CSE – 575 STATISTICAL MACHINE LEARNING

PROJECT 1 : DENSITY ESTIMATION AND CLASSIFICATION

OBSERVATION AND ANALYSIS:

In this project, we classify the digits from the original MNIST dataset specifically the images of digits 0 and 1. The train and test data for both digits are already given. We have to perform the four tasks mentioned, in order to classify the test data. The first task involves finding the average of brightness (feature 1) and standard deviation brightness of each image (feature 2) by finding the average or standard deviation of all pixel brightness values within a whole image array respectively. Here we are assuming the 2 feature values are independent and each image data is drawn from normal distribution. This paves a way for Task 2, calculation of the required 8 parameters which are mean, variance of feature 1 and 2 for both digits 0 and 1. Task 3 involves using the Naïve Bayes Classifier for finding the probability for a digit to be 0 or 1. The image is classified as 0 if the probability of zero is higher than the probability of one or vice versa. We are given the prior values of 0.5. We get the probability by using prior and estimated parameters in Gaussian Naïve Bayes formula for normal distribution. I used separate function to calculate this probability as it can be called any number of times. For Task 4, we need to calculate accuracies for digit 0 and 1 separately.

RESULT:

Parameters:

- Mean of feature1 for digit0 - 44.1095181122449
- Variance of feature1 for digit0 - 115.50944423874682
- Mean of feature2 for digit0 - 87.34421113558953
- Variance of feature2 for digit0 - 101.47007402345261
- Mean of feature1 for digit1 - 19.33325663265306
- Variance of feature1 for digit1 - 30.882705158921024
- Mean of feature2 for digit1 - 61.3042036283702
- Variance of feature2 for digit1 - 81.15514112470386

Accuracy:

- Accuracy for digit 0: 0.9173469387755102
- Accuracy for digit 1: 0.9233480176211454