NAME: ASHUTOSH UPADHYE

ID: 1001581542

SUBJECT: MACHINE LEARNING CSE 6363

DATE: 09/18/2018

Iris Data set was used for the linear regression algorithm. The data contains 150 records with 4 features of the flowers and the fifth one to be the species class.

Normally the N for the K-fold classification is around 1. For this the training data and testing data both are divided into 50-50 or exact half and the training takes place on training data.

First we calculate the beta matrix for all the predictor variables. Its dimensions would be 1\*4

Beta = inverse (X\_values.T.dot(X\_values)).dot(X\_values.T).dot(Y\_values.T) gives the beta matrix.

The dot product of beta matrix with vector values gives the prediction matrix.

Predicted\_matrix = B.dot(X\_values)

For calculating the stability and accuracy of the model there are two measures:

1)RMSE-Root Mean Square Error

2)R2 Score

The R2 score of the model was around 92% which dictates the accuracy of the model

The RMSE of the model is quite less which describes the variance of the variables in the model.

The values of RMSE and R2 score are printed on the console of the project.

For a low number of records i.e.150 the number of k for the k-fold validations should be less as possible. The R2 score and RMSE are proper factors for the validity of the model.

K-fold cross validation is also a factor for describing the validity of the model and how accurate one’s prediction is.

Just run the Linear\_Regression.py and the output will be printed on the console.

The iris data set is loaded in a .csv excel file.

The output on the console shows the predicted output with labels.