309551101郭育麟

Pattern Recognition

***Part 1***

1.

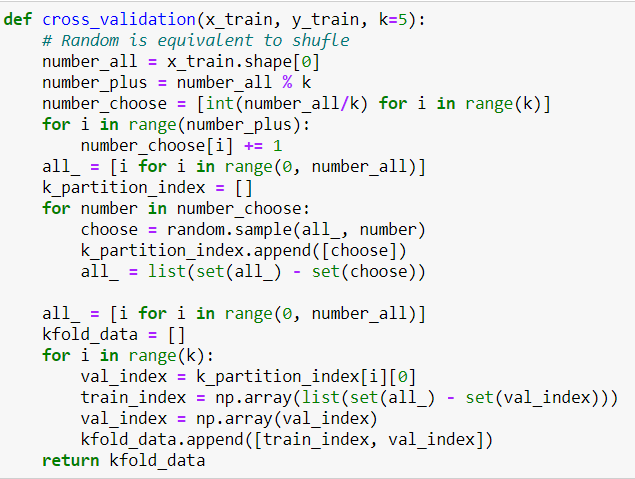
(10%) K-fold data partition: Implement the K-fold cross-validation function. Your function should take K as an argument and return a list of lists (len(list) should equal to K), which contains K elements. Each element is a list contains two parts, the first part contains the index of all training folds (index\_x\_train, index\_y\_train), e.g. Fold 2 to Fold 5 in split 1. The second part contains the index of validation fold, e.g. Fold 1 in split 1 (index\_x\_val, index\_y\_val)

Note: You need to handle if the sample size is not divisible by K. Using the strategy from sklearn. The first n\_samples % n\_splits folds have size n\_samples // n\_splits + 1, other folds have size n\_samples // n\_splits, where n\_samples is the number of samples, n\_splits is K, % stands for modulus, // stands for integer division. See this post for more details

Note: Each of the samples should be used exactly once as the validation data

Note: Please shuffle your data before partition

K is the parameter and represent that has K choice for training set and validation set . Due to need to handle the sample size is not divisible by K , I first create an array to represent the number of each partition should choose , then use “random” to choose index k times . Then have k partition , so I could choose one partition for validation set and another k-1 partition for training set , and each partition would be once the validation set .

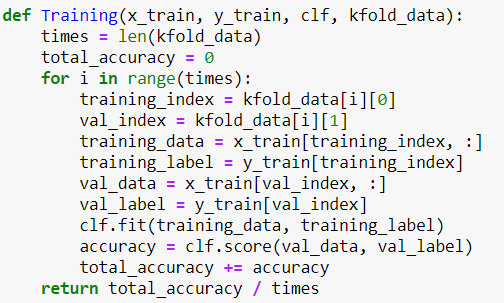


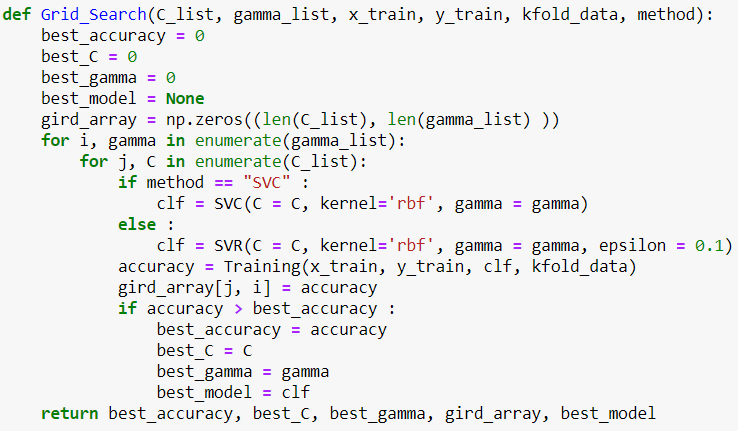
2.

(30%) Grid Search & Cross-validation: using sklearn.svm.SVC to train a classifier on the provided train set and conduct the grid search of “C” and “gamma”, “kernel’=’rbf’ to find the best hyperparameters by cross-validation. Print the best hyperparameters you found.

Note: We suggest use K=5

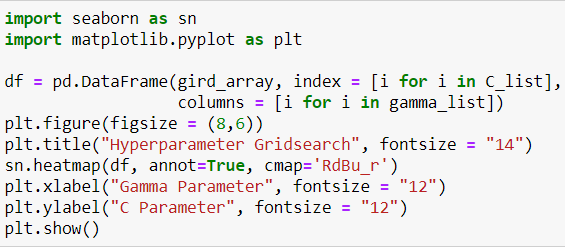
In this step , let “C” and “Gamma” has multiple choice , which C = [0.01, 0.1, 1.0, 10.0, 100.0, 1000.0, 10000.0] , Gamma = [0.0001, 0.001, 0.01, 0.1, 1, 10, 100, 1000] and use grid search to find the best model parameters .

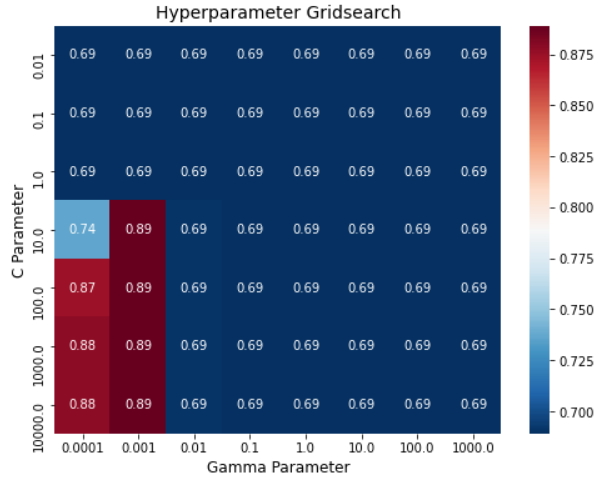




3.

(10%) Plot the grid search results of your SVM. The x, y represent the hyperparameters of “gamma” and “C”, respectively. And the color represents the average score of validation folds.

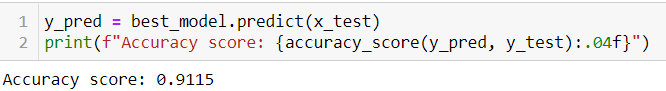




4.

(15%) Train your SVM model by the best hyperparameters you found from question 2 on the whole training set and evaluate the performance on the test set.

Note: Your accuracy scores should be higher than 0.85



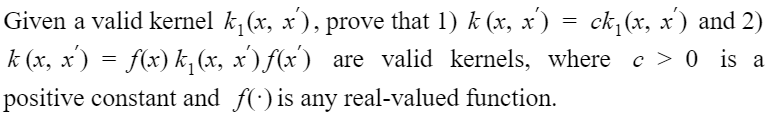
5.

(15%) Consider the dataset used in HW1 for regression. Please redo the above questions 2 ~ 4 with the dataset replaced by that used in HW1, while the task is changed from classification to regression. You should use the SVM regression model RBF kernel with grid search for hyperparameters and K-fold cross-validation (you can use any K for cross-validation). Then compare the linear regression model you have implemented in HW1 with SVM by showing the Mean Square Errors of both models on the test set.



**Part2**

1.



is a valid kernel

(1)

(2)