculate the parameters inside step by step:

The function of RBF model:

Here we use Gaussian function:=exp(-), and we use K-means clustering(K=Ntr/10) to calculate the , and the σ can be set as the distance between two randomly chosen figure. First, I randomly choose 70% data as the training set, and rest is the test set.Therefore, the non-linear part is fixed, so we just need to calculate the by equation. After building up the model, we can get what it predict at each training data, the result of it can be plotted as follows:



The error between the target and the prediction on training data is: 50.3693.

**Exercise 7**

Then calculate what model predicts on testing data, the similar scatter plot can be seen as follows:



The error calculated on testing data is : 35.1721. The difference between two errors is 13.19.

Then we change the number of K, we no longer use K as Ntr/10. Instead, we can set K as Ntr/n(n from 2 to 12), and compare the different values of error differences. The result is shown in the following table:

|  |  |  |  |
| --- | --- | --- | --- |
| K= Ntr/n(n=1) | -17.89 | K= Ntr/n(n=7) | 27.1 |
| K= Ntr/n(n=2) | -19.82 | K= Ntr/n(n=8) | 30.2 |
| K= Ntr/n(n=3) | 3.90 | K= Ntr/n(n=9) | 4.68 |
| K= Ntr/n(n=4) | -13.89 | K= Ntr/n(n=10) | 13.19 |
| K= Ntr/n(n=5) | 16.63 | K= Ntr/n(n=11) | -3.44 |
| K= Ntr/n(n=6) | 14.41 | K= Ntr/n(n=12) | 26.64 |



We can clearly see that as the K get increased, the error difference would be smaller than before.

**Exercise 8**

Compare the result between the linear and non-linear models. Here we use mean squared prediction error to measure these two models, the outcome is shown in the following boxplot:



According to the boxplot, we can conclude that the RBF’s predicting performance is better than the traditional linear model’s in data we use(mean is smaller and the range is also relatively small). So non-linear model indeed can improve the prediction performance.

**Exercise 9**

To implement the RBF model again and prove my above observation, I download another dataset from the UCI Machine Learning repository. The dataset is about Combined Cycle Power Plant Data Set, it has 4 dimensions, aiming to predict the net hourly electrical energy output (PE) of the plant. I implement the linear and RBF model to this dataset. The result below also indicates that the RBF model would perform better than the linear one.





