Machine Learning Spring 2022
Assignment 4 Report

Ali Khalid
MSDS21001

June 4, 2022

1 Task # 1

1.1 Show verification performance with the dimensionality reduction. Which dimension is best for classification using the nearest neighbor.

```
Test Set Accuracy:  1.0
Test Set Accuracy:  0.9233333333333333
Test Set Recall Score:  0.9233333333333333
Test Set F1 Score:  0.9237866273243099
```

Figure 1: Verification Scores

1.2 Write your algorithm for speaker verification

- Find mean vector of the data
- subtract mean vector from all data points
- find the covariance matrix of data
- find eigen vectors and values of data
- select number of eigen vectors to be used for transformation by plotting proportion of variation of eigen vectors and selcting the elbow point
- project the data on eigen vectors
- The text time, project the data point using the eigenvectors of training data and find the 'k' nearest neighbours and assign the majority label to test point.
- Find the label for both points using the above step. If the label is same for both data points than we can verify that both data points belong to the same class and vice versa.

1.3 Analyze your results and report precision, recall, accuracy, F1 score, confusion matrix, ROC Curve for each class and for all classes combined ROC curve. Also, report area-Under-the-Curve

```
Test Set Accuracy:  1.0
Test Set Accuracy:  0.9233333333333333
Test Set Recall Score:  0.9233333333333333
Test Set F1 Score:  0.9237866273243099
```

Figure 2: Verification Scores

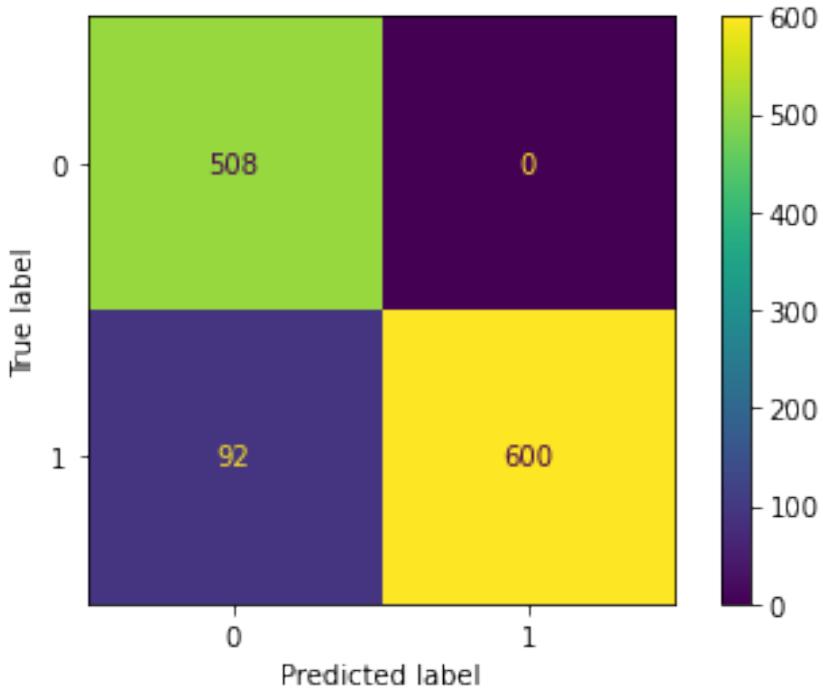

Figure 3: Verification Confusion Matrix

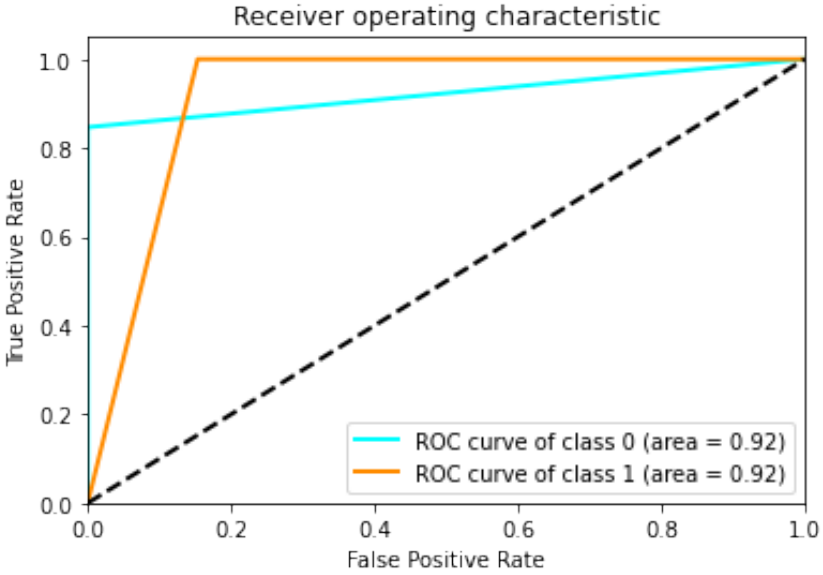

Figure 4: Verification ROC

2 Task # 2

2.1 Write your algorithm for speaker identification

- Split data into test and train set.
- Separate the data points of each class.
- Find the eigen vectors of each class using PCA.

- At test time, project the test data to eigen vectors of all the classes and then reconstruct it. Find the loss between the reconstructed test point and the original test point using root mean square error for each class separately. The label of the class having least error is assigned to the test data.

2.2 Analyze your results and report precision, recall, accuracy, F1 score, confusion matrix, ROC Curve for each class and for all classes combined ROC curve. Also, report area-Under-the-Curve

```
Test Set precison:  0.8949486285124583
Test Set Accuracy:  0.8966666666666666
Test Set Recall Score:  0.8966666666666666
Test Set F1 Score:  0.8948654819768748
```

Figure 5: Verification Scores

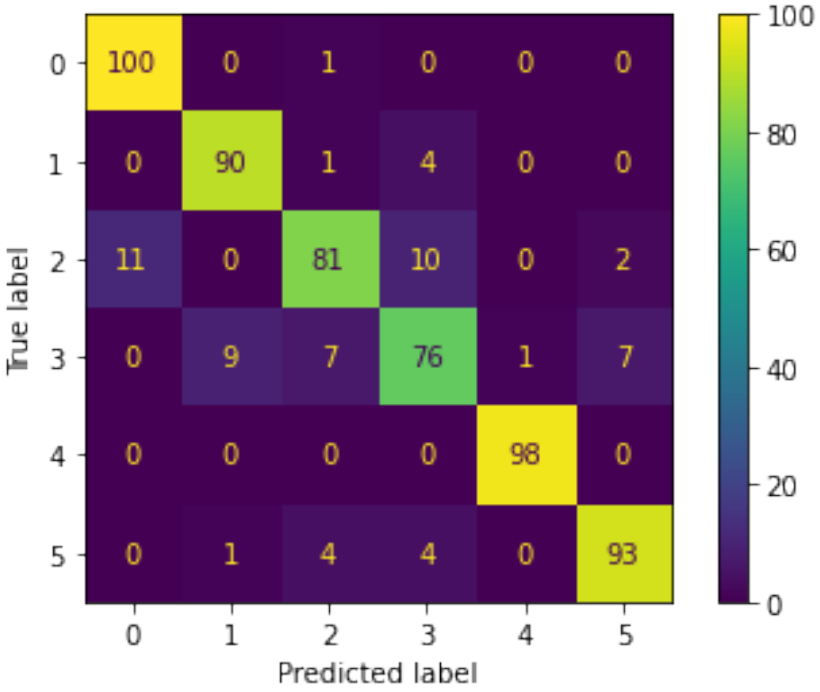

Figure 6: Verification Confusion Matrix

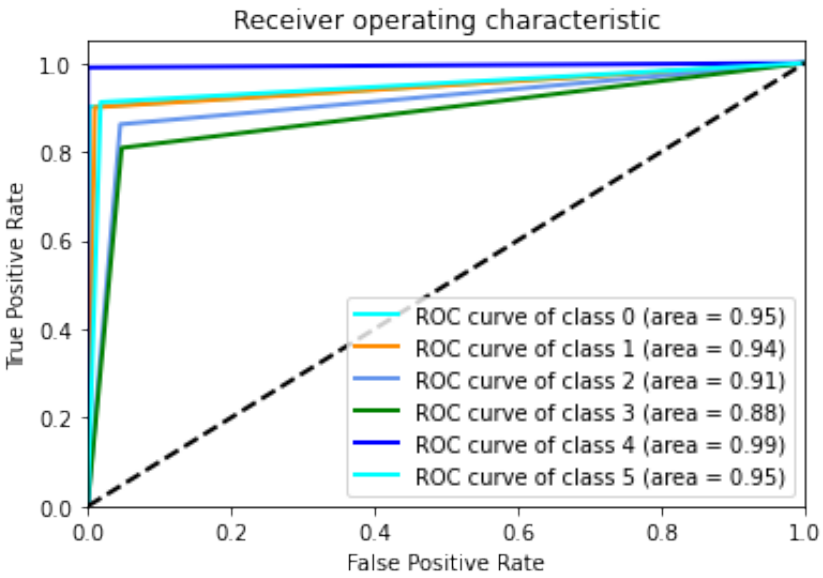

Figure 7: Verification ROC