IPR CSE 555 Problem Set 1: Discriminant Analysis

Aniruddha D. Karlekar
Department of Computer Science
University at Buffalo
Buffalo, NY 14214
akarleka@buffalo.edu

Abstract

In this problem we will apply discriminant analysis to recognize the digits in the MNIST data set (http://yann.lecun.com/exdb/mnist/). As a bonus problem we will construct "Fisher digits". We will train our model using the training data sets ("train-images-idx3-ubyte.gz" and "train-labels-idx1-ubyte.gz") and test the performance using the test data set ("t10k-images-idx3-ubyte.gz" and "t10k-labels-idx1-ubyte.gz").

1 Task 1

The images are 28 x 28 pixels in gray-scale. 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 "train-images-idx3-ubyte.gz") have Gaussian distribution. Draw the mean and standard deivation of those features for the 10 categories as 28 x 28 images using the training images ("train-images-idx3-ubyte.gz"). There should be 2 images for each of the 10 digits, one for mean and one for standard deviation. We call those "mean digits" and "standard deviation digits" in CSE455/555.

Method:

1. Reading and Loading the dataset.
2. Calculate the mean of all pixels for each unique digit. A mean gives us the average of all values

$$\bar{x} = \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right) = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

3. Calculate the Standard Deviation of all pixels for each unique digit. The Standard Deviation gives us the dispersion of values from the mean value.

$$s = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2},$$

2 Task 2

Classify the images in the testing data set ("t10k-images-idx3-ubyte.gz") using 0-1 loss function and Bayesian decision rule and report the performance. Why it doesn't perform as good as many other methods on LeCun's web page? Before coding the discriminant functions, review Section 2.6.

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.

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. For our task, we have used quadratic discriminant analysis that has the form:

|

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

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.

Quadratic discriminant analysis (QDA) is closely related to linear discriminant analysis (LDA), where it is assumed that the measurements from each class are normally distributed. Unlike LDA however, in QDA there is no assumption that the covariance of each of the classes is identical. When the normality assumption is true, the best possible test for the hypothesis that a given measurement is from a given class is the likelihood ratio test.

Result:

After successfully implementing the Quadratic discriminant function, we were able to classify the MNIST dataset with an accuracy of 82.49%.

3 Conclusion

Why it doesn't perform as good as many other methods on LeCun's web page?

0-1 loss function and Bayesian decision rule doesn't perform as good as many other methods on LeCun's web page because:

- The naïve Bayes classifier makes the assumption that all features are independendent given the output class. This can cause the result to have a poor performance.
- For any possible feature, we need to estimate a likelihood value based on the frequency of occurance. This may result in probabilities that tend towards 0 or 1, rendering the classifier ineffective.
- As it is fed increasing quantities of training data, the performance of the Naive Bayes classifier plateaus above a certain threshold. Thus at one point it does not lead to significant benefits when we feed the classifier more training data.

References

[1] https://en.wikipedia.org/

[2]Professor Wen Dong’s class notes and ppt