COMP6245 Foundations of Machine Learning – Lab 2

Supritha Konaje
Email: ssk1n21@soton.ac.uk
Student ID: 32864477

Linear Classification using Perceptron Algorithm

Linear Classifications use a linear combination of input data to categories them into labels. We can plot them in a plane or line. When data is linearly separable, we can use linear classification.

1. Applying Perceptron Algorithm

One of the algorithms for linear classification is Perceptron Algorithm. Here, we take the input vectors(x) that have their weight(w). The result after we get the value from $w^T x$, helps us decide which of the plane x will lie.

We first initialize, w with a random vector. Then we iterate over the value over the data that is both the negative and positive data. The whole concept of the algorithm is based on the value that we get after performing the dot product of w and x. If it is less than 0 or greater than 1, then we will be updating the randomly initialized w by subtracting or adding the x else we won't be changing it.

We have two bi-variant Gaussian sample of size 100 each. Means as m1 = $\begin{bmatrix} 0 \\ 5 \end{bmatrix}$ m2 = $\begin{bmatrix} 5 \\ 0 \end{bmatrix}$. The covariance matrix is C = $\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$. We first plot the scatter plot for the samples. Using the algorithm, we will be training the test data.

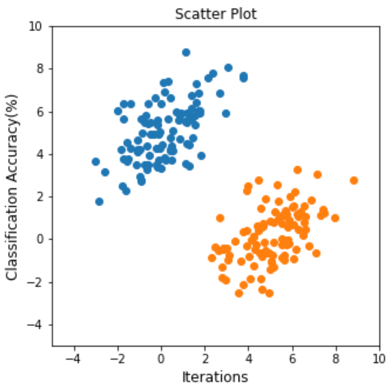

Figure 1: Scatter Plot for Gaussian Densities

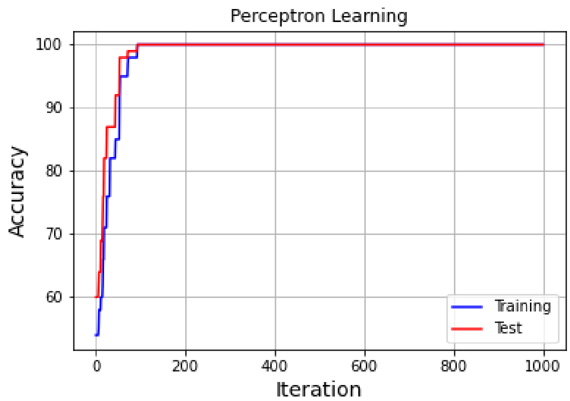

Figure 2: Training Phase

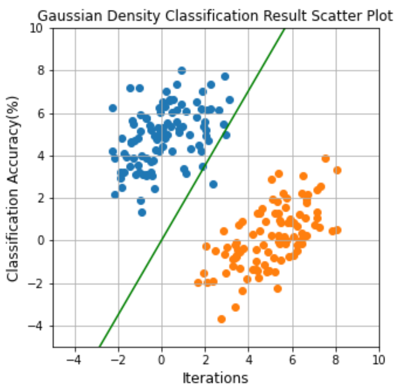

Figure 3: Classification Result of Gaussian Densities

The means were changed for the next task that are plotted graphs in Figure 4, Figure 5 and Figure 6. Means for the new task are m1 = (2.5,2.5) and m2 = (10,10). To get the decsion boundary, we need to add a constant so that we get the linear equation for the boundary line. This type of classification can be done for the distributions whose line doesn't pass from the origin as we can see from Figure 5.

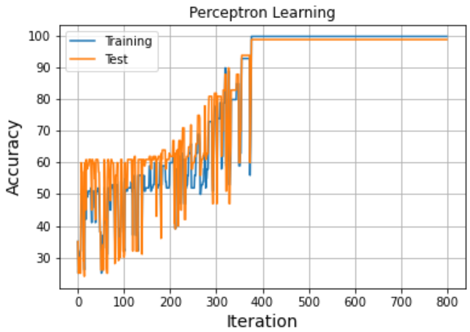

Figure 4: Training phase with accuracy 100% and 99%

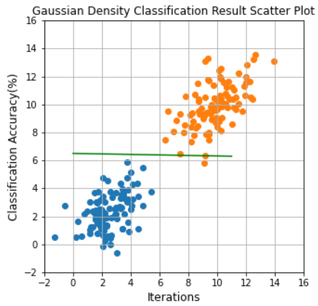

Figure 5: Classification Result

2. Bigger Dataset for Banknote Authentication:

We were given the task to apply Perceptron Algorithm for a bigger data set that we took from UCI Machine learning Repository. The dataset was taken from [1] https://archive.ics.uci.edu/ml/datasets/banknote+authentication.

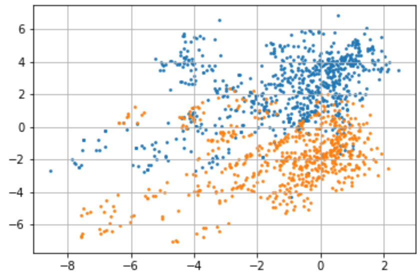

Figure 6: Orange Scatter plot represents real notes, Blue scatter represents fake notes.

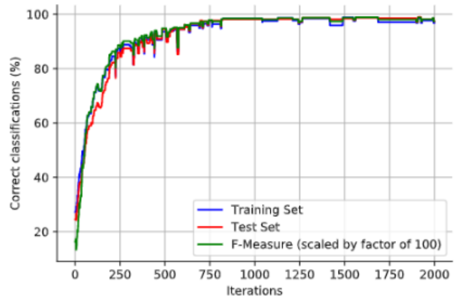

Figure 7: Training Phase.

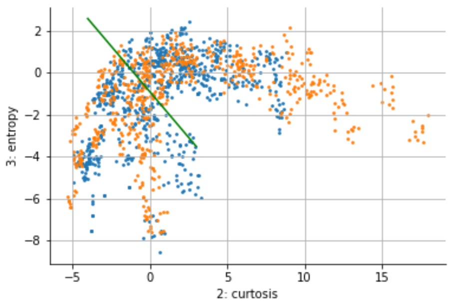

Figure 8: One of the attributes of image (Entropy)

In Figure 8, we get one of the attributes defined by the data set. Similarly, we can get the other attributes by substituting the values for the variables to plot the decision boundary.

Reference:

[1] Dua, D. and Graff, C. (2019). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science