CSE-575
Project Report- Density Estimation & Classification
Gaurav Kumar – ASU ID- 1229081284
gkumar28@asu.edu

1. Introduction

In this project, we explore density estimation and classification using the Naive Bayes algorithm. The main objectives of the project are to extract features from the original dataset, calculate the parameters for the Naive Bayes classifiers, and utilize these classifiers for classification tasks. We apply these concepts to a specific problem involving handwritten digits recognition.

2. Problem Statement

The problem at hand involves the recognition of handwritten digits, specifically digits "0" and "1." The task is to build Naive Bayes classifiers that can differentiate between these two digits based on their features. The dataset is divided into training and testing sets for both digits "0" and "1." Our goal is to estimate the densities of the features and calculate the parameters required for the Naive Bayes algorithm. Additionally, we aim to classify the test data points accurately and evaluate the performance of our classifiers.

3. Methodology

The project consists of the following key tasks:

Task 1: Feature Extraction In this task, we extract two features for each image: the average brightness and the standard deviation of brightness. These features are calculated for both digits "0" and "1" training sets. The data is then converted into 2-dimensional data points for further analysis.

Task 2: Parameters Calculation Here, we calculate the parameters necessary for the Naive Bayes classifiers. The parameters include the mean and variance of the features for both digits "0" and "1." These parameters will be used to estimate the densities of the features. Values calculated are as below:

mean_feature1_digit0	44.19002474489796
variance_feature1_digit0	116.01946334433511
mean_feature2_digit0	87.40524535470902
variance_feature2_digit0	101.99378035755481

mean_feature1_digit1	19.401106887755102
variance_feature1_digit1	30.780551107226092
mean_feature2_digit1	61.42827980309535
variance_feature2_digit1	81.07081100836506

Task 3: Classification of Test Data Using the calculated parameters, we implement the Naive Bayes algorithm for classification. The algorithm computes the probability of each class (digit "0" or "1") based on the given features. The class with the higher probability is chosen as the prediction. We apply this classification to the test data points for both digits "0" and "1."

Task 4: Accuracy Evaluation To evaluate the performance of our classifiers, we calculate the accuracy of our predictions on the test data. The accuracy is determined by comparing the predicted labels to the actual labels of the test data points. The acceptable ranges for the parameters and accuracy values are also taken into consideration. Values calculated are as below:

Accuracy for digit 0	91.73% (0.9173)
Accuracy for digit 1	92.33% (0.9233)

4. Results and Discussion
Upon executing the tasks, we obtain the calculated parameters for the Naive Bayes classifiers. These parameters enable us to classify the test data points accurately. The project demonstrates the successful implementation of the Naive Bayes algorithm for density estimation and classification tasks. The classification accuracy achieved is within the acceptable range, validating the effectiveness of the approach.

5. Conclusion
In this project, we successfully utilized the Naive Bayes algorithm for density estimation and classification tasks on a handwritten digit's recognition problem. We extracted features from the dataset, calculated the required parameters, and applied the Naive Bayes algorithm for classification. The results show that the approach is accurate and efficient in classifying digits "0" and "1" from the test data. This project provides insights into the application of probabilistic methods for classification tasks.