Machine Learning Lab 3

Junming Zhang   29299527   jz1g17@ecs.soton.ac.uk
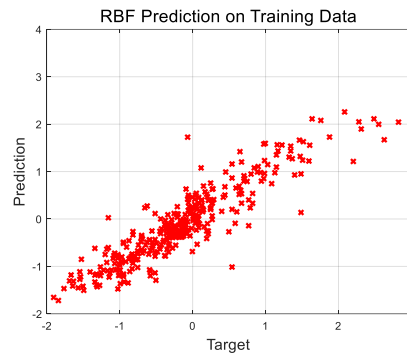
Problem 1 -6

Figure 1. The prediction on training data

Function of RBF model: Gaussian function:

$$g(x) = \sum_{k=1}^{K} \lambda_k \emptyset(\|x - c_k\|) \qquad \emptyset(\alpha) = \exp(-\alpha/\sigma^2)$$

If apply Gaussian function to RBF model with c_k calculated by K--means clustering(K=Ntr/10). The value of σ is set as the distance between two random scatters. In this case, 70% data are used for training set and rest of it is for test set. Finally, the λ_k can be calculated.

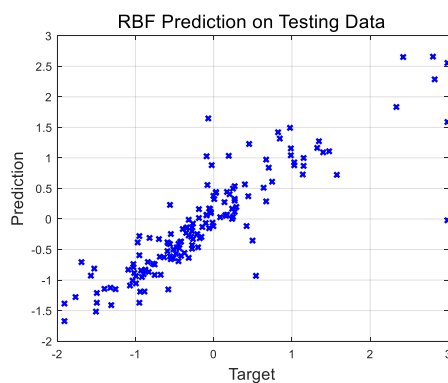The error between the target and the prediction on training data is: 50.37.

Problem 7

Figure 2. Second prediction diagram

Figure 3. The line chart of error with K

The analysis object is the predicted model of testing data. The distribution of it can be shown in Figure 2.

The error between the target and the prediction on this data is 35.17 and diffrerence between them is 13.19. For observing the different of error changed with K, a K value line chart is plotted as Figure 3.
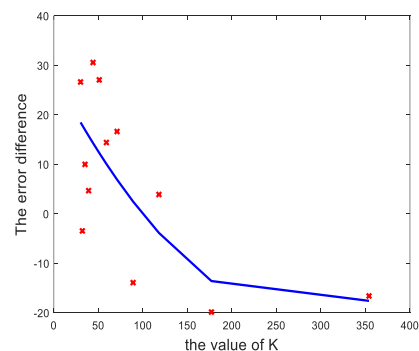
For K= Ntr/n(n=1~12), the corresponding value of Δ_{error} are show as Table 1.

1	2	3	4	5	6	7	8	9	10	11	12
17.89	19.82	3.90	13.89	16.63	14.41	27.10	30.20	4.68	13.19	3.44	26.64

Table 1. The K value and corresponding error

It could find the error difference is decrease with the increasing K.
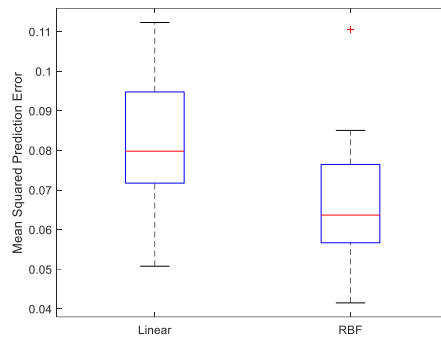
Problem 8

Figure 4. Boxplot of linear and non-linear models

The mean squared prediction error is used to compare the result between linear and non-linear models. A boxplot which show in Figure 4 is apply here.

It could find the RBF prediction is better than tradition linear model because the mean and floating range are smaller. Which means the non-liner model improve the performance of prediction.