Pattern Recognition System

Categorization of input data into identifiable classes

Vinay Mohan Behara

CEG7570 -Pattern Recognition

Wright State University

1. Abstract

Pattern recognition can be defined as the categorization of input data into identifiable classes via the extraction of significant features or attributes of data from a background of irrelevant details.

2. Introduction

The project is divided into three parts. In the first part, we normalize the features, remove the outliers and plot the graphs so that we can distinguish classes from each other. We develop methods that will normalize and display the data set.

3. Approach

The dataset with 1372 observations, 4 features and 2 classes is imported. The steps involved in this process are as follows:

Step1:

Once the data set is imported, in the process of normalization, we find the Mean, Variance and normalize these values to 0 and 1 respectively.

The formulas used are

$$\bar{x}_k = \frac{1}{N} \sum_{i=1}^{N} x_{ik}, \quad k = 1, 2, \ldots, l$$

$$\sigma_k^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_{ik} - \bar{x}_k)^2$$

$$\hat{x}_{ik} = \frac{x_{ik} - \bar{x}_k}{\sigma_k}$$

In the above step we calculate the mean , variance and normalize these values. These normalized values are taken into a matrix.

Step 2:

After normalizing values, the process involved in removing outliers is if the normalized observations are greater than 3.0 or smaller than -3.0, we remove those observations in entire feature set. That is if we have these conditions satisfied, we remove the entire row in implementation. Now we take the output of all these values and form a matrix along with the class added to next row of the created matrix.

Step 3:

We can display the remaining observations in a two dimensional feature space. There are four features which have normalized observations. The user is asked to select any two observations with his choice on x axis and y axis respectively. Red and Blue colours are used to distinguish the classes from each other.
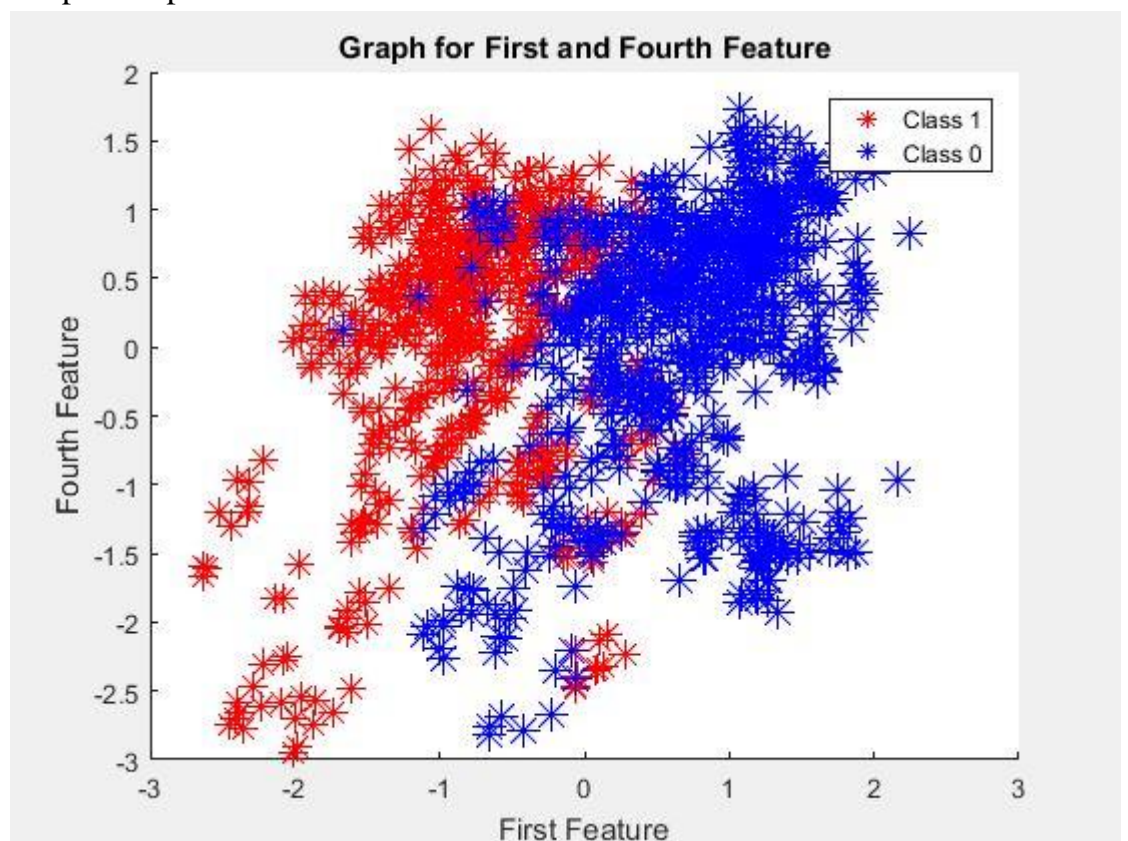
4. Test Cases
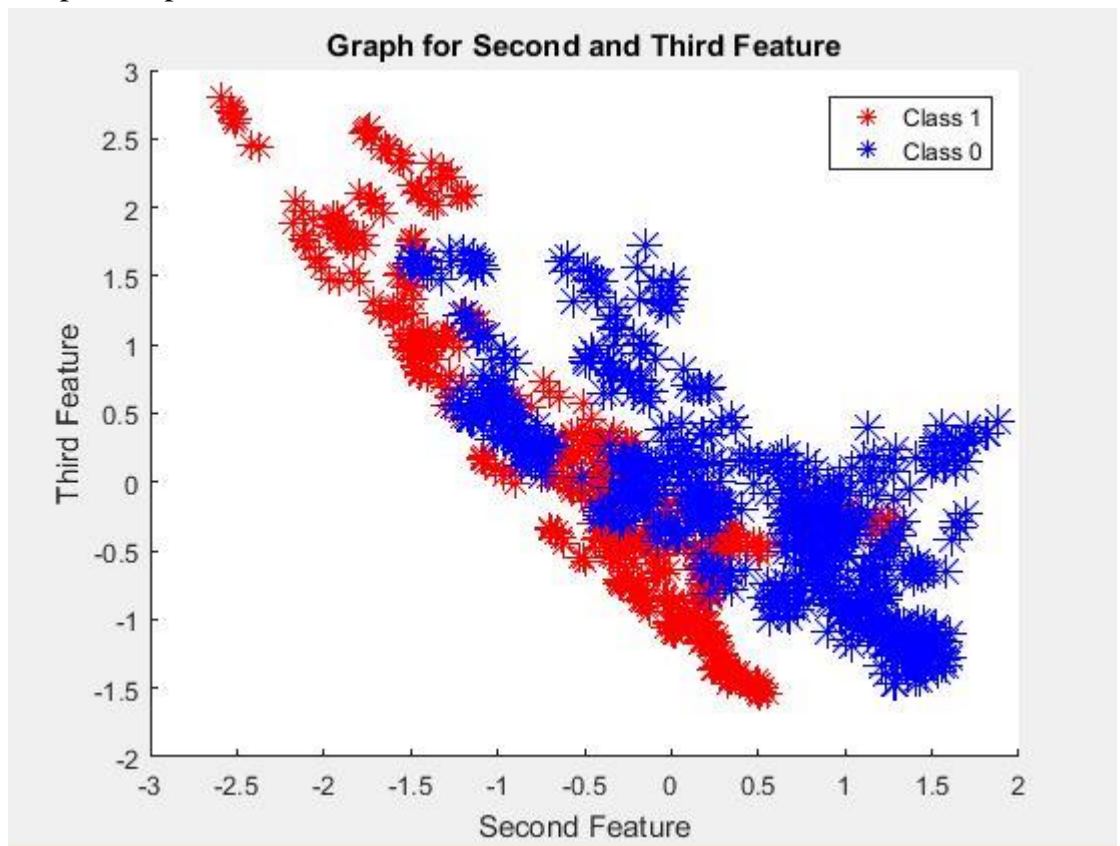- Output Graph for first feature and second feature
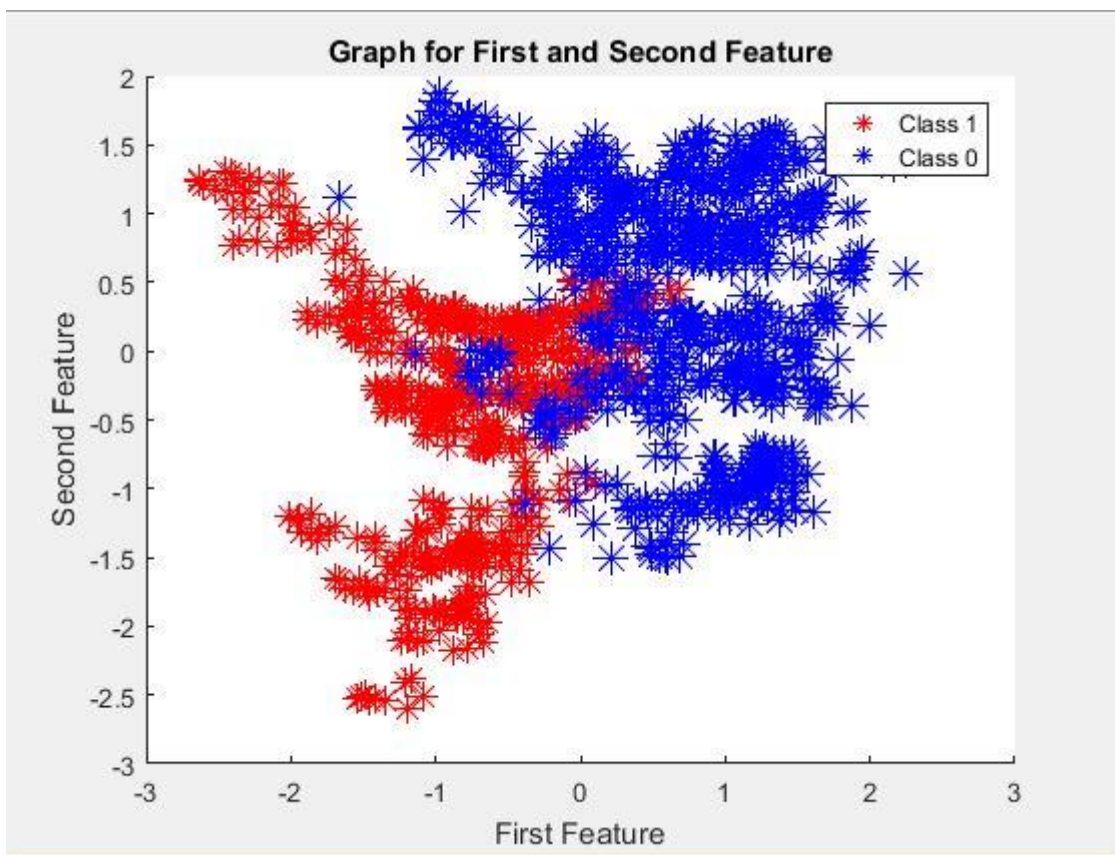
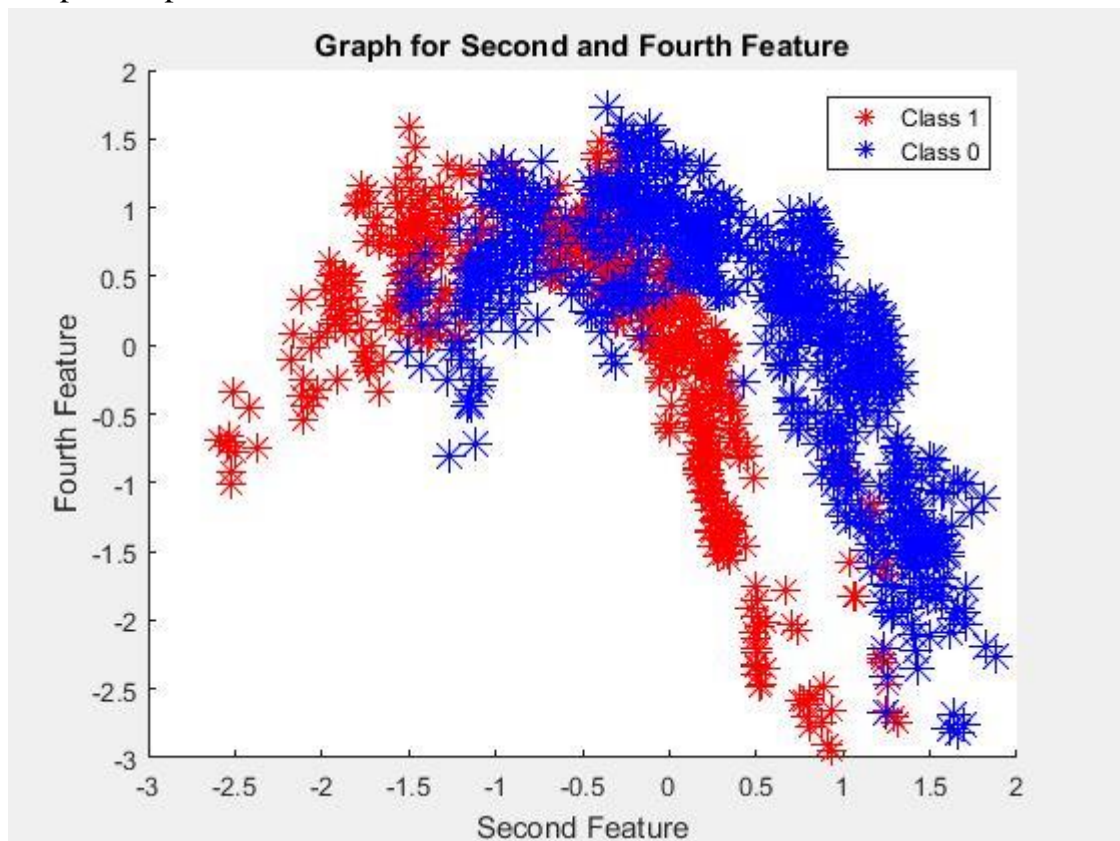- Output Graph for first feature and third feature
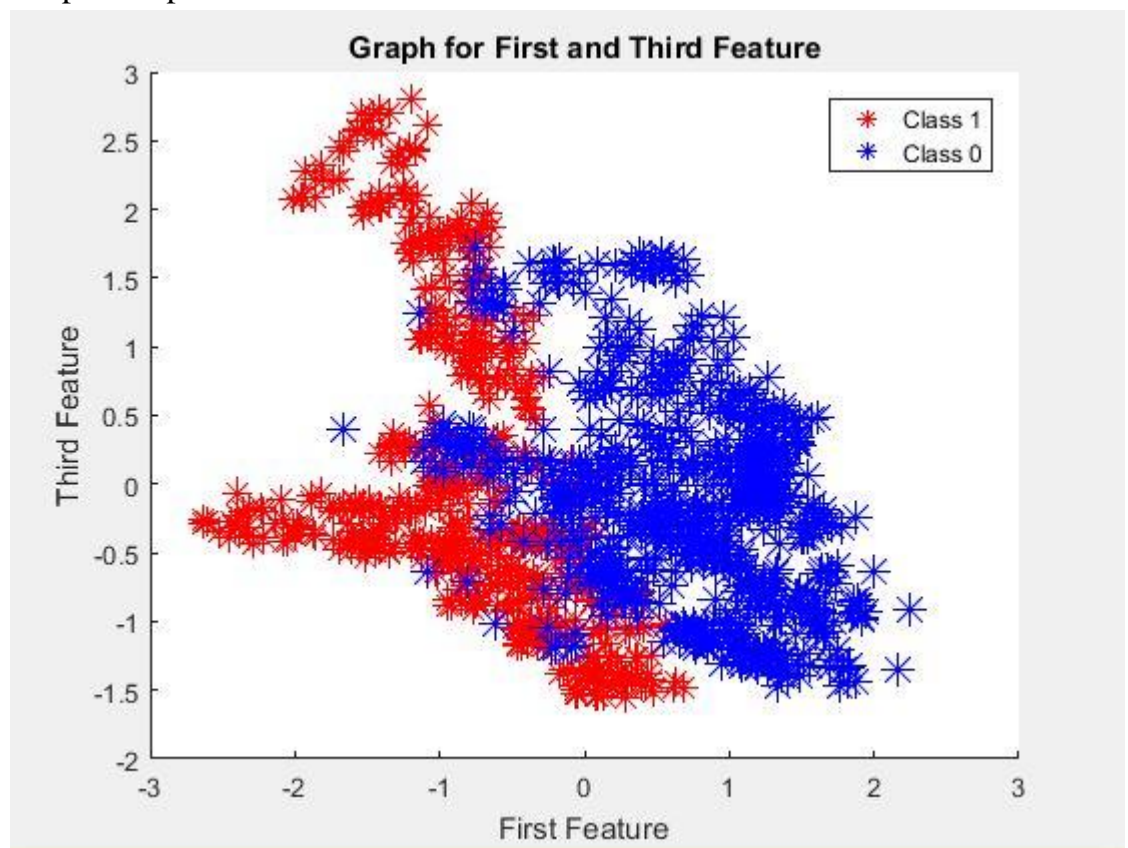
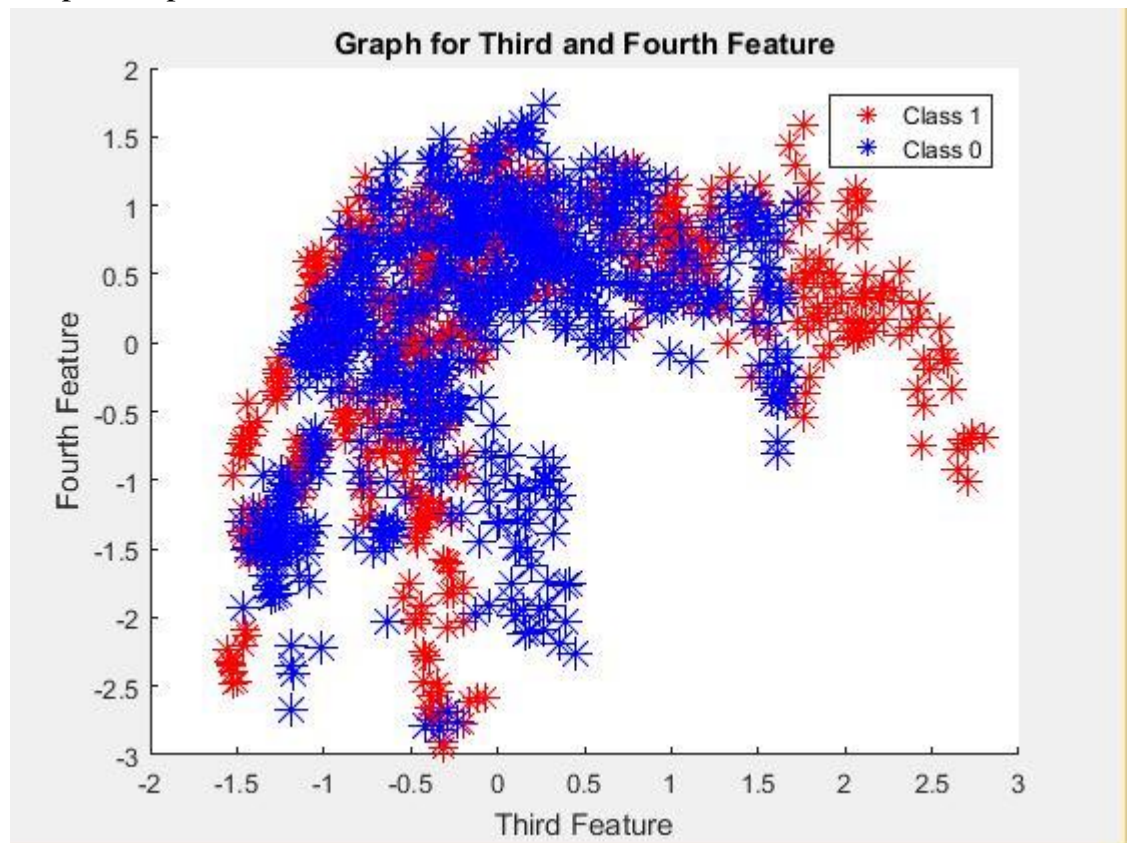- Output Graph for first feature and fourth feature

- Output Graph for second feature and third feature

- Output Graph for second feature and fourth feature

- Output Graph for third feature and fourth feature

Graph for Third and Fourth Feature

5. Conclusion

The main goal of Normalization is to remove outliers. The plotting of graph is done with the remaining values without outliers in a two dimensional space.