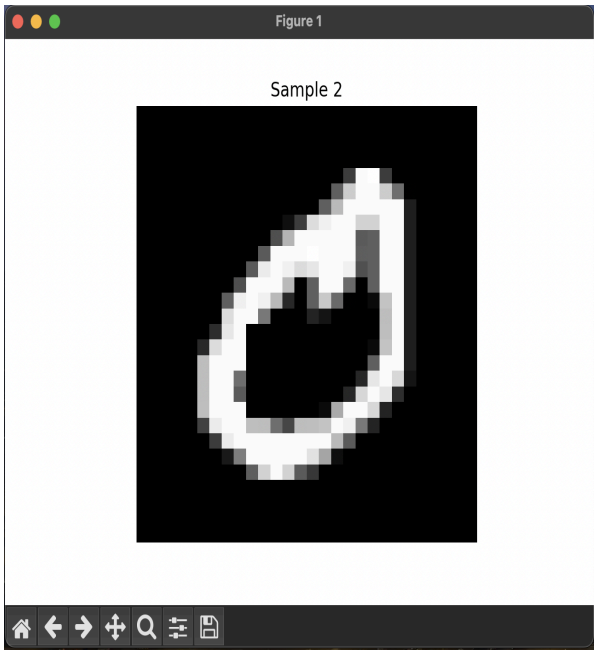
SML Assignment 2
Report

Aditya Prasad
2022036

Question 1)

Output:




```
adityaprasad@Adityas-MacBook-Pro-2 ~ % /usr/
Class 0
2024-02-21 23:23:16.520 Python[21491:1084825
torableState: and returning YES.

Class 1

Class 2

Class 3

Class 4

Class 5

Class 6

Class 7

Class 8

Class 9

Accuracy: 0.7808
adityaprasad@Adityas-MacBook-Pro-2 ~ %
```

This code is an implementation of a Gaussian Quadratic Discriminant Analysis (QDA) classifier for the MNIST dataset. Here's a breakdown:

Assumptions:
1. Assumes that the MNIST dataset is loaded with training and testing images (`x_train`, `x_test`) along with their corresponding labels (`y_train`, `y_test`).
2. Assumes that each class follows a Gaussian distribution with a separate covariance matrix for each class.

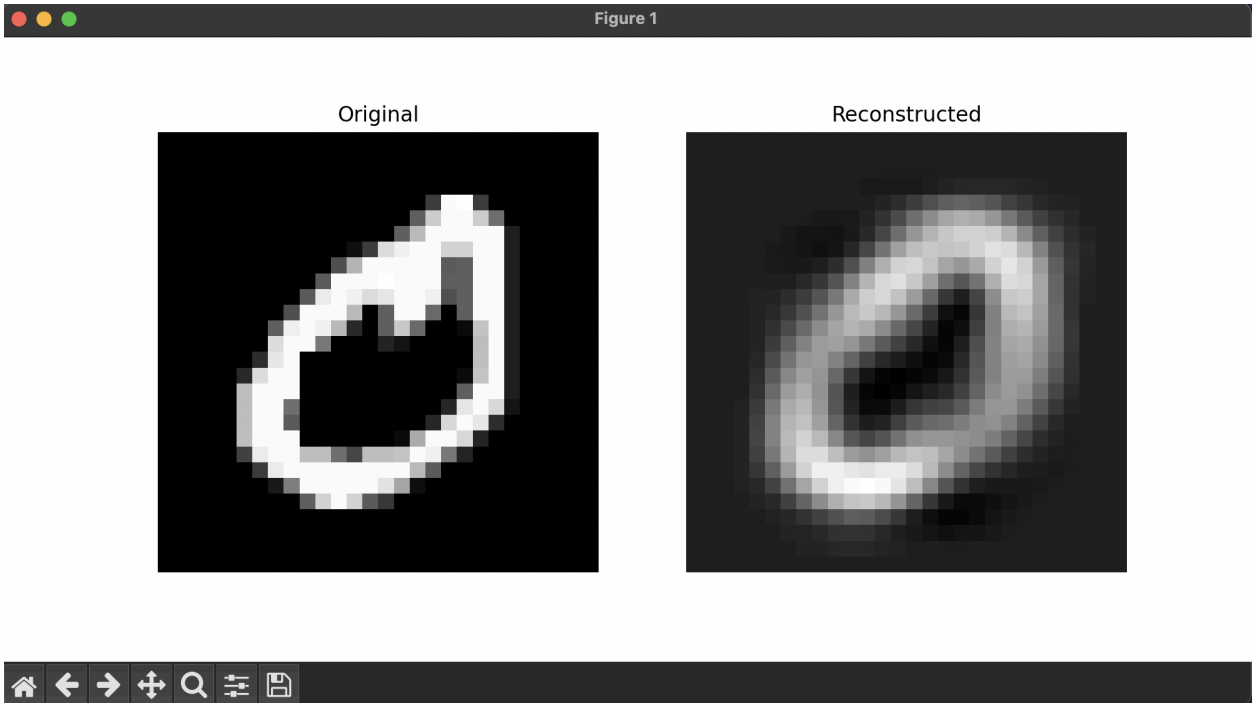
Approach:
1. Data Loading: Load the MNIST dataset.
2. Display Samples: Define a function to display sample images.
3. Calculate Accuracy: Define a function to calculate accuracy.
4. Collect Samples: Collect 5 samples for each of the 10 classes from the training dataset.
5. Training: Calculate the mean and covariance matrix for each class. Regularize the covariance matrix to avoid singularities.
6. Testing: For each test sample, calculate the quadratic forms using the QDA discriminant function.
7. Prediction: Assign the class label corresponding to the maximum quadratic form value.
8. Accuracy Calculation: Calculate the accuracy of the classifier on the test dataset.

Explanation in Short:
The code implements a QDA classifier for the MNIST dataset by first training on the training data and then testing on the test data. It assumes Gaussian distribution for each class and calculates mean and covariance matrices for each class during training. During testing, it calculates quadratic forms using these parameters and assigns the class label corresponding to the maximum quadratic form value.

Results:
The code outputs the accuracy of the classifier on the test dataset.
The accuracy of the classifier on the test dataset is printed as the final output. This accuracy indicates how well the QDA model performs in classifying handwritten digits from the MNIST dataset.

Question 2)

Output:




```
adityaprasad@Adityas-MacBook-Pro-2 ~ % /u
MSE for p=5: 5212.671600878399
Accuracy for p=5: 0.7304
2024-02-21 23:33:25.075 Python[21660:1092
torableState: and returning YES.
MSE for p=10: 5800.342064124673
Accuracy for p=10: 0.8927
MSE for p=20: 6310.612026098812
Accuracy for p=20: 0.9507
adityaprasad@Adityas-MacBook-Pro-2 ~ %
```

This code performs dimensionality reduction using Principal Component Analysis (PCA) on the MNIST dataset, reconstructs the images from the reduced dimensions, and then applies Quadratic Discriminant Analysis (QDA) for classification, evaluating the accuracy of classification for different values of reduced dimensions.

Assumptions:
1. The MNIST dataset is available and properly loaded.
2. PCA is used for dimensionality reduction.
3. QDA is a suitable classification technique for the reduced data.
4. Accuracy is a valid metric for evaluating classification performance.

Approach:
1. The code begins by loading the MNIST dataset and reshaping the training and testing data.
2. PCA is performed on the training data to compute eigenvectors and eigenvalues of the covariance matrix.
3. The training data is projected onto the principal components to obtain a lower-dimensional representation.
4. For each specified value of reduced dimensions (p), the code reconstructs images from the reduced dimensions and calculates the Mean Squared Error (MSE) between original and reconstructed images.
5. Mean and covariance are calculated for each class using the reduced training data.
6. QDA is applied to the reduced testing data to classify digits.
7. Accuracy is calculated for each value of p.
8. Original and reconstructed images are displayed for visual inspection.

Results:
1. - The Mean Squared Error (MSE) indicates how well the reconstruction captures the original images. Lower MSE suggests better reconstruction.
2. - The accuracy of classification using QDA is calculated for each value of p, providing insight into the effectiveness of dimensionality reduction for classification.
3. - Reconstructed images are displayed to visually inspect the quality of reconstruction.

Overall, the code provides a comprehensive analysis of dimensionality reduction using PCA and its impact on classification accuracy using QDA on the MNIST dataset.