

CS5691 Pattern Recognition and Machine Learning

Programming Assignment 4

Team 16 - Akshith CS19B005, Shashank CS19B043

May 2, 2022

Feature Extraction: In the handwriting and spoken-digit datasets, training examples are sequences of varying lengths. To extract features of same length, we used the following procedure:

1. Identify the length of the smallest sequence in all of the training data and the development data. This step can be performed during the testing phase, as we have all the training data available in memory during testing. Let this length be x . The objective is to reduce all the training & development sequences to this length x .
2. To reduce the length of a sequence from y to x , choose a window of size $y - x + 1$. Position this window at the leftmost end of the sequence of length y , so that it contains the first $y - x + 1$ elements of it. The average of these elements becomes the first element in the reduced sequence.
3. Move the window one step to the right and re-compute the average to find the next element of the reduced sequence. Repeat this till the window reaches the rightmost end of the longer sequence of length y . The size of the reduced sequence becomes x .

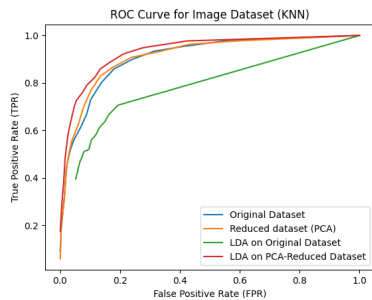
Constant length features extracted from data are then used in all the classification algorithms of the assignment (for Handwriting and Spoken-Digit datasets).

1 K-Nearest Neighbours Classifier

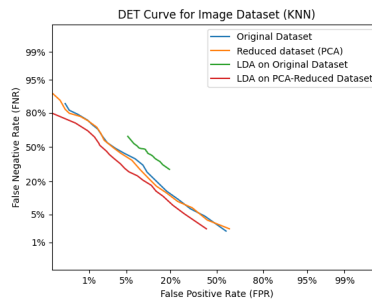
Interpret each training example as a multi-dimensional vector. To classify a test vector, compute the distances of it to each of the training examples. Choose the classes corresponding to the lowest K (empirical) distance training points. The relative frequency of each class in these K classes gives the likelihood of the test point belonging to that class. Classify the test vector into that class having the highest likelihood. These likelihoods can be used to plot the ROC and DET curves with varying thresholds.

1.1 Image Dataset

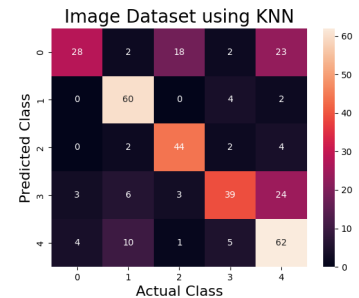
Consider each image as an 828 (36×23) dimensional vector. When KNN is used for classification on the Image dataset directly, as described above, we got an accuracy of 67.8% for $K = 30$. Using PCA to reduce the dimension of the feature vector from 828 to 80, an accuracy of 71.2% was obtained for $K = 30$. When LDA was applied, we got an accuracy of 56.3% on the original dataset and 74.42% on the reduced dataset. Figure 1 shows the plots for the Image Dataset with different combinations of PCA and LDA.



(a) ROC Curve



(b) DET Curve



(c) Confusion matrix for K = 20 (original dataset)

Figure 1: Plots for the Image Dataset

1.2 Synthetic Dataset

Figure 2 shows the scatter plot of the synthetic data assigned to our team. Every training example is a 2

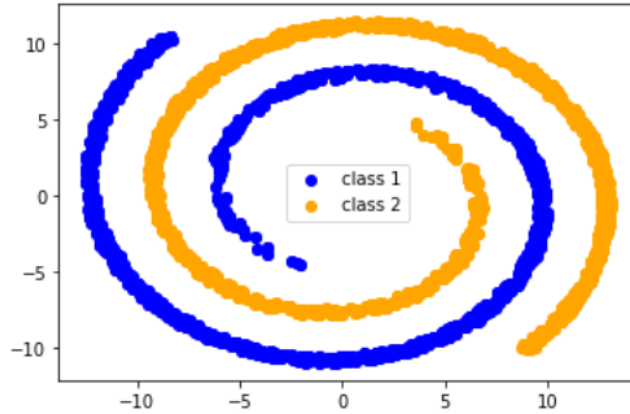


Figure 2: Scatter Plot of Synthetic Dataset

dimensional vector. Similar to the procedure followed in the Image dataset, experiments were carried out with PCA and LDA. When KNN is used on the Synthetic dataset directly, we got an accuracy of 100% for $K = 1, 5, 10$. Using PCA to reduce the dimension of the feature vector from 2 to 1, an accuracy of 72.1% was obtained for $K = 15$. When LDA was applied on both the original and reduced (using PCA) datasets, we got an accuracy of 66.9% and 72.1% respectively. Figure 3 shows the plots for the Synthetic Dataset with different combinations of PCA and LDA.

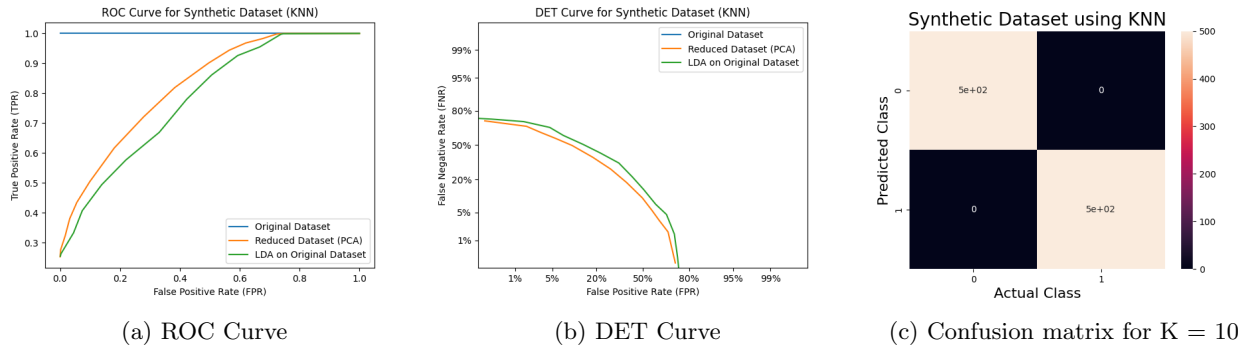


Figure 3: Plots for the Synthetic Dataset

1.3 Handwriting Dataset

KNN was applied on the normalized Handwriting dataset after constant feature extraction was done (explained previously). Similar to previous datasets, different experiments were performed with & without PCA and LDA. When KNN is used on the Handwriting dataset directly, we got an accuracy of 85% for $K = 10$. Using PCA to reduce the dimension of the feature vector to 150, an accuracy of 85.0% was obtained for $K = 10$. When LDA was applied on both the original and reduced (using PCA) datasets, we got an accuracy of 94% on both. Figure 4 shows the plots for the Handwriting Dataset with different combinations of PCA and LDA.

1.4 Spoken-Digit Dataset

KNN was applied on the Spoken-Digit dataset after constant feature extraction was done (explained previously). Similar to previous datasets, different experiments were performed with & without PCA and LDA. When KNN is used on the Handwriting dataset directly, we got an accuracy of 90% for $K = 15$. Using PCA to reduce the dimension of the feature vector to 100, an accuracy of 90% was obtained for $K = 20$. When LDA was applied on both the original and reduced (using PCA) datasets, we got an accuracy of 85% and 93.3% respectively. Figure 5 shows the plots for the Handwriting Dataset with different combinations of PCA and LDA.

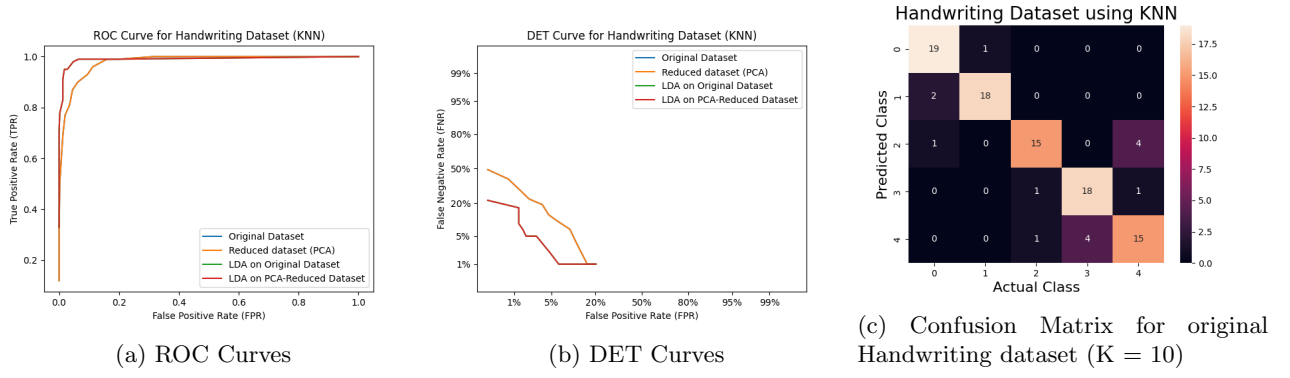


Figure 4: Plots for the Handwriting Dataset

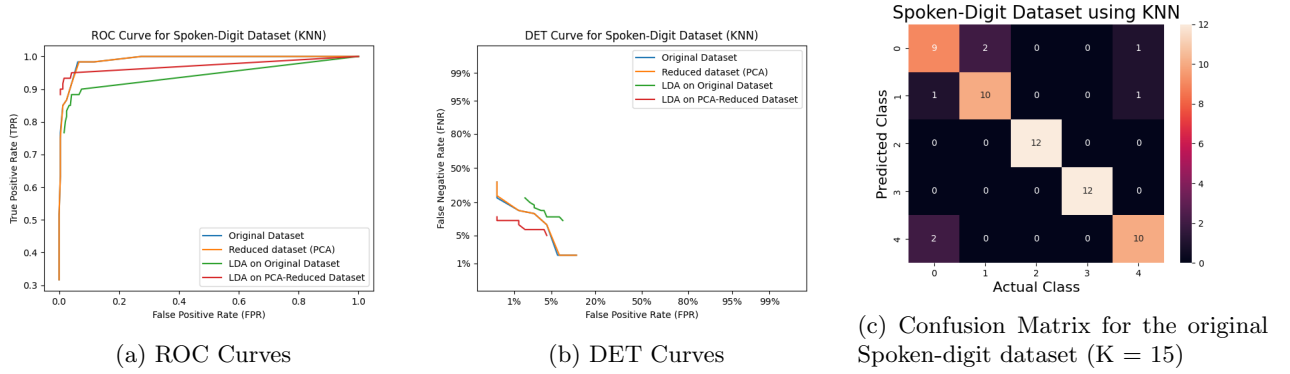


Figure 5: Plots for the Spoken-Digit Dataset

2 Support Vector Machines

Classification using Support Vector Machines was done using the functions available in the `sklearn` library. We tried classification using different kernels: `linear` (ovo and ovr), `poly`, and `rbf`. Among these, `rbf` kernel gave the best performance on all the datasets in terms of accuracy. Using the `rbf` kernel, we tried classification with different regularization parameters. The regularization values corresponding to highest accuracy was chosen in each of the plots for all the datasets.

2.1 Image Dataset

SVM was applied on the Image dataset. Different experiments were performed with & without PCA and LDA. Using SVM, We got an accuracy of 76.7% for $C = 10$.

Use PCA to reduce the dimension of the feature vector from 828 to 50. An accuracy of 77.8% is obtained for $C = 10$.

When LDA was applied on both the original and reduced (using PCA) datasets, we got an accuracy of 59.2% and 72.7% respectively. Figure 6 shows the plots with these combinations.

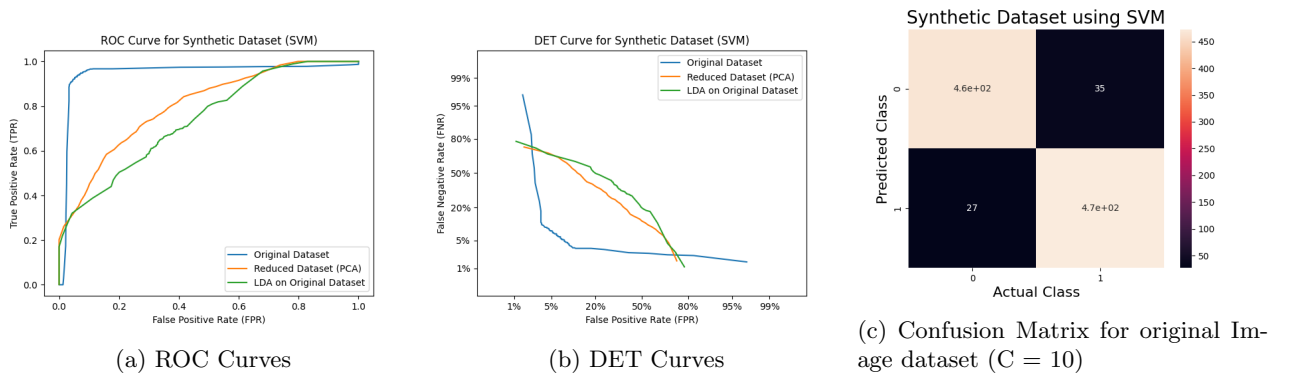


Figure 6: Plots for the Image Dataset

2.2 Synthetic Dataset

SVM was applied on the Synthetic dataset. Different experiments were performed with & without PCA and LDA.

Using SVM, We got an accuracy of 93.8% for $C = 2$.

Use PCA to reduce the dimension of the feature vector from 2 to 1. An accuracy of 72.1% is obtained for $C = 2$.

Using LDA, we obtained an accuracy of 65.4% on both original and reduced datasets. Figure 7 shows the plots with these combinations.

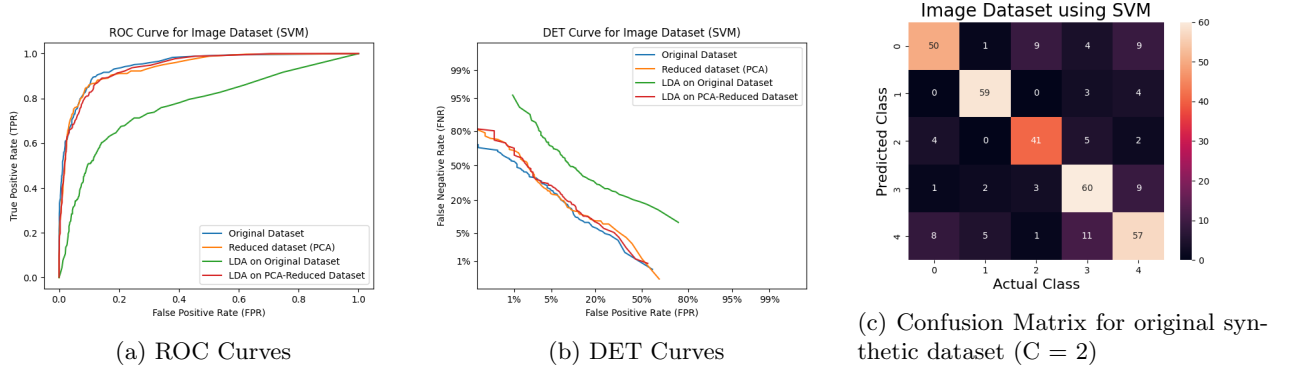


Figure 7: Plots for the Synthetic Dataset

2.3 Handwriting Dataset

SVM was applied on the Handwriting dataset. Different experiments were performed with & without PCA and LDA.

Using SVM, We got an accuracy of 92% for $C = 10$.

Use PCA to reduce the dimension of the feature vector to 100. An accuracy of 89% is obtained for $C = 10$.

Using LDA, we obtained an accuracy of 93% on both reduced and original datasets. Figure 8 shows the plots with these combinations.

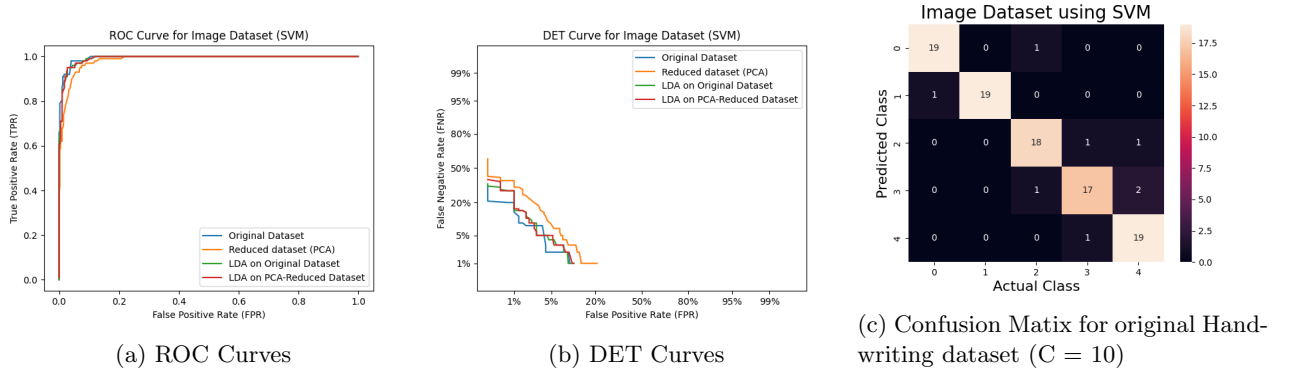


Figure 8: Plots for the Handwriting Dataset

2.4 Spoken-Digit Dataset

SVM was applied on the Spoken-Digit dataset. Different experiments were performed with & without PCA and LDA.

Using SVM, We got an accuracy of 96.6% for $C = 10$.

Use PCA to reduce the dimension of the feature vector to 100. An accuracy of 96.67% is obtained for $C = 10$.

Using LDA, we obtained an accuracy of 80% on the original dataset and 91.67% on the reduced dataset. Figure 9 shows the plots with these combinations.

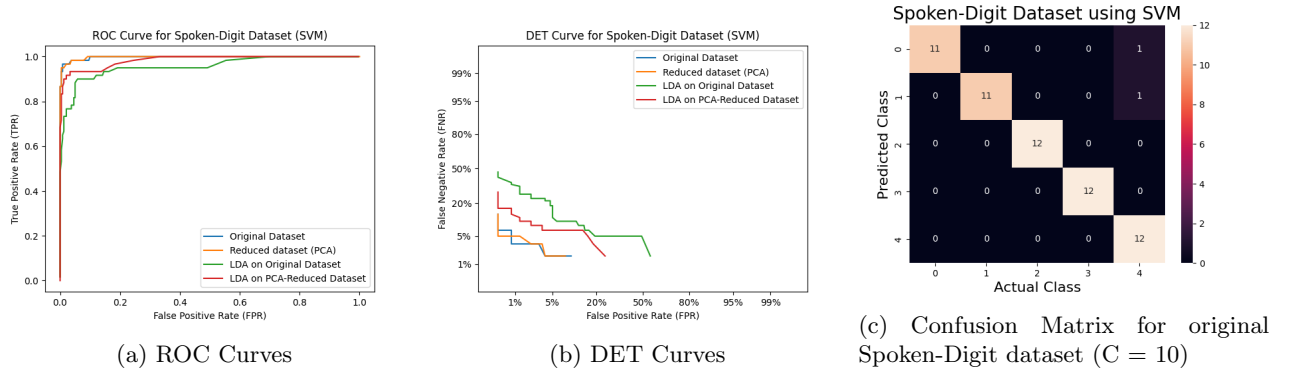


Figure 9: Plots for the Spoken-Digit Dataset

3 Logistic Regression

Classification using Logistic Regression is done with linear and second order basis functions. These basis functions create non-linear boundary surfaces. Gradient descent is used with varying number of iterations. We used softmax activation function to estimate likelihoods and used them to plot ROC and DET curves.

3.1 Image Dataset

Logistic Regression was applied on the Image dataset. Different experiments were performed with & without PCA and LDA.

Using logistic regression without LDA or PCA, we got an accuracy of 55.7% .

Using PCA to reduce the dimension of the feature vector to 60, an accuracy of 72.4% is obtained.

Using LDA to reduce the dimension of feature vector to 4, an accuracy of 56% is obtained.

Using PCA followed by LDA, an accuracy of 75.57% is obtained.

Figure 10 shows the plots with these combinations.

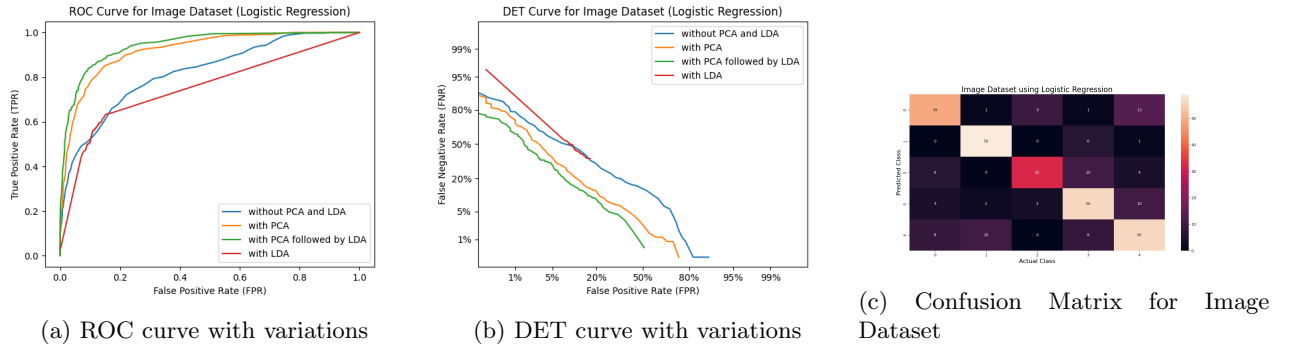


Figure 10: Plots for the Image Dataset

3.2 Synthetic Dataset

While applying Logistic Regression on the Synthetic dataset, we performed different experiments with & without PCA and LDA.

Using logistic regression without LDA and PCA, we got an accuracy of 62.8%

Use PCA to reduce the dimension of the feature vector from 2 to 1, an accuracy of 63.9% was obtained.

Using LDA to reduce the dimension of feature vector to 1, an accuracy of 65.9% is obtained.

Figure 11 shows the plots with these combinations.

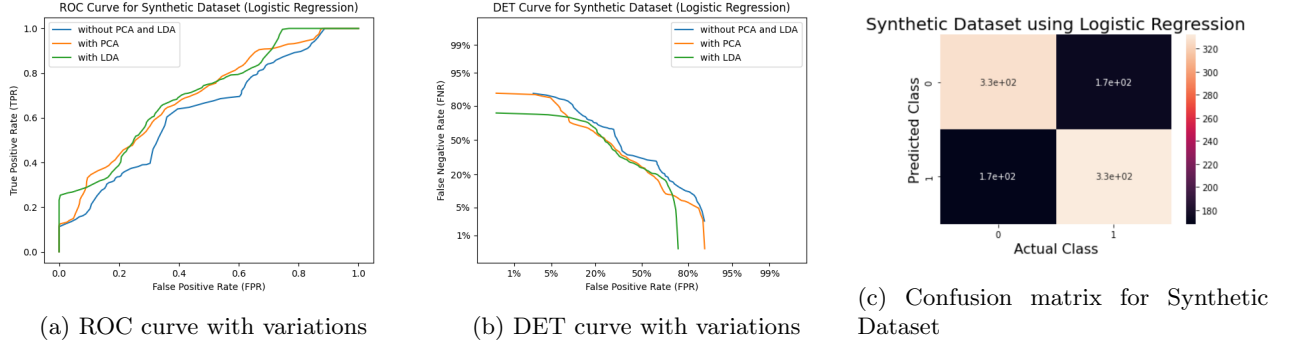


Figure 11: Plots for the Synthetic Dataset

3.3 Handwriting Dataset

Logistic Regression was applied on the Handwritten dataset. Different experiments were performed with & without PCA and LDA.

Using logistic regression without LDA and PCA, We got an accuracy of 92% .

Use PCA to reduce the dimension of the feature vector to 10, An accuracy of 91% is obtained

Using LDA to reduce the dimension of feature vector to 4, an accuracy of 91% is obtained.

Using PCA followed by LDA, an accuracy of 89% is obtained

Figure 12 shows the plots with these combinations

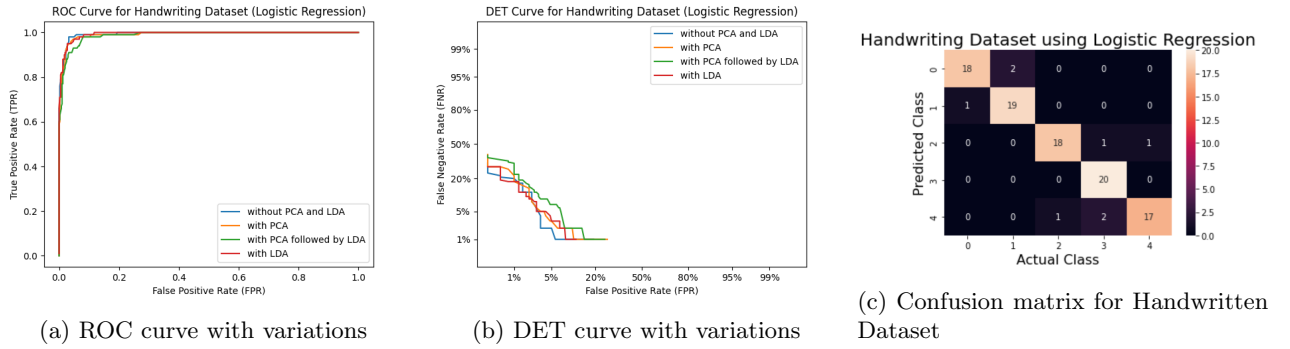


Figure 12: Plots for the Handwriting Dataset

3.4 Spoken-Digit Dataset

Logistic Regression was applied on the Spoken-Digit dataset. Different experiments were performed with & without PCA and LDA.

Using logistic regression without LDA and PCA, We got an accuracy of 91.66% .

Use PCA to reduce the dimension of the feature vector to 50, An accuracy of 93.33% is obtained

Using LDA to reduce the dimension of feature vector to 4, an accuracy of 85% is obtained.

Using PCA followed by LDA, an accuracy of 98.33% is obtained

Figure 13 shows the plots with these combinations

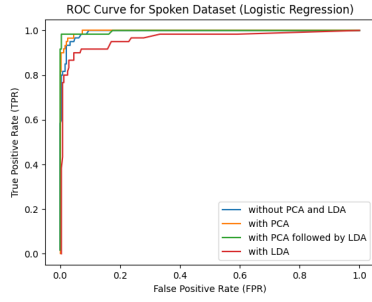
4 Artificial Neural Network (ANN)

To build a Multilayer Perceptron or ANN Classifier, we used `MLPClassifier` available in the `neural_network` module of the `sklearn` library. We experimented with different no. of hidden layer sizes, activation functions, and maximum iterations. For each dataset, parameters that gave the best accuracy on the development data were selected to build the classifier. Exact values of these parameters can be found the code submitted.

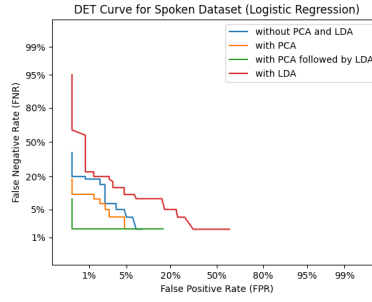
4.1 Image Dataset

Using ANN on the Image dataset, we performed experiments with & without PCA and LDA.

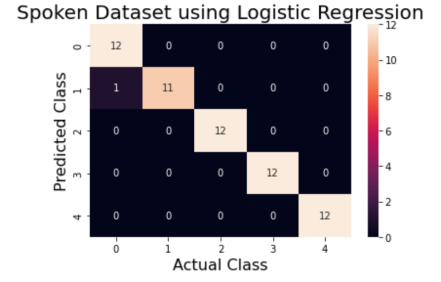
On the original Image dataset, an accuracy of 74.42% using `random_state = 10` and `max_iter = 1000`.



(a) ROC with variations



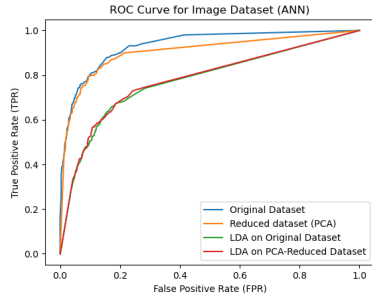
(b) DET with variations



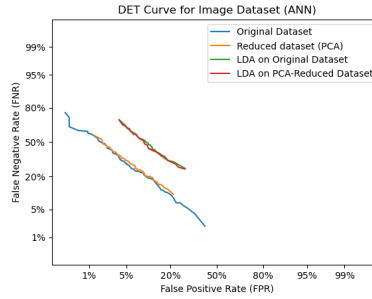
(c) Confusion Matrix for Spoken-Digit Dataset

Figure 13: Plots for the Spoken-Digit Dataset

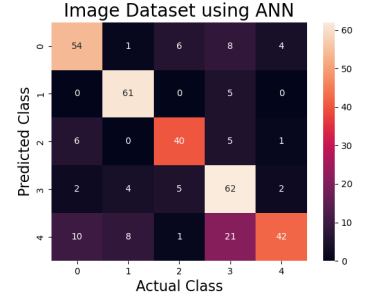
We used PCA to reduce the dimension of the feature vectors from 828 to 100. An accuracy of 72.7% is obtained for the same parameter values as above. Using LDA, we obtained an accuracy of 54% on the original dataset and 56.6% on the reduced dataset. Figure 14 shows the plots with these combinations.



(a) ROC Curves



(b) DET Curves

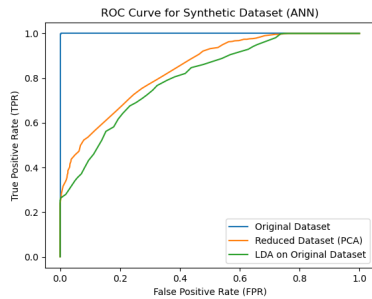


(c) Confusion Matrix for original Image dataset

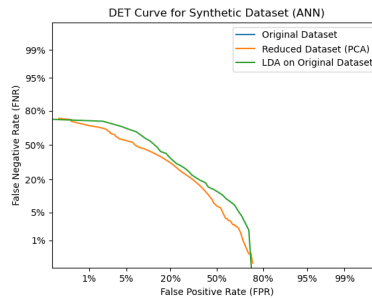
Figure 14: Plots for the Image Dataset

4.2 Synthetic Dataset

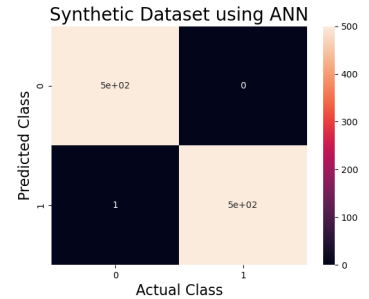
Using ANN on the Synthetic dataset, we performed experiments with & without PCA and LDA. On the original Synthetic dataset, an accuracy of 99.9% using `random_state = 3` and `max_iter = 1000`. We used PCA to reduce the dimension of the feature vectors from 2 to 1. An accuracy of 74% is obtained for the same parameter values as above. Using LDA, we obtained an accuracy of 71.89% on both the original and reduced datasets. Figure 15 shows the plots with these combinations.



(a) ROC Curves



(b) DET Curves



(c) Confusion Matrix for original Synthetic dataset

Figure 15: Plots for the Synthetic Dataset

4.3 Handwriting Dataset

Using ANN on the Handwriting dataset, we performed experiments with & without PCA and LDA. On the original Handwriting dataset, an accuracy of 95% using `random.state = 10` and `max.iter = 1000`. We used PCA to reduce the dimension of the feature vectors to 100. An accuracy of 94% is obtained for the same parameter values as above. Using LDA, we obtained an accuracy of 90% on both the original and reduced datasets. Figure 16 shows the plots with these combinations.

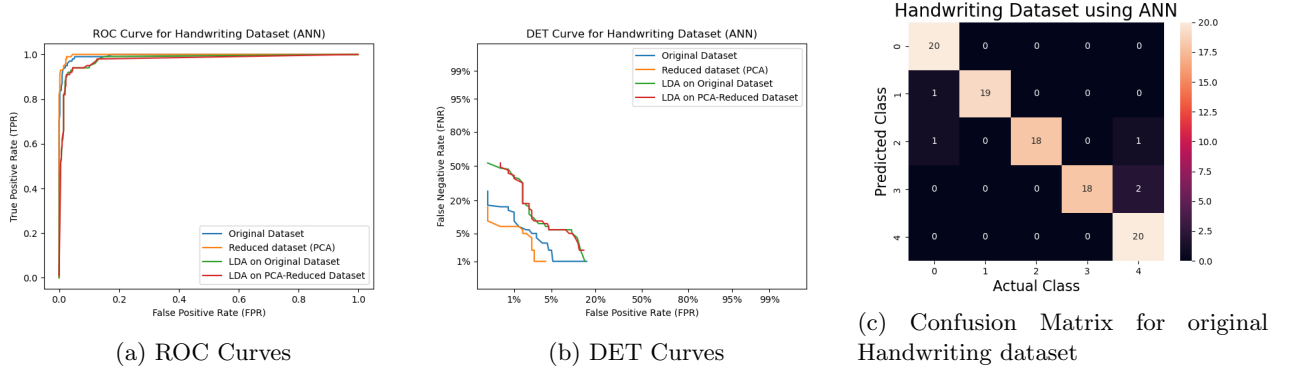


Figure 16: Plots for the Handwriting Dataset

4.4 Spoken-Digit Dataset

Using ANN on the Spoken-Digit dataset, we performed experiments with & without PCA and LDA. On the original Spoken-Digit dataset, an accuracy of 93.33% using `random.state = 10` and `max.iter = 1000`. We used PCA to reduce the dimension of the feature vectors to 100. An accuracy of 93.33% is obtained for the same parameter values as above. Using LDA, we obtained an accuracy of 85% on both the original and reduced datasets%. Figure 17 shows the plots with these combinations.

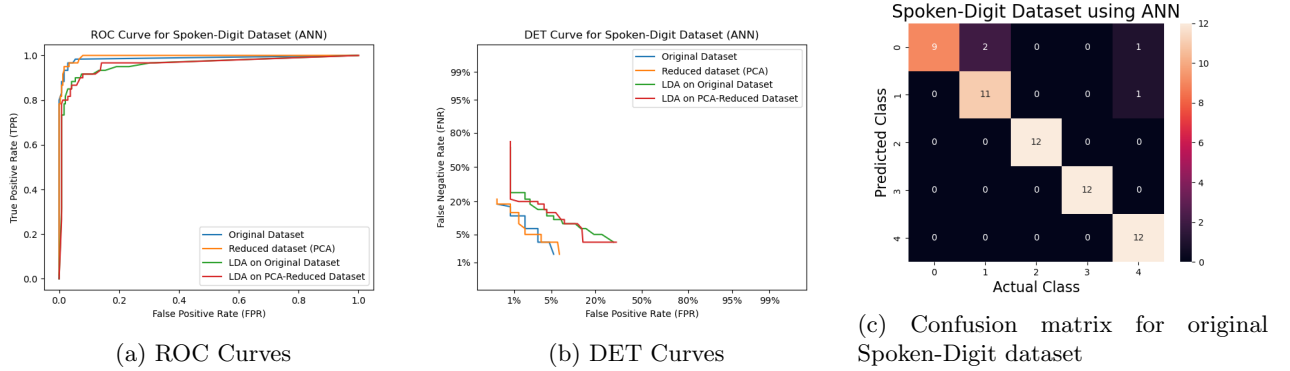


Figure 17: Plots for the Spoken-Digit Dataset