Machine Learning Assignment 2

Group Members:

S.Bhaala Natarajen 19S005

N.Harish Kumar 19S012

S.Pon Rishikesh 19S024

# Dataset Description:

The dataset includes information about the shape of the bill, as well as the label. It is made up of 200 banknotes in total, 100 for genuine/counterfeit each.

Attributes:  
-counterfeit: Whether a banknote is counterfeit (1) or genuine (0)  
-Length: Length of bill (mm)  
-Left: Width of left edge (mm)  
-Right: Width of right edge (mm)  
-Bottom: Bottom margin width (mm)  
-Top: Top margin width (mm)  
-Diagonal: Length of diagonal (mm).

# Algorithms used:

Logistic Regression:

Logistic Regression is a machine learning algorithm which is used for the classification problems. It is a predictive analysis algorithm and based on the concepts of probability. We can call Logistic Regression as a Linear Regression model but the Logistic Regression uses more complex cost function, this cost function can be defined as the ‘sigmoid function’ or also known as ‘logistic function’ instead of a linear function.

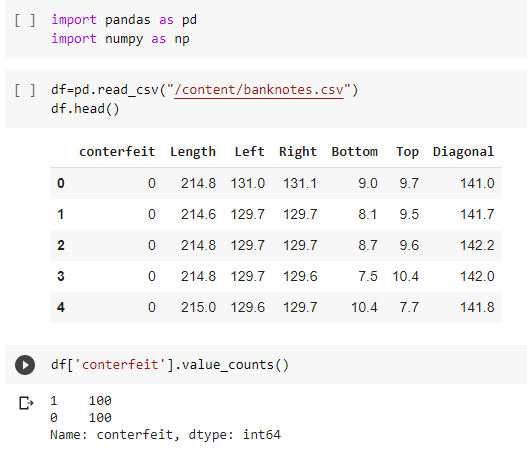
Gradient Descent:

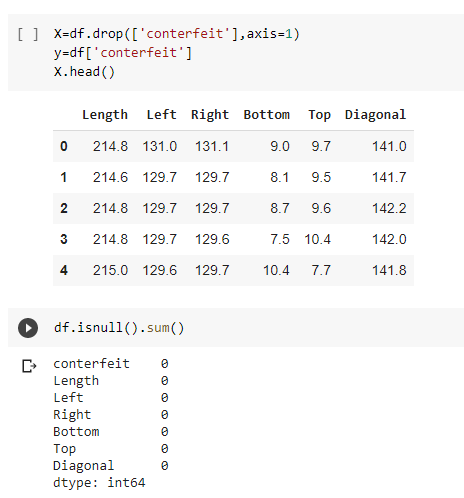
Gradient Descent is an optimisation algorithm that follows the negative gradient of an objective function in order to locate the minimum of the function. It is a simple and effective technique that can be implemented with just a few lines of codes. It also provides the basis for many extensions and modifications that can result in the better performances. The algorithm also provides the basis for widely used extensions called Stochastic Gradient Descent used to train Deep Learning Neural Networks.

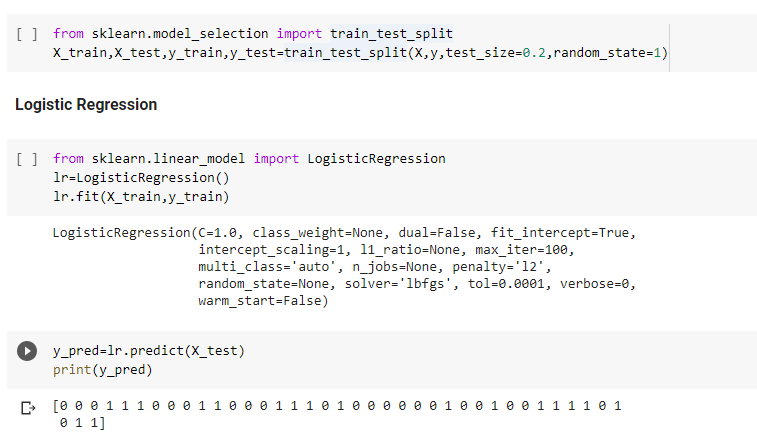
Linear Discriminant Analysis:

Linear Discriminant Analysis (LDA) is a dimensionality reduction technique. As the name implies, this technique reduces the number of dimensions (i.e. variables) in a dataset while retaining as much information as possible. Linear Discriminant Analysis or LDA uses the information from both features to create a new axis and projects the data to the new axis in such a way as to minimize the variance and maximise the distance between mean of two classes.

# Code and Inference:

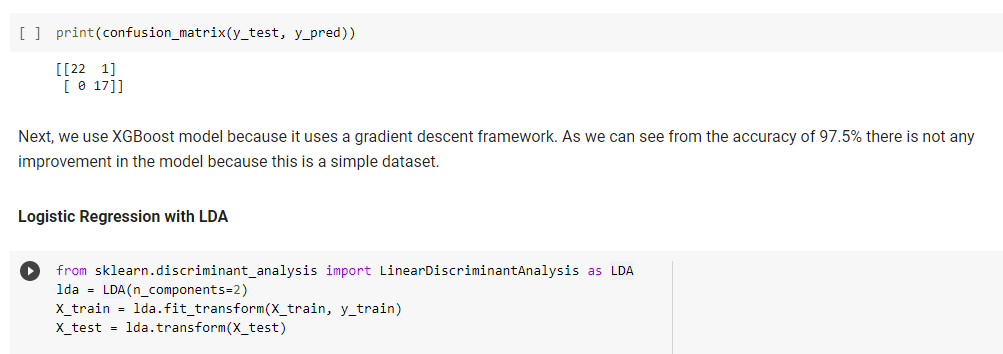


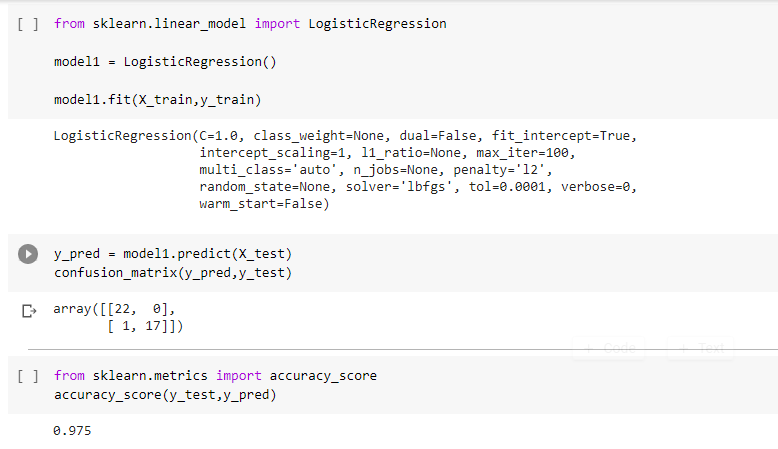


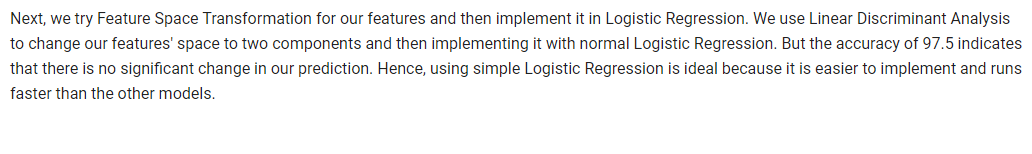












# Conclusion:

Hence, using simple Logistic Regression is ideal because it is easier to implement and runs faster than the other two models.