Report For Radical Basis Function

RUN YING JIANG (rj1u20@soton.ac.uk)

I. Introduction

In this lab, we implement a Radial Basisi Functions model on **Diabetes dataset** [1] and compared its performance with linear regression model and the RBF model imported from **sklearn**. Besides, ten-fold cross validation is used to quote uncertainty in comparing the performances.

II. Comparison of Linear regression and RBF model

We implement linear regression and RBF model seperately and scatter target data and predict data on the plot to have a more intuitive view of comparison results. **Figure 1 :RBF result** and **Figure 2: Linear model result** .
We find that compared with linear regression, RBF model has a better performance on training set but not work as well as linear regression in testing set.

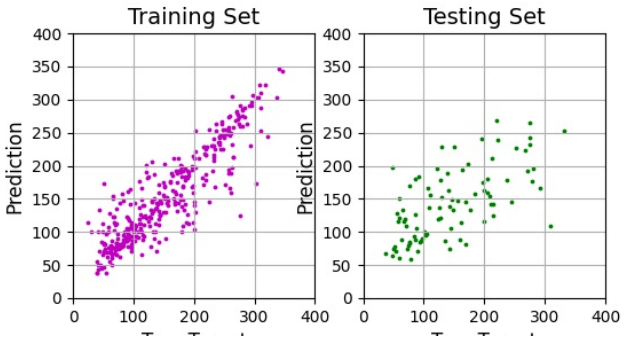

Fig. 1. RBF result: The right columns shows scatter of target data and predict data of rbf model on training set while the left shows the scatter on test set. The horizontal axis is the target data and the vertical axis is the predicted data.

Fig. 2. linear regression result: The right columns shows scatter of target data and predict data of linear regression model on training set while the left shows the scatter on test set.The horizontal axis is the target data and the vertical axis is the predicted data.

A. Question on basis function σ

I suppose set σ to be the distance between two randomly chosen points would cause error. Because it is a random situation which means σ could be very close to zero. And we will get a bad performance of the model in this case.

III. Comparison of original data and normalized data on RBF model

We normalize each feature of input data with a mean of 0 and standard deviation of 1 and implement RBF model on the standard-normalized dataset.**Figure 3 :Normalization result**. By comparing the result of original data **Figure 1 RBF result**. we can know that the standard normalization leads to a worse result in RBF model, which is more converge on the training set but overfitting on testing set.

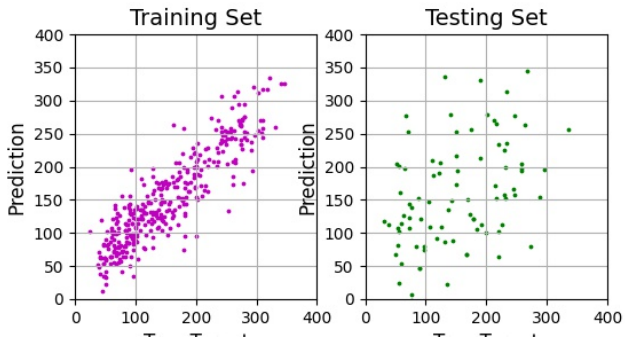

Fig. 3. Normalization result: The dataset has been standard-normalized.The right columns shows scatter of target data and predict data of RBF model on training set while the left shows the scatter on test set.The horizontal axis is the target data and the vertical axis is the predicted data.

IV. Cross validation results on RBF and Linear Regression Model

We implement ten-fold cross validation which is splitting the data into ten part, training on nine tenths of the data ,testing on the held out tenth set and repeating the process ten times. For five-fold cross validation, we repeat the process five times. By implementing ten-fold cross validation, we iterate the whole dataset and get a more stable result.The cross validation result is in **Figure 4:Cross validation result on RBF and Linear Regression Model**
We can see from the figure that the RBF model has a sligtly better performance than linear regression model.

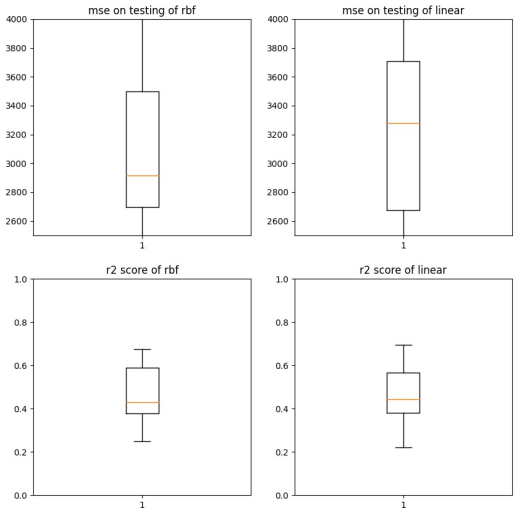

Fig. 4. Cross validation result on RBF model and Linear Regression Model: The right column shows the boxplot of RBF model while the left shows the boxplot of Linear Regression model. The plot in the upper-left corner shows the mean squared errors of the cross validation result of RBF model, which has a mean value of *3035.9300* while in the upper corner, the mean value of the Linear Regression model is *3224.0406*. Besides, the bottom left corner shows the R2 score. By evaluating with the criteria of R2 score, RBF model get a mean value of *0.4554* and Linear regression model is relatively weaker with a mean value of *0.4072*

V. Cross validation results on RBF and sklearn Model

We import GaussianProcessRegressor model from sklearn, and set the kernel of the Regressor as RBF. And by running the model with different *learning scale in kernel*, we set value of parameter *learning scale* to 200. **Figure: 5 The performance of sklearn in different learning scale**

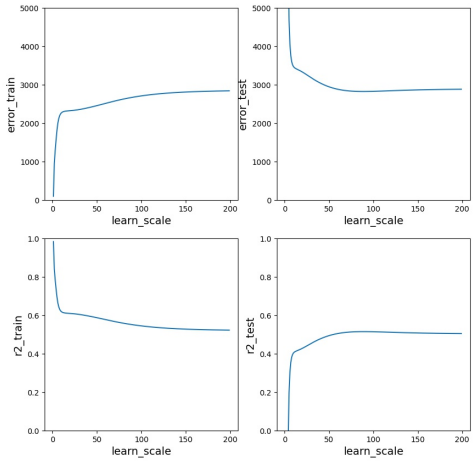

Fig. 5. The performance of in different learning scale: The horizontal axis is the learning scale value ranges from 1 to 200. The vertical value is the performance of model in mean squared error and r2 score

We compare the results of our RBF model with the RBF model imported from sklearn with a learning scale of 200 and find the sklearn model performs slightly better than our RBF model. **Figure: 6 Cross validation results on RBF model and sklern Model**

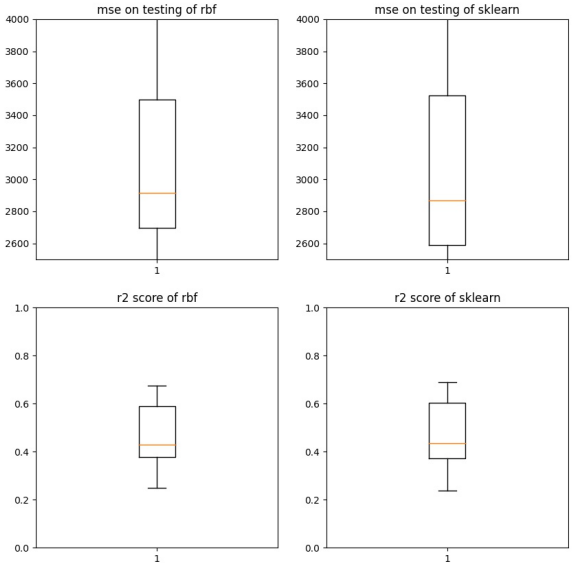

Fig. 6. Cross validation results on RBF model and sklern Model: The right column shows the boxplot of RBF model while the left shows the boxplot of sklearn model. The plot in the upper-left corner shows the mean squared errors of the cross validation result of RBF model, which has a mean value of *3035.9300* while in the upper corner, the mean value of the sklearn model is *2988.7916*. Besides, the bottom left corner shows the R2 score. By evaluating with the criteria of R2 score, RBF model get a mean value of *0.4554* and sklearn model is relatively better with a mean value of *0.4641*

References

[1] Dheeru Dua and Casey Graff. UCI machine learning repository, 2017.