**Student Name 1: Nishanth Marer Prabhu**

**NUID 1: 002624650**

**Student Name 2: Alexander Motes McNeil**

**NUID 2:**

Course: Introduction to Machine Learning and Pattern recognition

College: College of Engineering

**Final Project**

In this project we are trying to determine if based on the given data which team is going to win the match.

We will try to implement this using two process Expected Risk Minimization technique and a Multi-Layer Perceptron. We will use the data which contains 30 features; however, we will try to reduce the feature set by determining the correlation matrix.

Correlation Matrix:

Where, represent the variance of x and y

Furthermore, we will calculate the threshold for the correlation matrix knows given by the equation below:

Where N, represents the sample size.

If any entry in the correlation matrix between the resultant ‘Y’ output vector and the feature vector has a value greater than ‘r’, then we will consider that feature, however if the value of matrix between two features is greater than ‘r’, then there exists multi-collinearity.

To solve this, we will determine the partial co-relation co-efficient given by the equation below.

Where,

Once the above values are determined for the features vectors with multi-collinearity, we will select only the required features and proceed further.

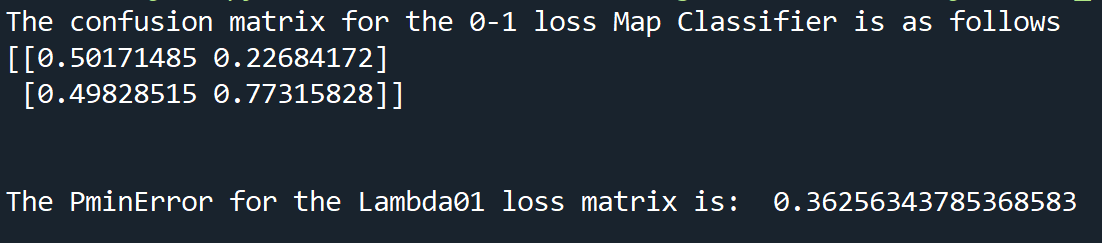
In this problem, we selected only **10 features** which showed significance.

**Algorithm:**

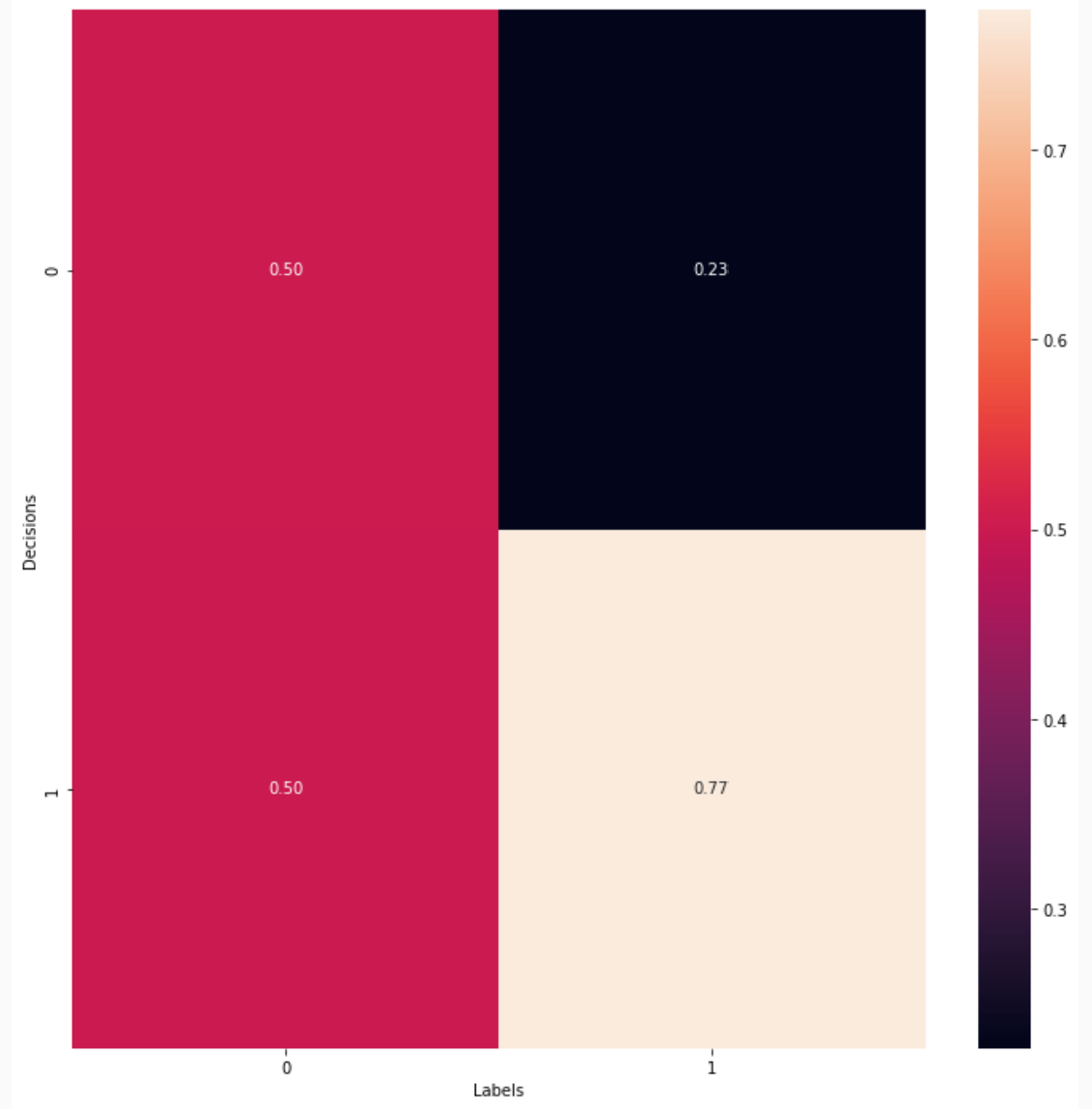
1. Read the data from the csv file and remove all the unwanted feature vector and consider only the required feature vectors.
2. Determine the class labels and further determine the class posterior values
3. Now find the mean and covariance matrix for each of the labels
4. Regularize the covariance matrix to ensure that they are invertible
5. Now, evaluated the PDF and obtain the result
6. In this process, we already have the class priors provided, now we need to compute the:

1. Then, we need to compute the product of the class priors with the class conditional pdf
2. Once we have the above product, we need to divide it by , to obtain the class posterior value
3. To determine the Risk Matrix, we need to multiply the Class Posterior with the loss matrix
4. Once the risk matrix has been calculated, we need to pick the decision with least risk for each data point.
5. Further we need to compute the confusion matrix and plot the graph which shows the correct and incorrect decisions made by the classifier.
6. Confusion Matrix: (The Column is the label and the rows are the decision made)

**Sample Run:**

****

**Confusion Matrix:**

****