Machine Learning

Assignment 1

Submitted by Dhruvin Doshi

UTA ID: 1001419924

Classification via Training a model to do Linear Regression

1. Introduction:

The given data was used to train a model with the help of one of the basic concepts i.e. Linear Regression with N-fold cross validation as means to improve the accuracy of the data predicted by the model. After the training and testing of the data, the following evaluations were made.

1. Problem:

Given a data set that consists of description of types of flowers as its class, sepal length and width, petal length and width, we wish to make future predictions for similar data which enables the model to classify the flower as one of the three flowers. We are also expected to perform N-fold Cross validation which helps in expanding the training set. One may repeat this procedure to improve accuracy.

1. Data:

The data is already given in a comma separated value file (.csv) which is directly imported using pandas library in python. The data consists of a 150 x 5 table.

1. Method:

The method is divided into Training, Testing and Evaluating the data given and observed and drawing conclusions from the same.

* Training

We compute Beta, which is a multi-variate least square estimator. Using the formula described and mentioned in the class. Using the same where A is the training set and Y is the output label. The dimensions of beta are 4x1.

* Testing:

In this we use the computed beta values in the training set and apply it to the Test set. We use the formula Ybar=beta\*T where T is the test set.

* Evaluating:

By comparing the Ybar values with test labels set, we check for the accuracy of the model which is discussed below.

1. Results:

The final answer is the result of all the iterations. I have used 5-fold cross validation to get a quantitative value which will be displayed at the end of code. It is represented as a percentage.