CSE 555: Pattern Recognition (Spring 2019)

Instructor: Dr. Wen Dong

Programming Assignment 1: Bayesian Decision Theory

March 1, 2019

Report By:

Siddheswar Chandrasekhar

Objective

In this problem we will apply discriminant analysis to recognize the digits in the MNIST data set. We will train our model using the training data sets and test the performance using the test data set.

The images are 28 x 28 pixels in gray-scale and the categories are 0, 1, ... 9. We concatenate the image rows into a 28 x 28 vector and treat this as our feature, and assume the feature vectors in each category in the training data have Gaussian distribution.
We then draw the mean and standard deviation of those features for the 10 categories as 28 x 28 images using the training images. There are two images for each of the 10 digits, one for mean and one for standard deviation. We call those "mean digits" and "standard deviation digits".

We then classify the images in the testing data set using 0-1 loss function and Bayesian decision rule and report the performance.

Plan of Action

1. We read the MNIST dataset into *x_train*, *y_train*, *x_test* and *y_test*

2. We categorize each image in the *x_train* on the basis of their labels i.e. *y_train* and find each category's mean and standard deviation.

3. We plot these mean and standard deviation images

4. We then implement a quadratic discriminant function and pass it *x_test* and check our prediction against *y_test* and measure the accuracy of our prediction.

MNIST Dataset

The MNIST database (Modified National Institute of Standards and Technology database) is a large database of handwritten digits that is commonly used for training various image processing systems. The database is also widely used for training and testing in the field of machine learning. It was created by "re-mixing" the samples from NIST's original datasets. The creators felt that since NIST's training dataset was taken from American Census Bureau employees, while the testing dataset was taken from American high school students, it was not well-suited for machine learning experiments. Furthermore, the black and white images from NIST were normalized to fit into a 28x28 pixel bounding box and anti-aliased, which introduced grayscale levels. [2]

The MNIST database contains 60,000 training images and 10,000 testing images. Half of the training set and half of the test set were taken from NIST's training dataset, while the other half of the training set and the other half of the test set were taken from NIST's testing dataset. There have been a number of scientific papers on attempts to achieve the lowest error rate; one paper, using a hierarchical system of convolutional neural networks, manages to get an error rate on the MNIST database of 0.23%. The original creators of the database keep a list of some of the methods tested on it. In their original paper, they use a support vector machine to get an error rate of 0.8%. An extended dataset similar to MNIST called EMNIST has been published in 2017, which contains 240,000 training images, and 40,000 testing images of handwritten digits and characters. [2]

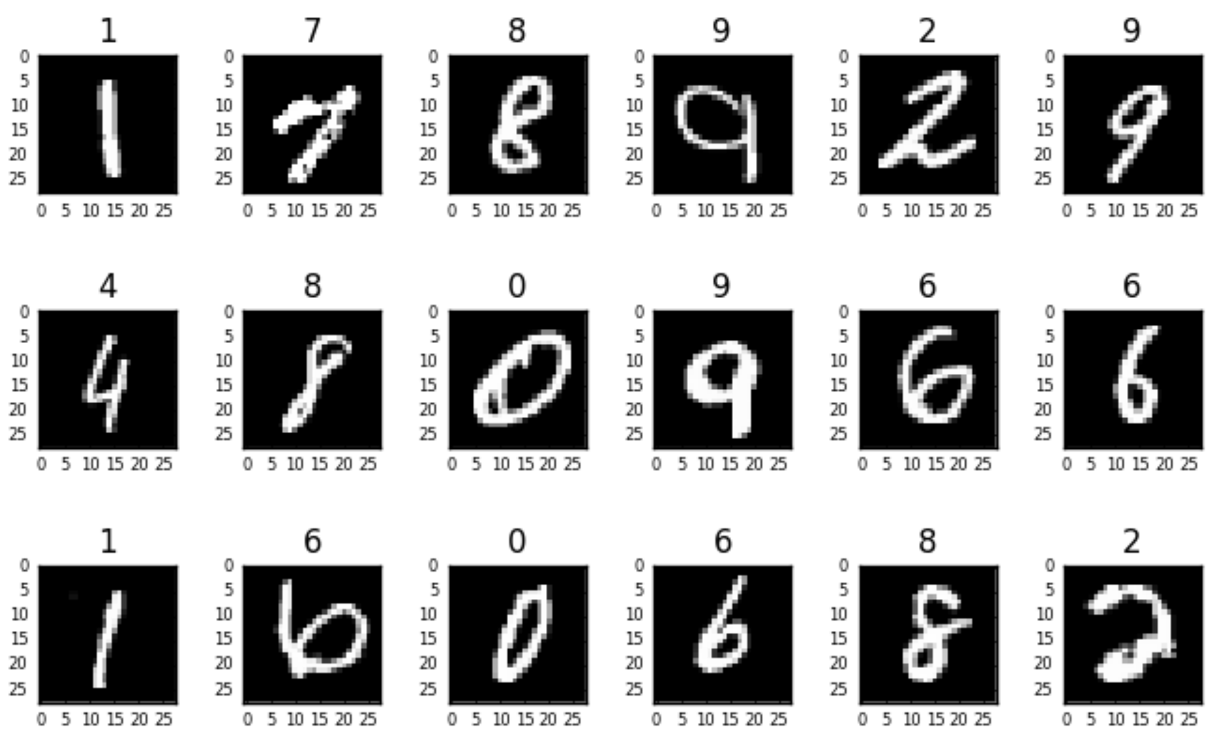

Fig 1 – MNIST sample images

Discriminant Analysis

Linear discriminant analysis (LDA), normal discriminant analysis (NDA), or discriminant function analysis is a generalization of Fisher's linear discriminant, a method used in statistics, pattern recognition and machine learning to find a linear combination of features that characterizes or separates two or more classes of objects or events. [1]

Discriminant analysis is used when groups are known a priori. Each case must have a score on one or more quantitative predictor measures, and a score on a group measure.
In simple terms, discriminant function analysis is classification - the act of distributing things into groups, classes or categories of the same type. [1]

We apply Quadratic Discriminant Analysis for our task, which has the form: [3]

$$g_i(x) = x^t W_i x + N_i^t x + B_{i0},$$

where $W_i = -\frac{1}{2} \sum_i^{-1}$, $N_i = \sum_i^{-1} \mu_i$ and $B_{i0} = -\frac{1}{2}\mu_i^t \sum_i^{-1} \mu_i + lnP(\omega_i) - \frac{1}{2}ln| \sum_i |$ and quadratic boundary.

Observations

Part 1

After categorizing each image in the *x_train* on the basis of their labels i.e. *y_train*, we find the mean image and standard deviation image of each category.

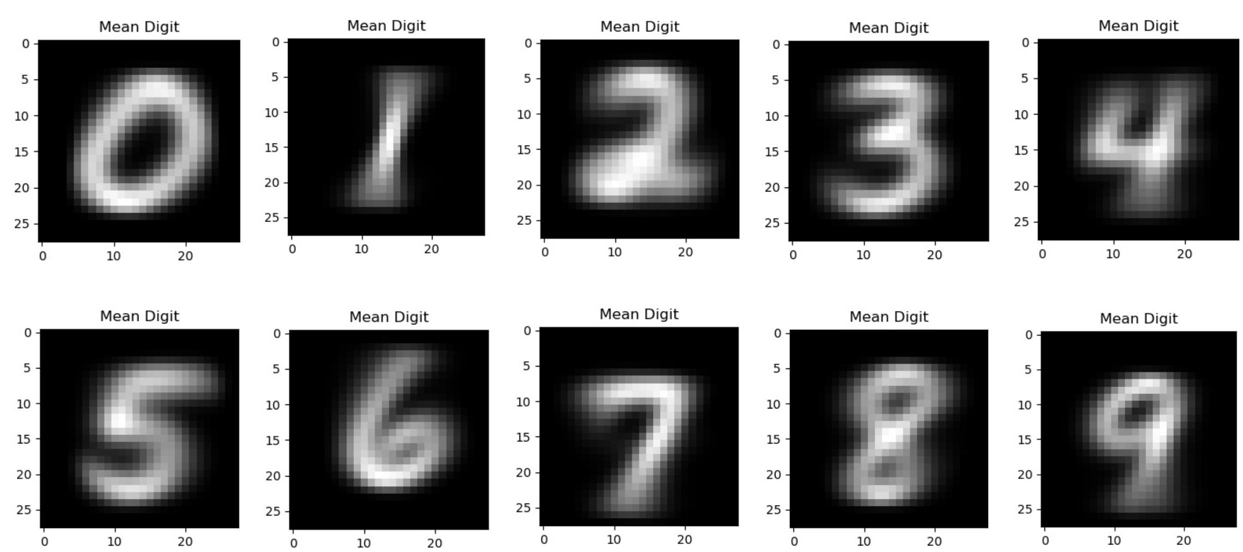

Fig 2 – Mean Images

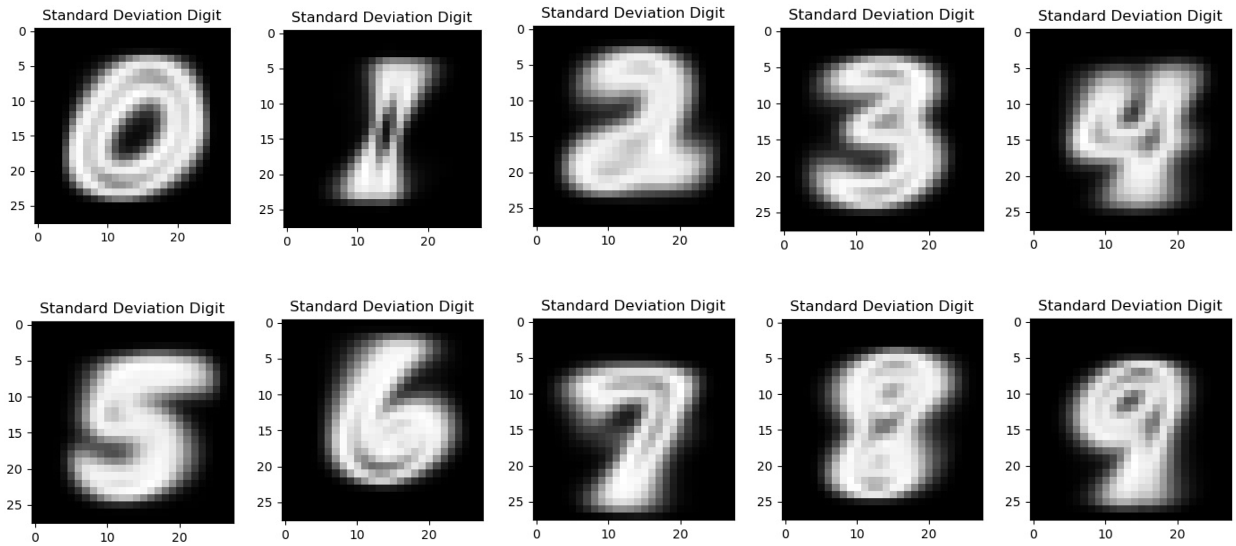

Fig 3 – Standard Deviation Images

Part 2

Our quadratic discriminant function, to which we pass *x_test* and check the prediction against *y_test*, marks an accuracy of 83.42%

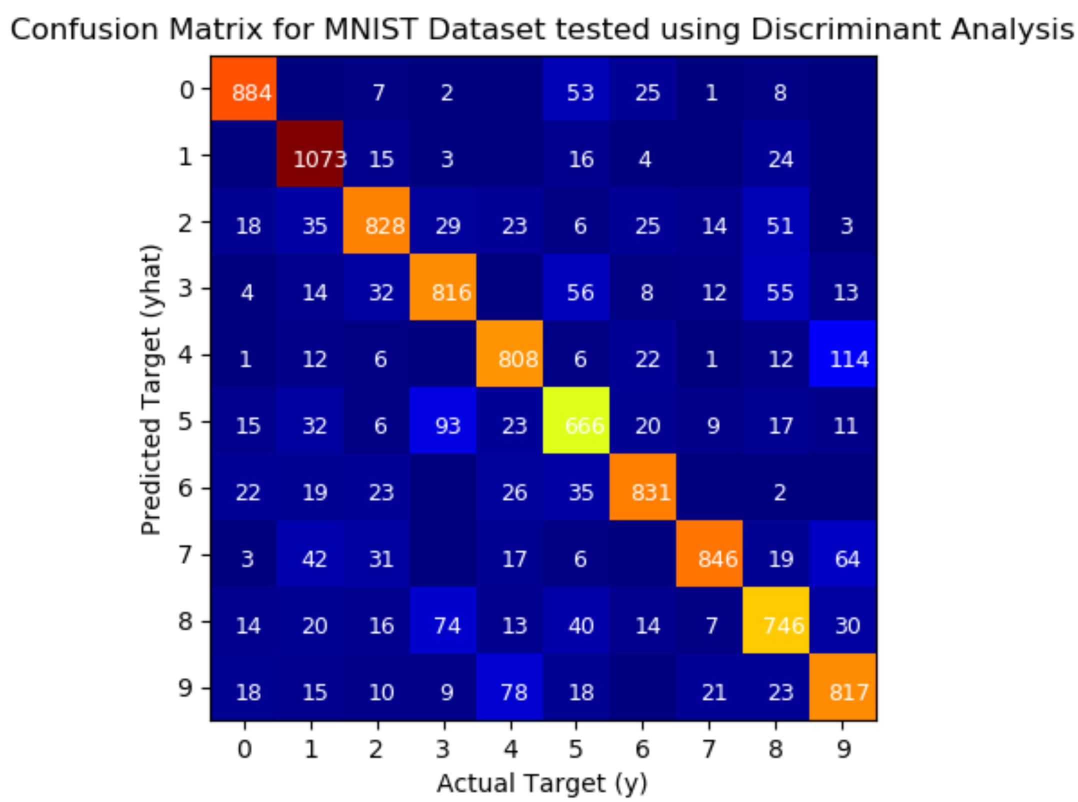

Fig 4 – Confusion Matrix

Inferences

From the above observations, we can make the following infers:

- When tested using Discriminant Analysis, we observe a higher error rate than other methods. This is because Discriminant Analysis is not a good measure of testing such datasets.

- Observing the confusion matrix, we notice some patterns of misclassification. For example, we observe that *'9'* is incorrectly predicted as *'4'* 114 times and *'4'* is incorrectly predicted as *'9'* 78 times. Also, *'8'* is incorrectly predicted as *'3'* 74 times and *'3'* is incorrectly predicted as *'5'* 93 times.
  This is because the digits, *9* and *4*, *3* and *8*, and in some cases *3* and *5* look very similar. And hence, their mean and standard deviation images also look very similar and hence is highly likely to be misclassified incorrectly.
  The main reason for this is that Discriminant Analysis makes prediction on the basis of similarity to the mean and standard deviation images and not individual features.

References

1. Linear Discriminant Analysis
   https://en.wikipedia.org/wiki/Linear_discriminant_analysis

2. MNIST Database
   https://en.wikipedia.org/wiki/MNIST_database

3. Professor Wen Don's Lecture notes on Bayesian Decision Theory