Name: **Hend Aly**
ASU ID: **1229580406**

Naive Bayes Classifier Project Report

Introduction

The goal of this project was to build a Naïve Bayes classifier to classify images of the handwritten digits "0" and "1" from the MNIST dataset. The project included several tasks, extracting features, calculating the parameters (Mean and Variance) for each feature, classification, and evaluating accuracy.

Methodology

Data Preparation

The MNIST dataset contains 70,000 images of handwritten digits, divided into 60,000 training images and 10,000 testing images. We used in this project datasets that were subsets of the MNIST database that focused on images of the digits "0" and "1". We used the training sets to develop the model, while the testing sets were used to evaluate the performance.

Feature Extraction

For each image, we extracted two features:

1. Mean Brightness: This feature represents an average of all pixel brightness values within a whole image array.

2. Standard Deviation of Brightness: This feature represents the standard deviation of all pixel brightness values within a whole image array.

Parameter Calculation

In this task, we used the training data to calculate the mean and variance for each feature. We used these parameters to model the probability distributions of the features for each class, assuming a normal (Gaussian) distribution.

```
Parameters for digit 0 (class 0):
Mean of Feature 1: 44.12499566326531
Variance of Feature 1: 113.85588999926534
Mean of Feature 2: 87.36115347709884
Variance of Feature 2: 100.58482244996796

Parameters for digit 1 (class 1):
Mean of Feature 1: 19.33574387755102
Variance of Feature 1: 31.09648836321741
Mean of Feature 2: 61.30217910218628
Variance of Feature 2: 82.04463062118892
```

Naïve Bayes Classification

I implemented the Naïve Bayes classifier by calculating the likelihood of each image belonging to either digit “0” or digit “1” using the features extracted. I used the Gaussian Probability Density Function (PDF) to compute these likelihoods for each feature and class. I made the final classification decision based on the product of these likelihoods and the prior probabilities of the classes.

$$\boxed{p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)}$$

Accuracy Calculation

Accuracy calculation for both test sets (digit “0” and digit “1”) was made to evaluate the performance of the classifier. The accuracy here is defined as the proportion of correct predictions that were predicted by the classifier out of the total number of predictions.

Analysis

The Naïve Bayes classifier achieved an accuracy of **91.73%** for the digit “0” test set and **92.33%** for the digit “1” test set. With these results, the classifier was slightly more effective in recognizing images of the digit “1” compared to the digit “0”. Overall, the Naïve Bayes classifier performed well, which indicates its effectiveness for basic image classification tasks.

Output:

['0406', 44.12499566326531, 113.85588999926534, 87.36115347709884, 100.58482244996796, 19.33574387755102, 31.09648836321741, 61.30217910218628, 82.04463062118892, 0.9173469387755102, 0.9233480176211454]

Conclusion

In this project, we used a Naïve Bayes classifier for image classification with the help of the MNIST dataset. Classifying the images of digits “0” and “1” was possible by using simple statistical features for the classifier. Although this model was able to achieve a reasonable level of accuracy, there might be more improvements to be made by enhancing feature selection or using more advanced machine learning algorithms.