COMP6245(20/21)-Foundation of machine learning(MSc)-Lab4
Pasinpat Vitoochuleechoti (pv1u20@soton.ac.uk)

Radial Basis Functions (RBF)

Implementation

At the beginning, I'm try to ran skeleton implementation of a Gaussian RBF model and we got the result as in Fig. 1.

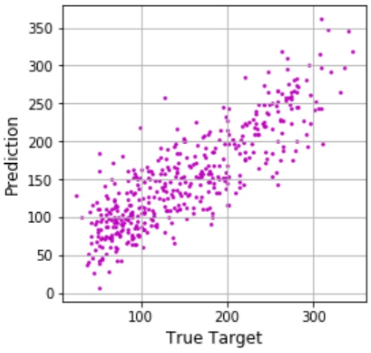

Figure 1: Skeleton Implementation of RBF Result

Improve the given implementation

Next step is improve the given implementation by follow the report. First, I normalize each feature of the input data to have a mean of 0 and standard deviation of 1 that mean we need to standardization input data to a z-score. We can find z-score by use equation:

$z_i = \frac{x_i - \bar{x}}{S}$ that

x_i is a previous sample in each feature of the input data

z_i is a a new sample in each feature of the input data or z-score

$\bar{x}$ is mean in each feature of the input data

S is standard deviation in each feature of the input data

Second, I change the parameter σ to be the average of several pairwise distances by used function in scipy library[1] and then find the mean by used numpy library[2].

Third, I used K-means clustering function from sklearn library[3] to set the basis function locations to the cluster centres. Then, When we implement the code we will got the new result as illustrate in Fig. 2.

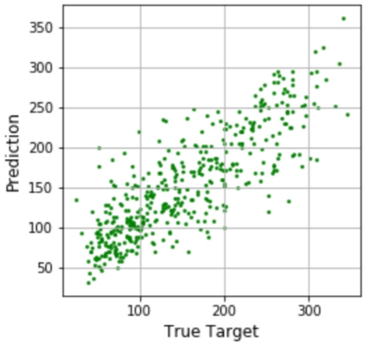

Figure 2: Result after make the following improvements

The last step of improve the given implementation we split the dataset into train and test dataset, I split it into 50 percent of train dataset and 50 percent of test dataset. The result of train and test dataset show in Fig. 3 and Fig. 4, respectively.

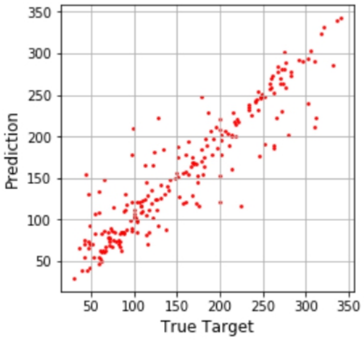

Figure 3: Training result after make the following improvements

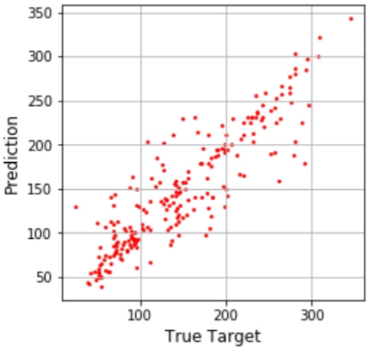

Figure 4: Testing result after make the following improvements

Implement ten-fold cross validation

When we finish all the improve the given implementation process, we will implement ten-fold cross validation and the result as present in the Fig. 5. The result look a bit strange because it is the combine of ten test result in ten-fold cross validation.

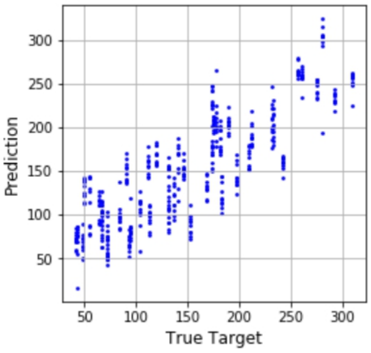

Figure 5: The combine result of the split testing data in ten-fold cross validation

Compare gaussian RBF model implementations with other model and show display the distributions of test set results by boxplots

The Fig. 6 is the result of linear regression models that not similar to any result of gaussian RBF model and the boxplot that show distribution of linear regression models, gaussian RBF model and true target is illustrate in Fig. 7.

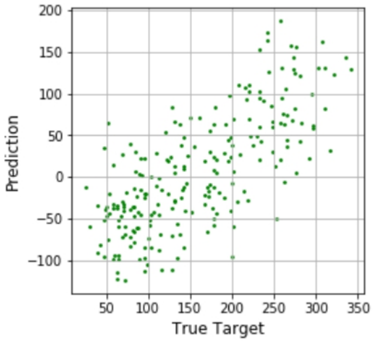

Figure 6: The result of linear regression models

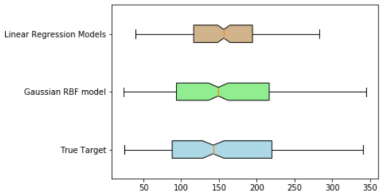

Figure 7: The boxplot of linear regression models, gaussian RBF model and true target

The last implementation is the sklearn inbuilt RBF model by use ten-fold cross validation and show the combine result of the split test data in ten-fold cross validation same as gaussian RBF model implementation. When we look at the result of sklearn inbuilt RBF model in Fig. 8, we will see the result less than Fig. 5 because the sklearn inbuilt RBF model might have performance better than gaussian RBF model. That make the result of sklearn inbuilt RBF model in each testing result of ten-fold cross validation have the same result and when we plot its we will see its overlap each other results. The boxplot that show distribution of sklearn inbuilt RBF model, gaussian RBF model and true target is illustrate in Fig. 9.

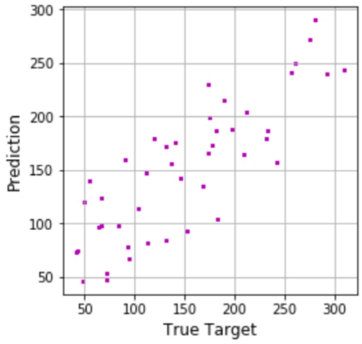

Figure 8: The result of sklearn inbuilt RBF model by use ten-fold cross validation

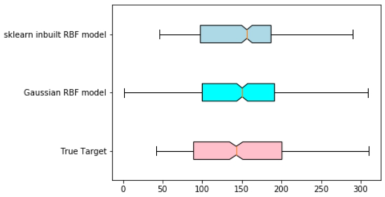

Figure 9: The boxplot of sklearn inbuilt RBF model, gaussian RBF model and true target (All use ten-fold cross validation)

Reference

[1]: Pairwise distances function in scipy library `https://docs.scipy.org/doc/scipy/reference/generated/scipy.spatial.distance.pdist.html`
[2]: Mean function in numpy library `https://numpy.org/doc/stable/reference/generated/numpy.mean.html`
[3]: K-means clustering function in sklearn library `https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html`