Pattern Recognition System

(Reduction of Features and Calculating the Squared Error and Percent Error)

Vinay Mohan Behara

Cs 7570- 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. In this second part, the features are reduced from d to 2. The Squared Error and Percent Error are calculated taking the new generated two feature set into consideration.

3. Approach

In the previous project we Normalized the values and removed the outliers. In this part we will import the normalized values from previous project and calculate the Squared Error and Percent Error from the obtained two feature values.

The formulas used are

Squared Error = Sum of two smallest Eigen Values

Percent Error = ((Sum of two smallest Eigen Values)/(Sum of all Eigen Values))*100

Step 1:

Once the normalized data without outliers is imported, we calculate the Covariance Matrix. For calculating Covariance Matrix we have a function named ‘cov’ in Matlab.

Step 2:

The Covariance Matrix is used to generate the Eigen Vectors and Eigen Values using the function ‘eigs’ and these values are sorted.

Step 3:
The Eigen Values are sorted and the two highest Eigen values are taken. Their Corresponding Eigen Vectors are taken into single separate column Matrices. Then the imported data matrix is multiplied with these two largest Eigen Vectors.

Step 4:
By multiplying in this way we are able to reduce the features from d to 2 and the two new features are obtained. We can display these two features in a two dimension by plotting a graph between the two features.

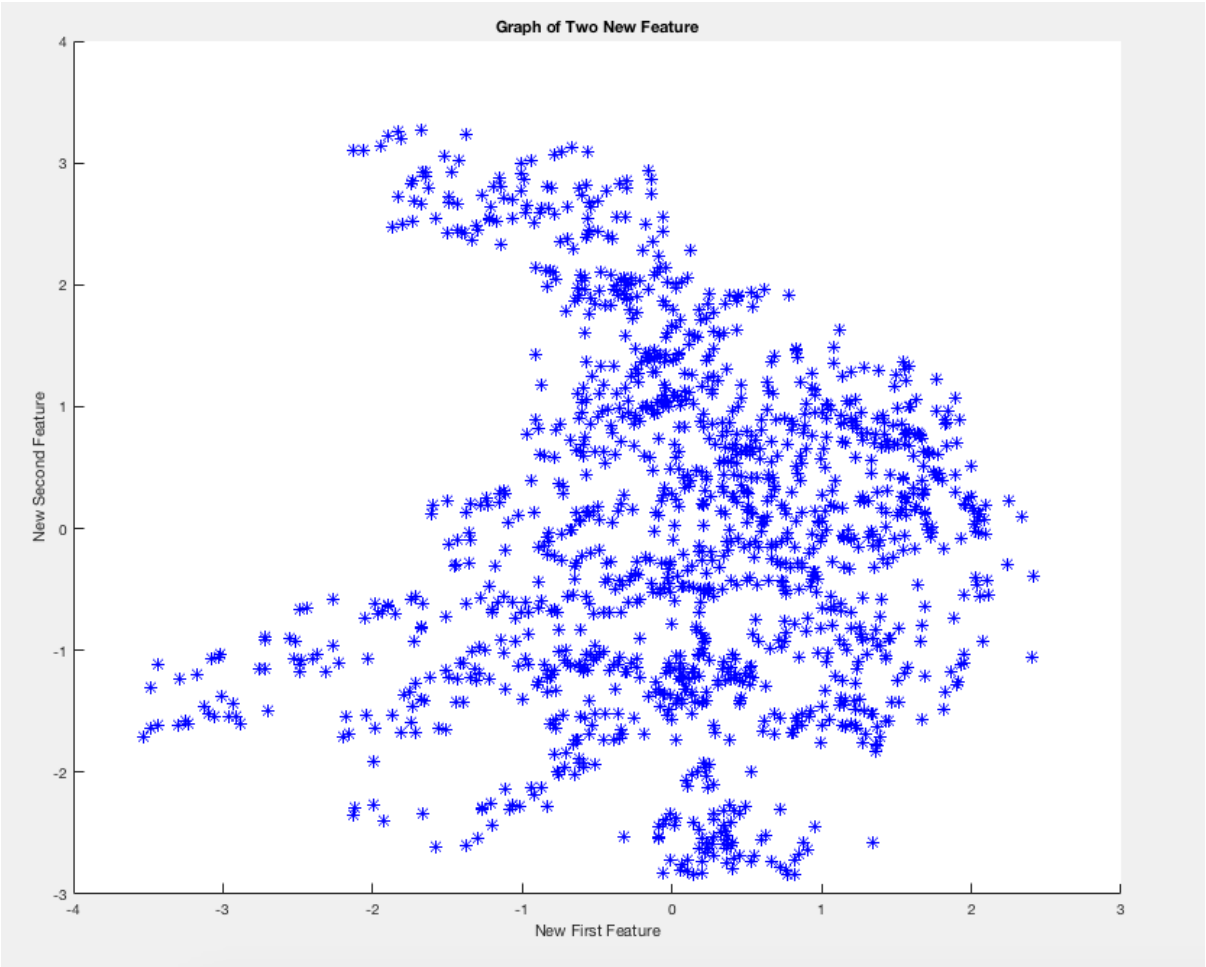
Step 5:
Now the Squared Error and Percent Error is calculated using the formulas .and the Percent error should be less than 50 percent for a decent result.

4. Output:

```
Command Window
   -0.0097   -0.0977   -0.6610    0.3010
   -0.6413   -1.1590    1.1790    0.7301
   -1.4067   -1.7565    2.5520   -0.0432
   -1.0471   -0.4398    0.2986    1.1365


Covar =

    0.9653    0.2311   -0.3145    0.2283
    0.2311    0.9049   -0.6689   -0.5023
   -0.3145   -0.6689    0.8408    0.3330
    0.2283   -0.5023    0.3330    0.9068


EigenVectors =

   -0.0564   -0.5369    0.8086   -0.2338
    0.7526    0.1584   -0.0269   -0.6386
    0.6016   -0.5266   -0.1388    0.5843
    0.2616    0.6398    0.5711    0.4429


EigenValues =

    0.1783         0         0         0
         0    0.3167         0         0
         0         0    1.1728         0
         0         0         0    1.9499
```

5. Test Case:

Graph for the new two features



6. Conclusion:

The Covariance Matrix is calculated for the output data set from previous project. The Eigen Vectors are calculated and the procedure is followed to reduce the features to 2 and the Squared Error and Percent Error are calculated.