Project Perceptron algorithm and Logistic classification

DryBean.xlsx contains data for 7 different dry bean varieties (Seker, Barbunya, Bombay, Cali, Dermosan, Horoz and Sira). The Perceptron algorithm and Logistic regression models were used to classify Seker vs Barbunya beans based on the dimensions and the shape forms. We compared the results with each method. It was shown that the more accurate classification is the single layer classification method than the Logistic regression method according to the given data set.

Introduction

This project is about classifications of Dry Beans. This problem is quite famous among most data scientists. This is a classification problem by using the given data. There are 13611 data rows in this data set. There are 7 different types of beans. Seker, Barbunya, Bombay, Cali, Horoz, Sira and Dermason are the types of beans in this data set. Also there are different types of dimensions and shapes of each bean. Area, Perimeter, Major Axis Length, Minor Axis Length, Aspect Ratio, Eccentricity, Convex Area, Equiv Diameter, Extent, Solidity, roundness, Compactness, ShapeFactor1, ShapeFactor2, ShapeFactor3, ShapeFactor4, Class are the main measurements were obtained by the data set.

The goal is to classify the two types of beans by using the given data set. Seker, Barbunya are the two types of beans that we need to classify. Perceptron method and Logistic growth methods were used to analyze the problem.

Perceptron is the single layer neural network. The main components of a single layer neural network are inputs, weights, summation function, activation function and outputs. Single layer neural network is shown in the fig 01. According to our problem there are 16 inputs. From these 16 inputs we identify the type of bean. We use $x_1, x_2, x_3, \ldots x_{16}$ as inputs and $w_1, w_2, w_3 \ldots w_{16}$as weights. Also (b) as a bias value. We used forward activation function to evaluate the inputs and found the output as an identified bean. There are mainly 3 types of activation functions sigmoid function, tanh function and ReLU Max(0,x) function. If we use activation function, the calculated answer is not the final answer we already have. So there is an error, E=(Y-y). So, in perceptron model, there is a back propagation method to reduce the error. The back propagation definition is recursive application of the chain rule along a computational graph to compute the gradients of all inputs, parameters. So this method will give the good estimation for classification of dry beans.

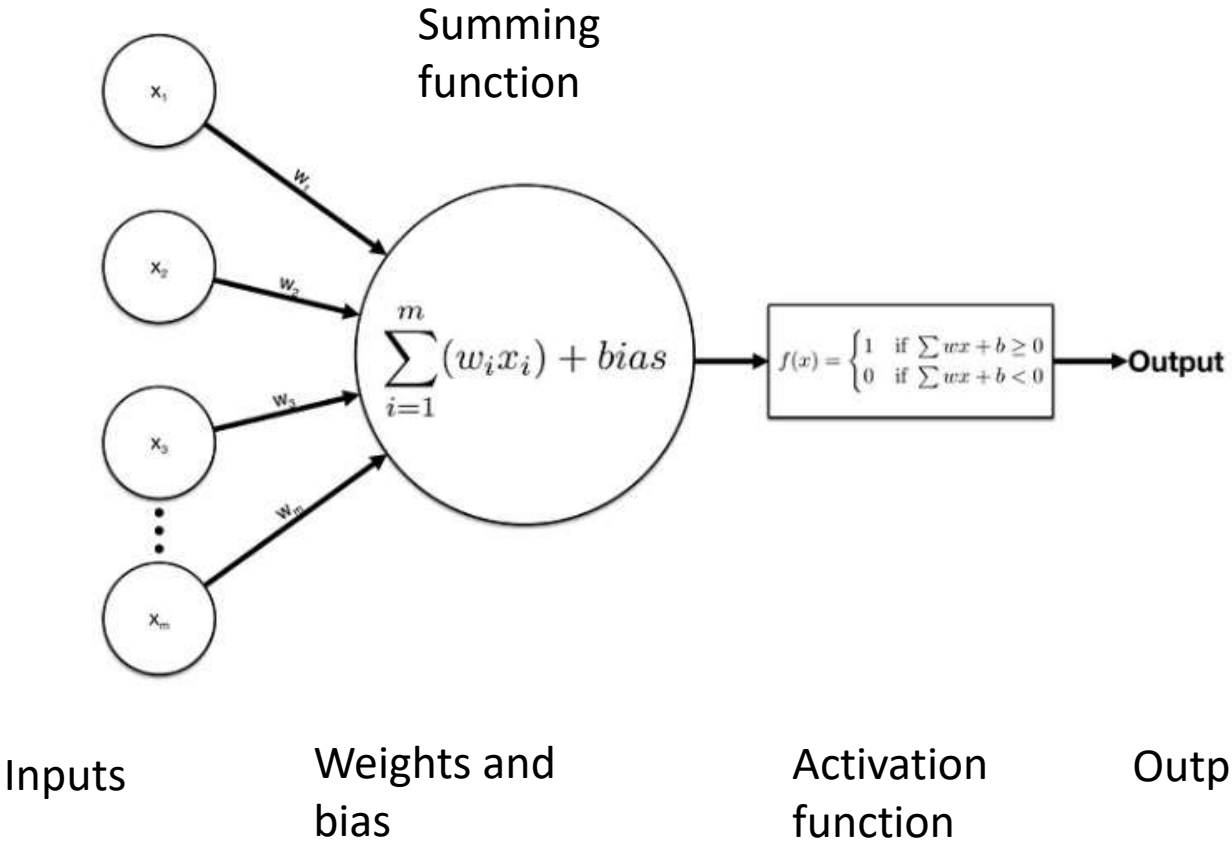


Fig 01. Single layer perceptron model

Logistic growth model is the conventional method for classification by using differential equation modeling. The logistic growth equation is

$$P(x) = \frac{1}{1+e^{-(\beta_0+\beta_1 x)}}$$

This logistic regression used in many classification models.

Methodology

Reading data

Reading an excel filefiles

```
data = pd.read_excel('D:\\WIU\\2023 Fall\\MATH 589\\Worksheet 3\\DryBean.xlsx')
print(data.head())
print(data.tail())
```

Data cleaning

a. In this data set checked the rows with missing data

```
print(data[data.isnull().any(axis=1)])
column_mapping = {col: i for i, col in enumerate(data.columns)}
```

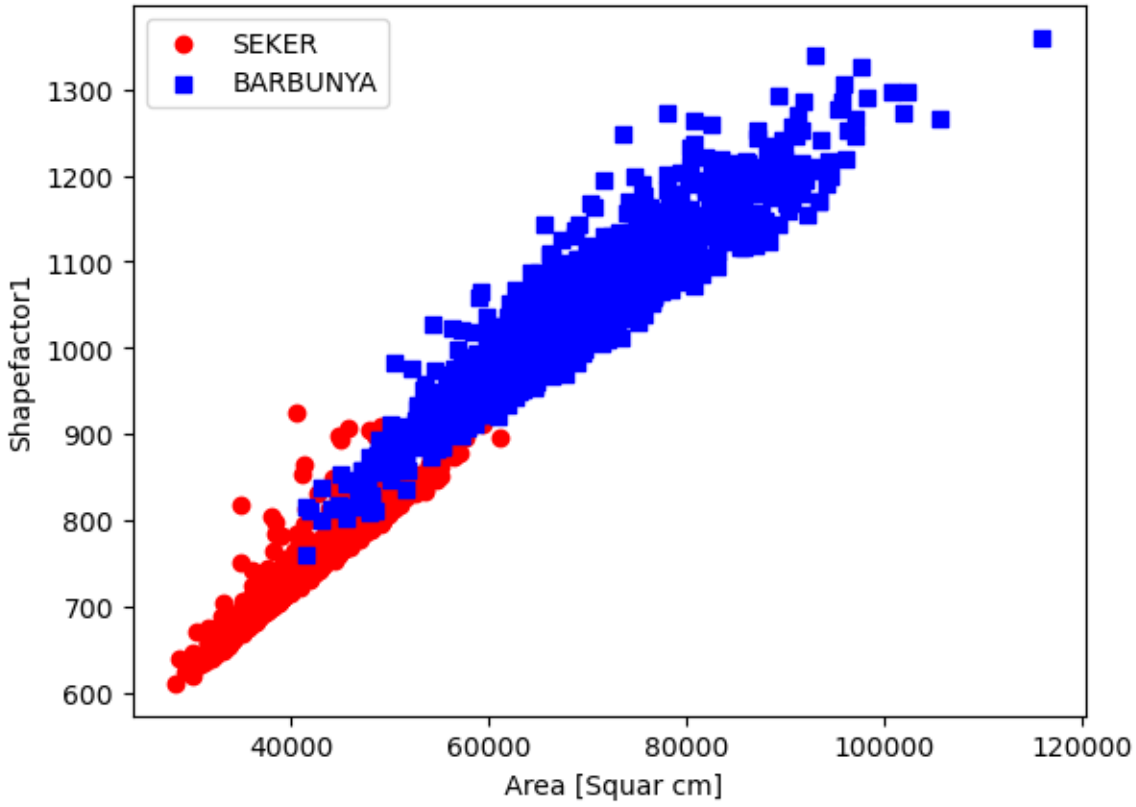
Data organization

Then changed the column names to index numbers. This is for easy understanding of codes. Rename the columns

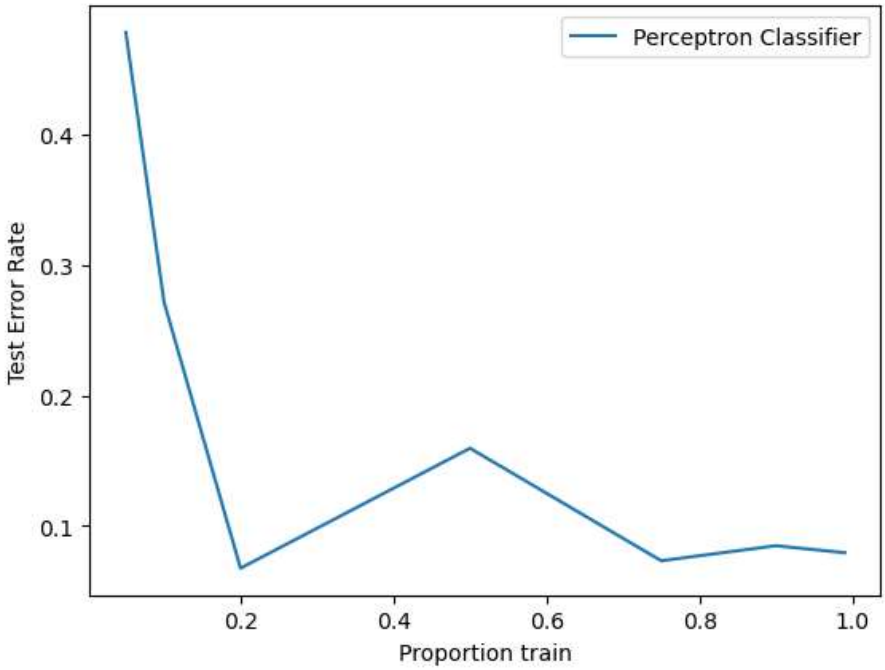
```
data.rename(columns=column_mapping, inplace=True)

# Print the DataFrame with columns renamed to numbers
print(data)
```

Analyzing Data

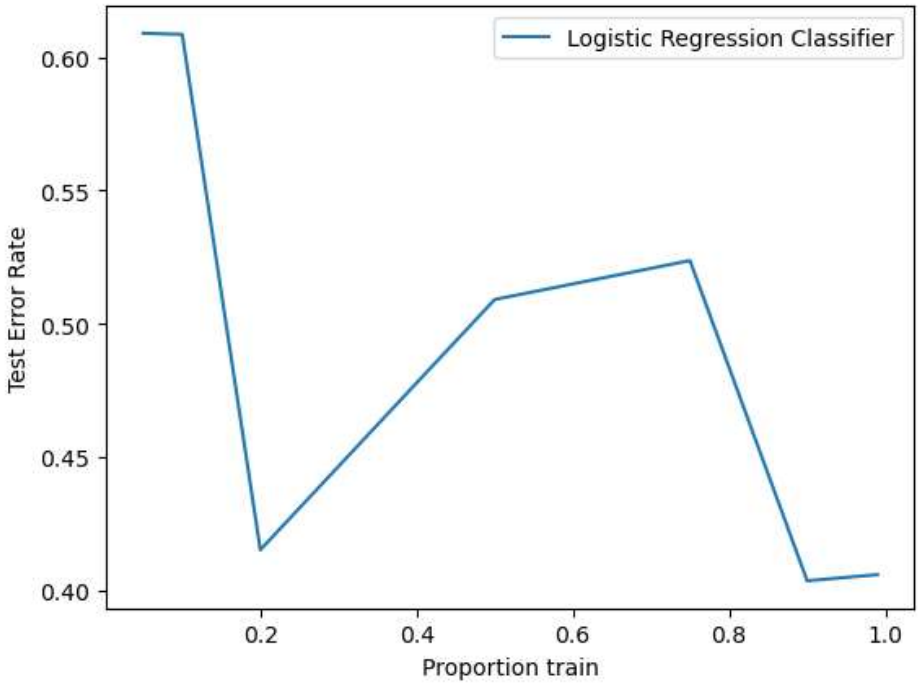
Then dropped the other types of bean data than Seker and Babunya.



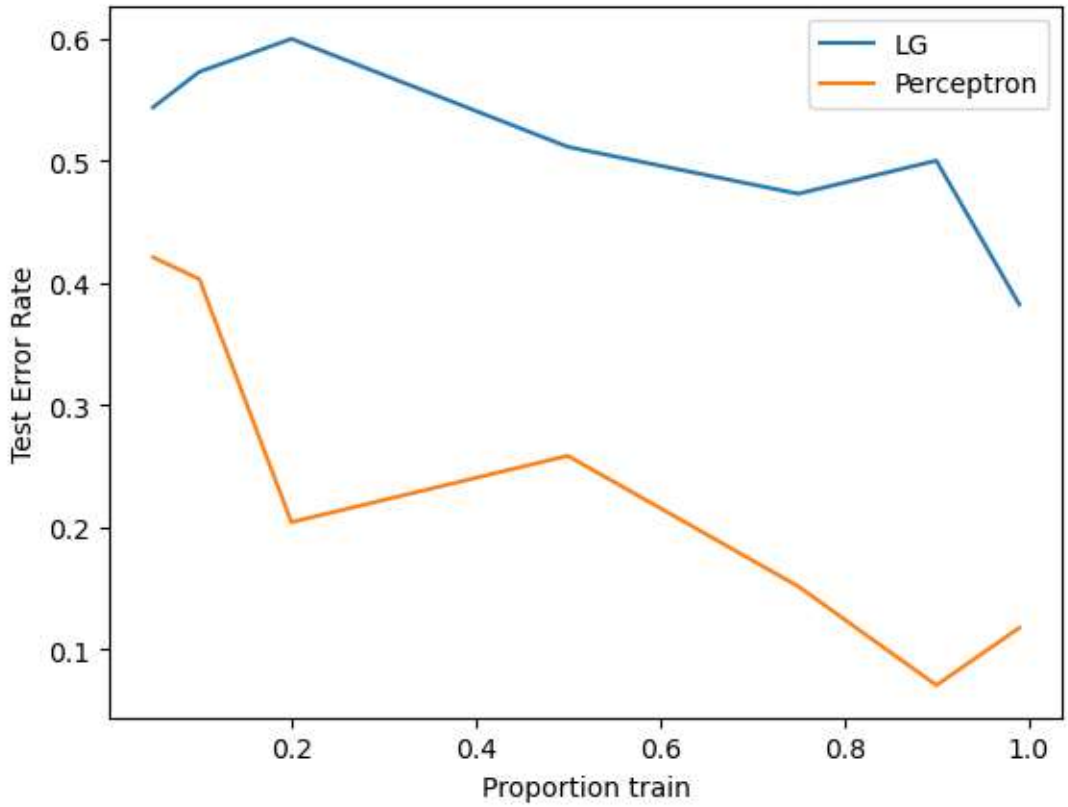
.
Then we used perceptron model,



Then we used the Logistic regression model



We compared the models



CONCLUSION

These two models classified the dry beans Seker and Babunya. Fig 3 and 4. Each classification method has errors according to the given proportionate data. The perceptron classification gives the least test error near the 0.2 to 0.1 range. When the proportionate high the error goes down and the half proportionate shows the much higher error in this model.

In the logistic regression model, the test error range is 0.55 to 0.40. This method also shows the least error when the proportionate goes high. When the proportion became 0.5 it was shown that the error rate. By comparison of these two methods, it can be seen that the single layer perceptron model is more accurate than logistic regression model.