

Machine Learning Lab 5

Zhikun Zhu; ID: 29356822

November 21, 2017

I Radial Basis Functions (RBF)

In this section, Radial Basis Functions as well as K-means algorithm are used to training the dataset which provided by [1].

Generally, the RBF function is given in Eq.1, which is a non-linear regression.

$$g(\mathbf{x}) = \sum_{k=1}^K \lambda_k \phi(||\mathbf{x} - \mathbf{c}_k||) \quad (1)$$

And for this specific case, the weights λ_k are fixed, Besides, Gaussian RBF is used, which is given by:

$$\phi(\alpha) = e^{-\frac{\alpha^2}{\sigma^2}} \quad (2)$$

The training results as well as prediction results are shown in Fig.1 and Fig.2, respectively. The error for training set is 0.1780 and its for test set is 0.2128.

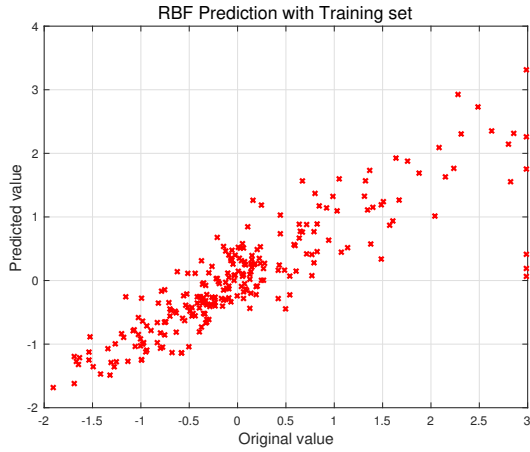


Figure 1: RBF prediction results with training dataset

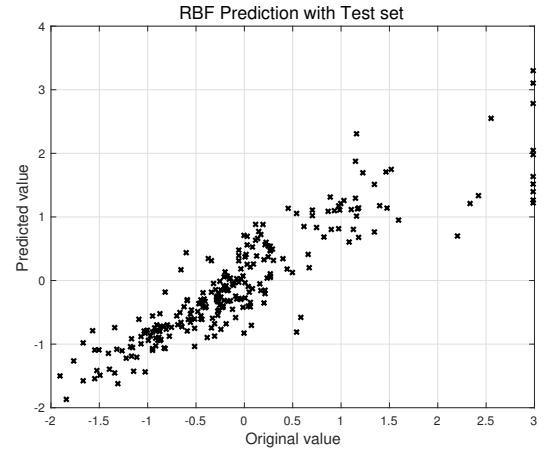


Figure 2: RBF prediction results with test dataset

II Effects of K-means

Then, we change K, from 1 to the size of training dataset, with a step of 10. And, finally, the error for both training set and test set are plotted in Fig.3 and Fig.4, respectively.

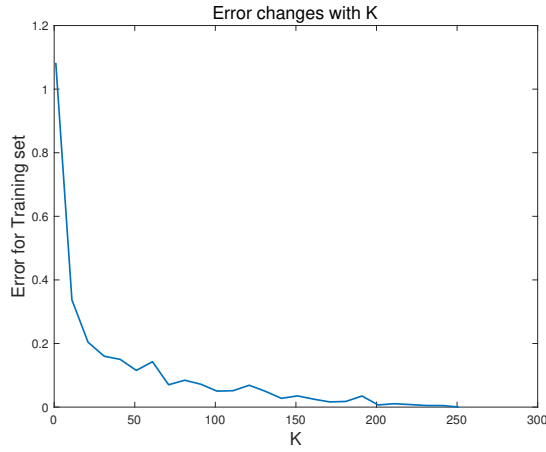


Figure 3: Error of Training set changes with number of K

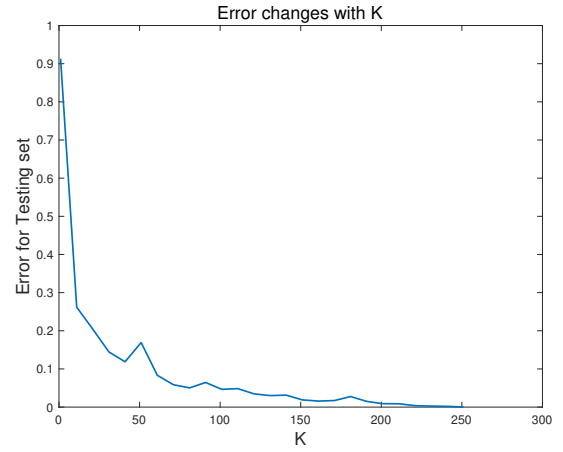


Figure 4: Error changes of Test set with number of K

III Comparison RBF with linear regression

Employ the algorithm which specified in Lab 4 to the training set as well as test set which is same with previous sections. The mean squared error on test data was obtained from 20 random circulation.

The boxplot for both Radial Basis Functions (RBF) method and Linear regression method are plotted. We then can compare two algorithms by the error of the prediction of house prices, which can be seen in Fig.5.

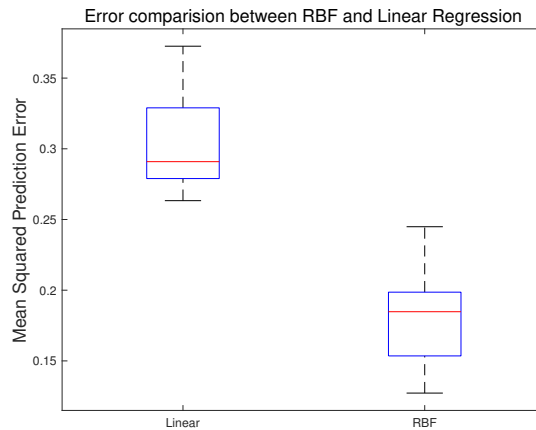


Figure 5: Error comparison between RBF and Linear Regression

References

- [1] David Harrison Jr and Daniel L Rubinfeld. Hedonic housing prices and the demand for clean air. *Journal of Environmental Economics Management*, 5(1):81–102, 1978.