

Fuzzy Logic Homework2

Covid Detection

Instructed by Dr.Tahayori

Reza Tahmasebi

saeed77t@gmail.com, Stu ID: 40160957

Table of Contents

[1 Covid detection based on relation matrix 3](#_Toc122720959)

[2 Data preprocess 3](#_Toc122720960)

[3 Method to create relation matrix 3](#_Toc122720961)

[3.1 Method one numeric range 3](#_Toc122720962)

[3.2 Method two the average of each feature 5](#_Toc122720963)

[3.3 Method three PCA feature extraction 8](#_Toc122720964)

[3.4 Method four LDA feature selection 10](#_Toc122720965)

[3.5 Method five, the variance of each feature 11](#_Toc122720966)

# Covid detection based on relation matrix

This project aims to create a relationship matrix based on our dataset. And to predict our test data labels with the composition of the relation matrix and test data.

# Data preprocess

First, the column *SESSO*  has been changed; the Female became 0, and the male became 1.

The second problem was the missing values that existed in the data set. I used the *KNNImputer* methodfrom *the sklearn* library to fill the blank values with the fournearest neighbors.

After that, I split the data to test data and train data (80% for training and 20% for testing). I used the *TestAndTrain* method from *PreprocessData* class to break my data.

# Method to create relation matrix

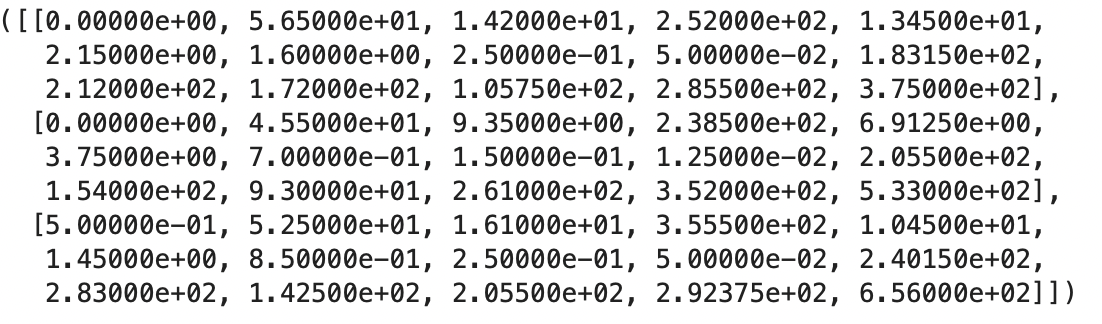
I implemented five methods to create my relation matrix, and each way gives different accuracies.

## Method one numeric range

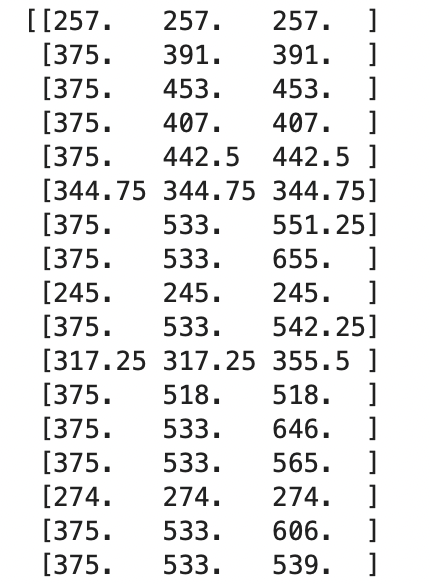
In this method, I split the training data into three parts based on the labels; training data became three parts because we had three labels.

After that, I determined a numeric range for each class of three parts. We had 15 features, and because of that, each matrix shape became 15x1.

The relation matrix was created with those three matrixes; the relation matrix shape was 3x15.

the relation matrix of method one

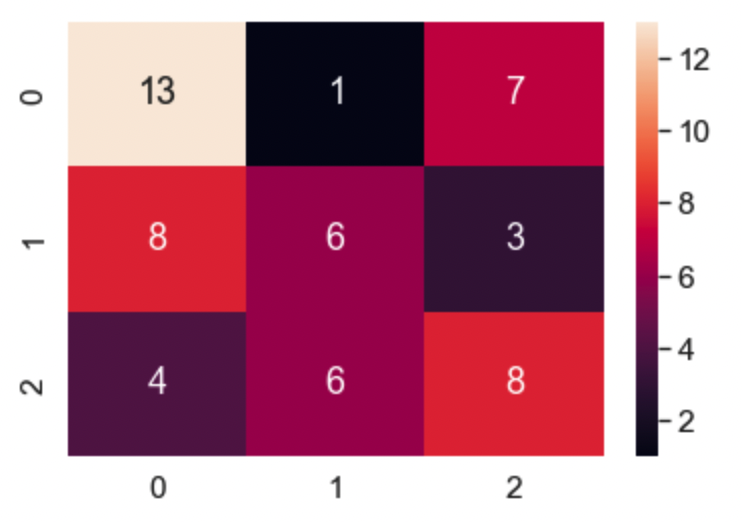
I used the *composition* method to decompose the relation matrix with the test dataset to create the Similarity matrix.



Example of method one Similarity matrix

After all those actions above, I ran the *argmax* method on each row of the similarity matrix to predict the labels.

The average accuracy of this method for the three classes was about 48.2%.

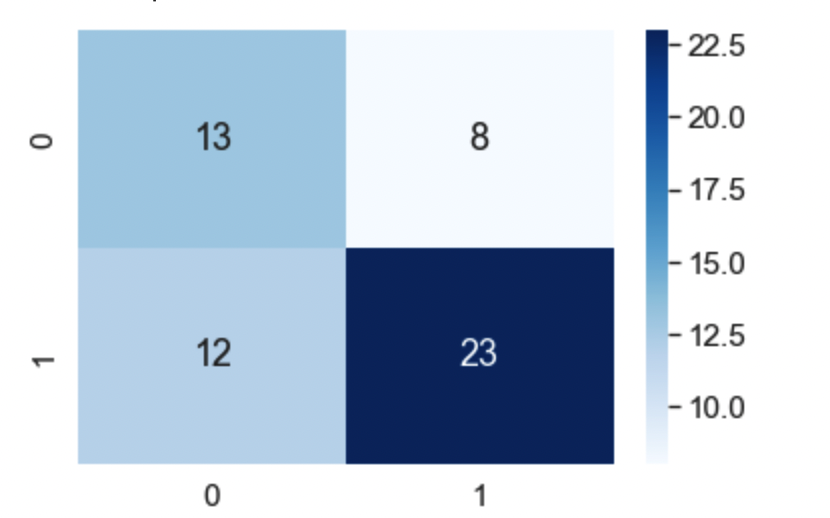


The confusion matrix of method 1 for three classes

Also, I took classes 1 and 2 as one, so class 0 is not sick, and class 1 is sick. And then, I applied method one to it. The results are as follows.

Average accuracy: 64.2%

And the confusion matrix is :



The confusion matrix of method 1 for two classes

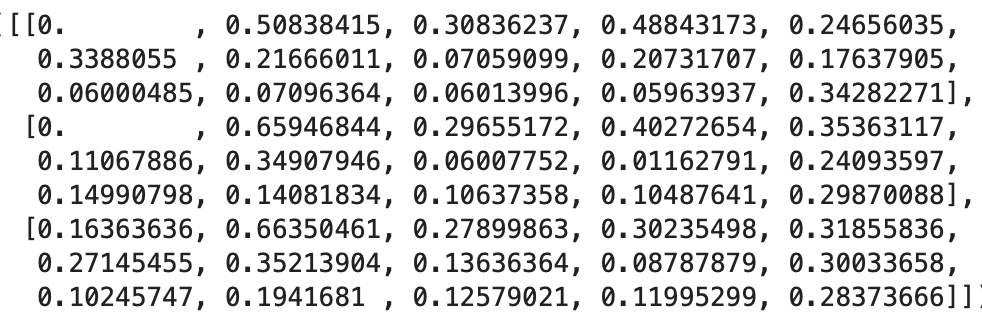
## Method two the average of each feature

In this method, I split the training data into three parts based on the labels; training data became three parts because we had three labels.

After that, I calculate the average of each part for all features. We had 15 features, and because of that, each matrix shape became 15x1.

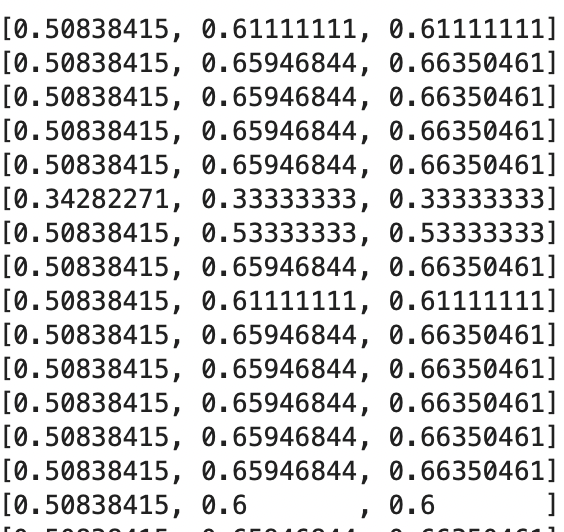
The data has been normalized in this method.

I put the matrixes together to create the relation matrix. The shape of the relation matrix is 3x15.



the relation matrix of method two

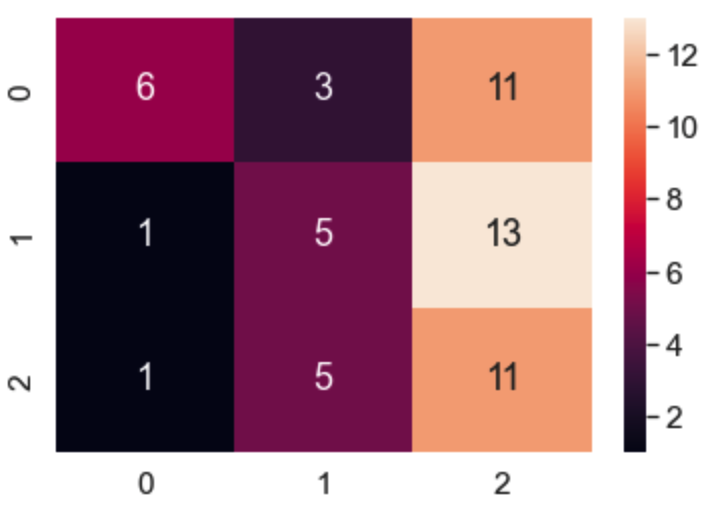
After creating the relation matrix, I used the *composition* method to decompose the relation matrix with the test dataset to create the Similarity matrix.



Example of method two Similarity matrix

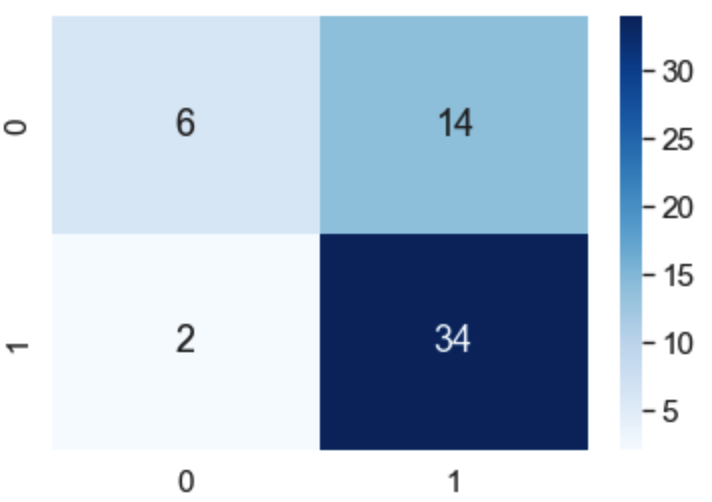
Like method one, the *argmax* method ran on a similarity matrix to predict the test data.

The accuracy of this model for three classes is 39.3%.



The confusion matrix of method two for three classes

I changed the three classes problem to two class problem and ran method two for it; the accuracy of method two for two classes problem is 71.42%, and the confusion matrix is :

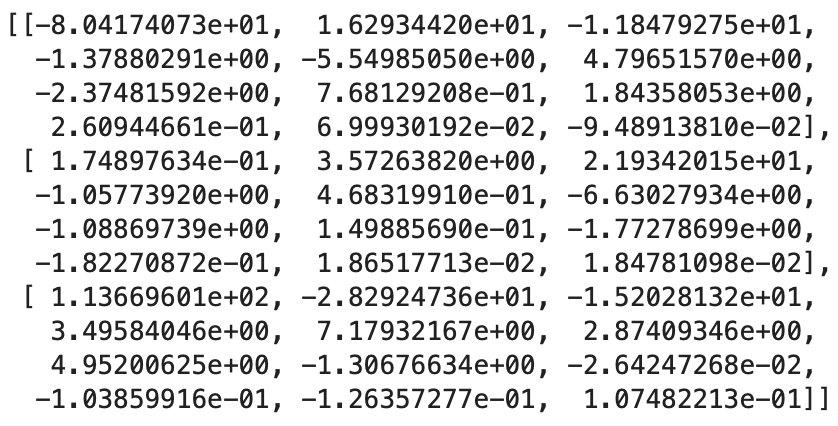


The confusion matrix of method two for two classes

## Method three PCA feature extraction

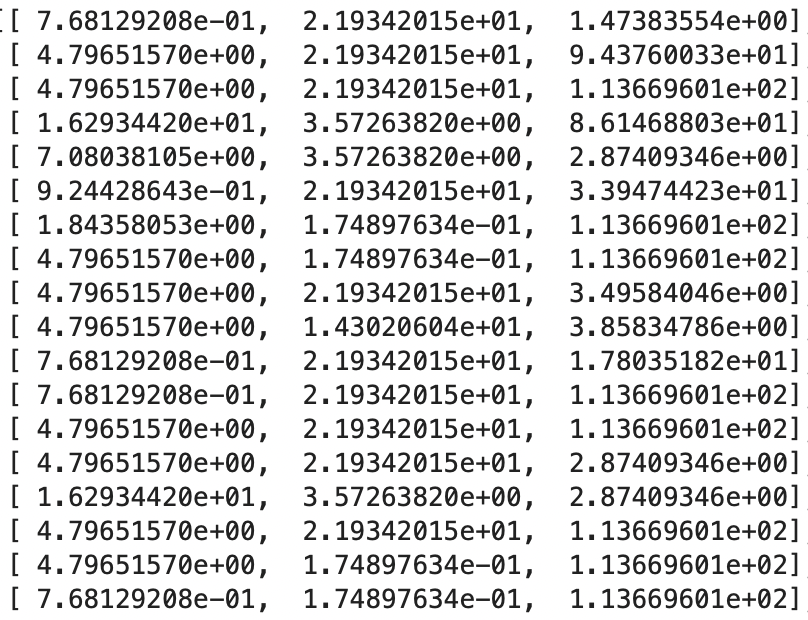
In this method, I used the *PCA*  method on the dataset, and then I split the data into three parts based on the class labels.

After that, I calculate the average of those three parts for each feature and put them beside each other to create the Relation matrix.



The relation matrix of method three

After creating the relation matrix, I used the *composition* method to decompose the relation matrix with the test dataset to create the Similarity matrix.

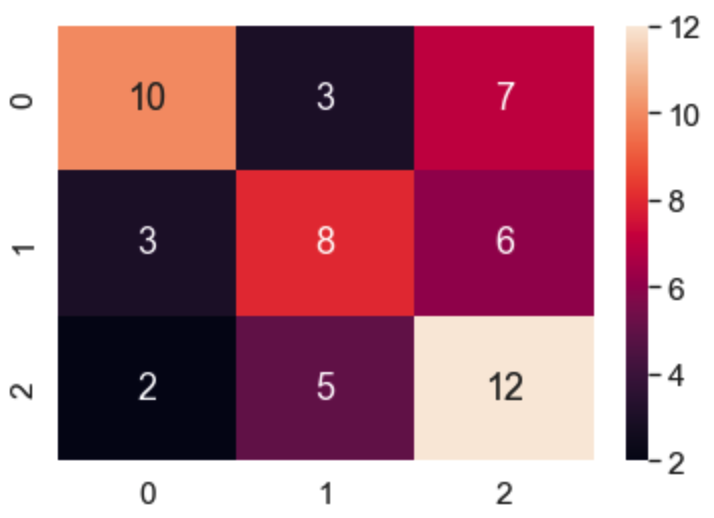


Similarity matrix of method three

And I used the argmax method to choose in what class our data belongs.

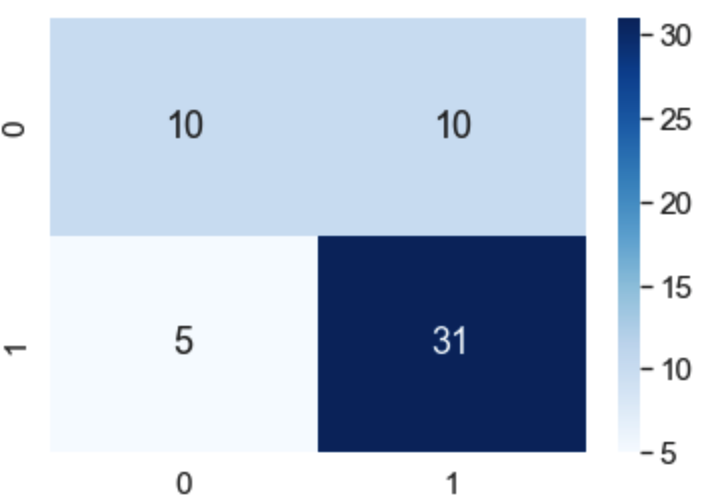
The accuracy of this method for a three-class problem is 53.5%.

The confusion matrix is :



The confusion matrix of three class problems for method three

This method has been running for two-class problems, and the accuracy of this method is:73.21%

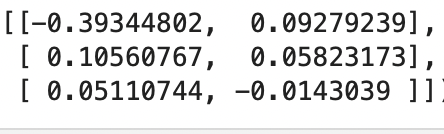


The confusion matrix of two class problems for method three

Note: PCA is a technique for unsupervised data, so PCA leads to overfitting.!

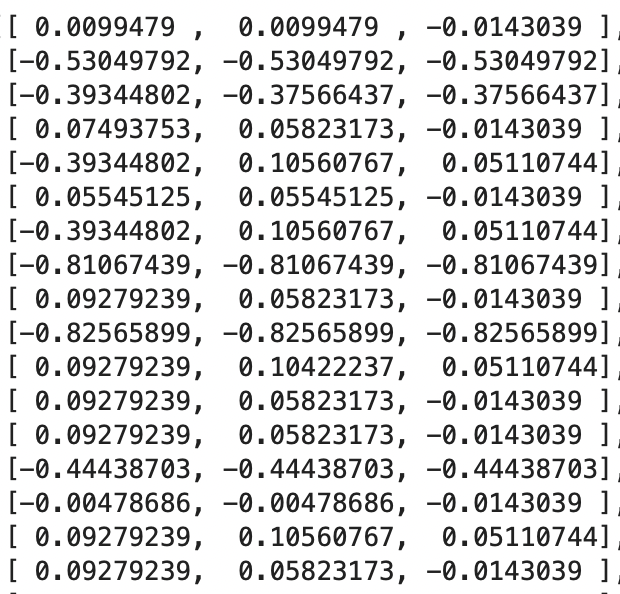
## Method four LDA feature selection

In this method, I used LDA to reduce the number of dimensions, and then like the other methods, I calculated the relation matrix with the average of each feature.



The relation matrix of method four

I used the *composition* method to decompose the relation matrix with the test dataset to create the Similarity matrix.



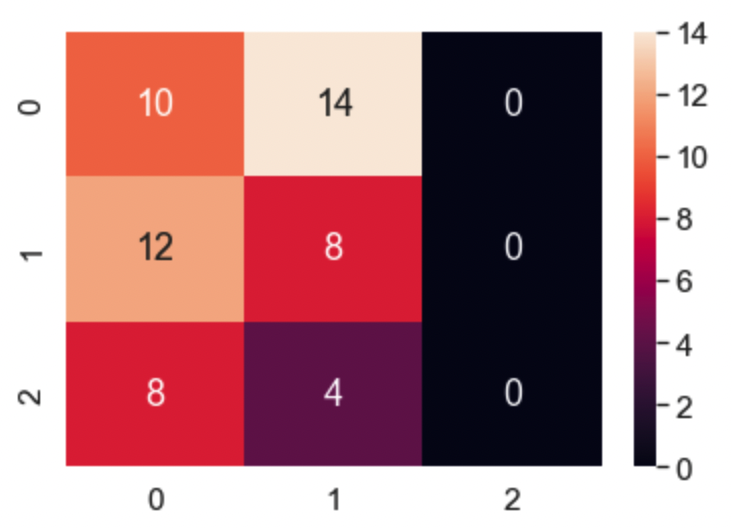
The similarity matrix of method four

And we predict the test data labels with a similarity matrix.

The results are as follows:

Accuracy: 32.1%

The confusion matrix :



The confusion matrix of method four, three classes

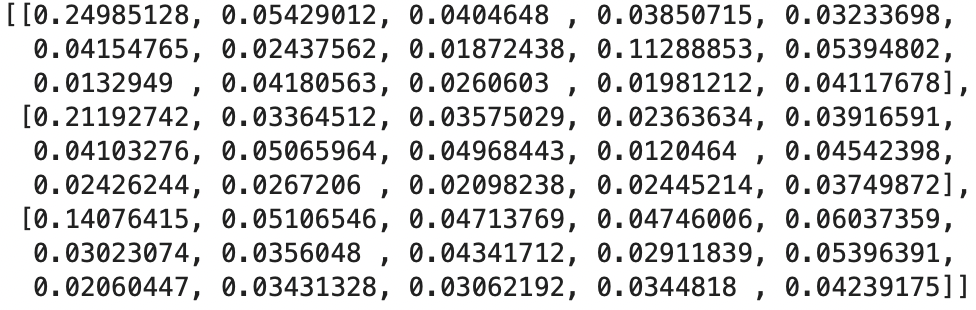
## Method five, the variance of each feature

In this method, I split the training data into three parts based on the labels; training data became three parts because we had three labels.

After that, I calculate the variance of each part for all features. We had 15 features, and because of that, each matrix shape became 15x1.

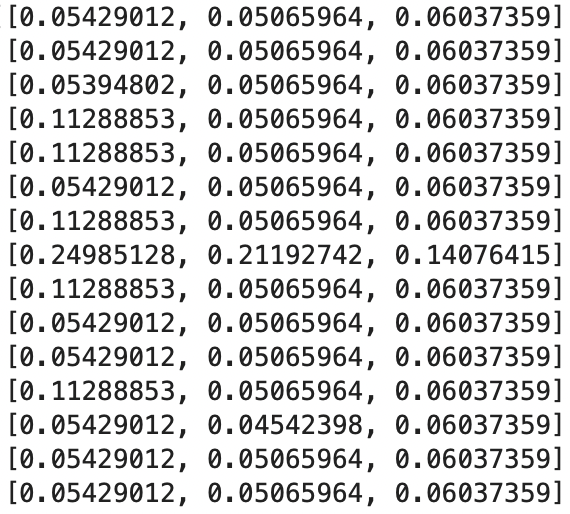
The data has been normalized in this method.

I put the matrixes together to create the relation matrix. The shape of the relation matrix is 3x15.



The Relation matrix of method five

Like the other methods, I used the *composition* method to decompose the relation matrix with the test dataset to create the Similarity matrix.



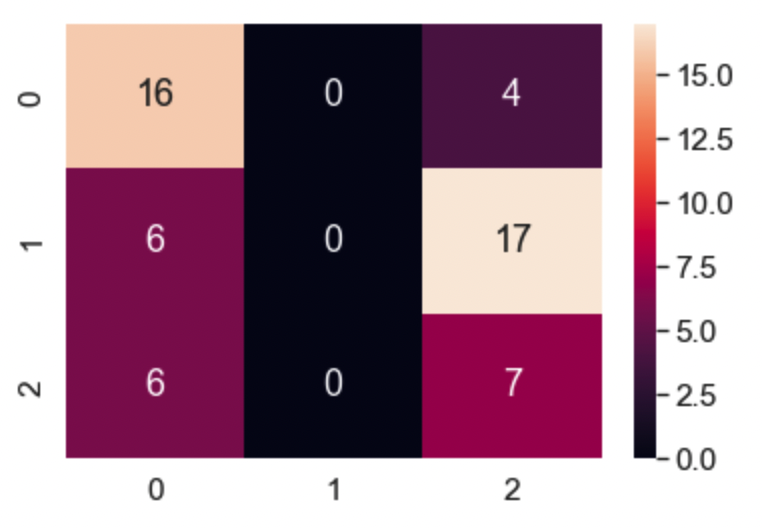
An example of a similarity matrix in method five

After that, the argmax method became handy, and like the other methods, I used it to determine the classes of the test data set.

The results for the three class method are as follow,

Accuracy: 41.07 5

The confusion matrix :

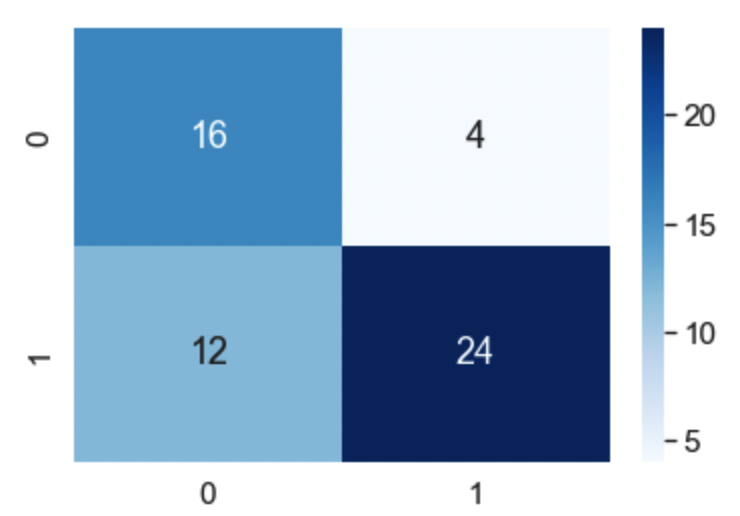


The confusion matrix of three class problems with method five

This method was also applied to two class problem, and the results are as follow :

Accuracy: 71.42%

The confusion matrix is :



The confusion matrix of two class problems with method five