**EE7207 Assignment1**

**ZHANG ZHENGHANG**

**E-mail: ZZHANG062@e.ntu.edu.sg**

**1.The process of building a RBF model:**

1. The suitable number of hidden layer neurons

center\_num=10

1. Neuron center determination

I use the K-means clustering to select prototypes of the training samples.

1. the radial basis function

Gaussian function:.d means the distance of sample and the neuron centre.

 determines the size of receptive field, which is equal to.

1. weight estimation



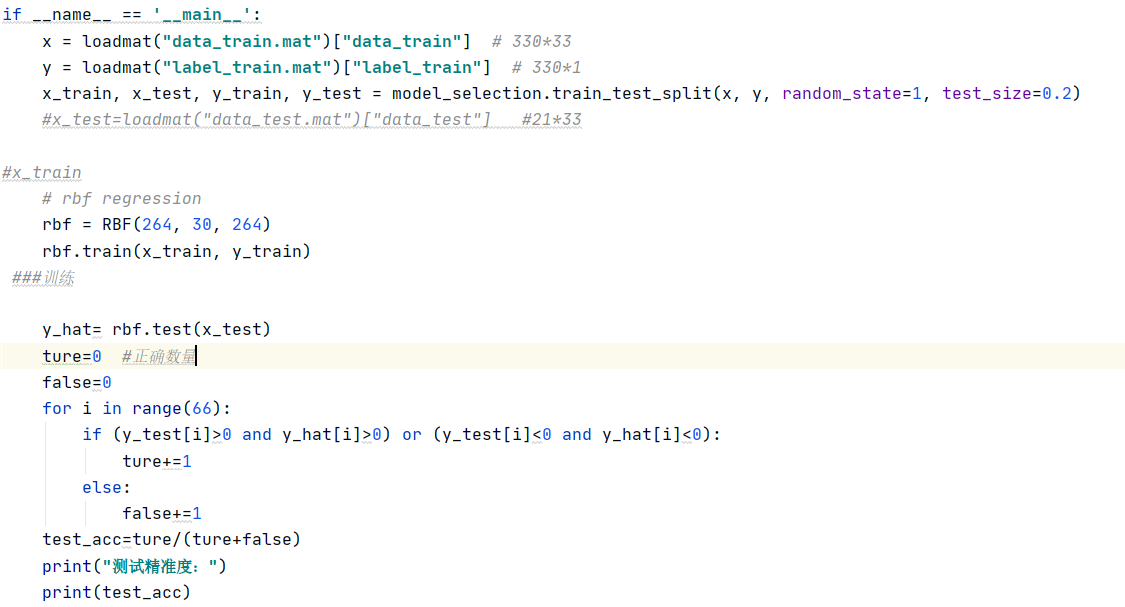
**Data preprocessing:**

Totally, there are 330 samples in the TRAIN set. I divided the set into two parts. The 80% samples used as train samples and the other 20% samples used to judge the performance of the model.

**Python code:**







**2.The process of building a SVM model:**

C means penalty parameter. The larger C is, it is equivalent to penalizing the slack variable. It is hoped that the slack variable will be close to 0, that is, the penalty for misclassification will increase, which tends to be the case of fully splitting the training set, so that the accuracy of the training set is very high, but generalization Ability is weak. The C value is small, the penalty for misclassification is reduced, fault tolerance is allowed, they are regarded as noise points, and the generalization ability is strong.

Therefore, I assume it equal to 0.8.

Gamma is a parameter of the Gaussian Kernel Function used in SVM. In python, it is usually set as ‘auto’. Through a series test on the validation set, I found it is better to set Gamma equal to 0.4.

**Data preprocessing:**

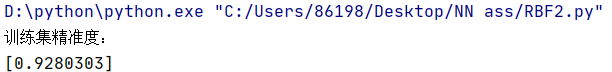
The same as RBF part.

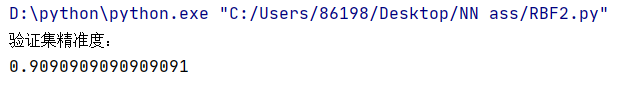
**Python code:**



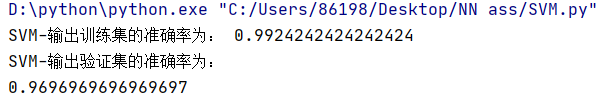
**3.Result and analysis:**

**A.RBF**





**B.SVM**



|  |  |  |
| --- | --- | --- |
| accuracy | **RBF** | **SVM** |
| **Train data** | **92.80%** | **99.24%** |
| **Validation data** | **90.90%** | **96.96%** |

**SVM has a better performance no matter on train data or validation data, which means the SVM has high prediction accuracy and good generalization ability in the mean time.**

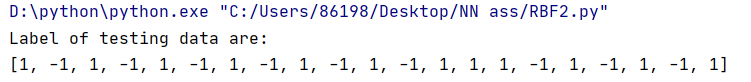
**Because the underlying theory of SVM avoids the need for heuristics often used in the conventional RBF.**

**And in Gaussian kernel SVM, the number of the neurons and their centre vectors are determined automatically.**

**4.Predict:**

The are 21 samples without labels given in another file. By the above two models, the prediction is shown below.

**A.RBF**



**B.SVM**