# Project1

Submission data: September 9, 2019

Roey Ben Hayun

# Density Estimation and Classification using Naïve Bayes Classifier

## Data Set and formatting

I have used loadmat method to load the test and train data.

Training set dims:

28x28x5923

28x28x6742

Testing set dims:

28X28X980

28x28x1135

For ease of matrix calculation using numpy, I have changed the 3d ndarray order:

Training set dims:

5923x28x28

6742x28x28

Testing set dims

980X28x28

1135X28X28

## Feature extraction

Two i.i.d features extracted:

X1 – mean pixel intensity of each image

X2 – average of variance of each row (for each image)

## Parameter estimation

The number of MLE parameters for the model we need to calculate (assuming gaussian distribution) is 8 (I ignored the prior probability since its equal for ‘0’ and ‘1’)

Two MLE parameters per each digit per each digit : mle\_params (mue,variance) \* 2\_features \* 2\_digits = 8

### MLE Estimators

### MLE Estimators values

### Normal distribution

Gaussian Naïve Bayes assume:

### Classify the testing sample

In Naïve Bayes, the 2-D features (i.e. mean of pixel intensity and the average row variances) are conditionally independent.

Using the formula above, and the testing set for digit 0 and 1. I have extracted features X1 and X2, calculated the conditional probability using the normal distribution and the MLE parameters previously calculated and compared the resulted probability vector

For digit 0 test set I have calculated:

This resulted a vector with True and False entries. I have counted the number of trues. This gives the number of corrected predictions.

I have followed the same procedure

### Classification Accuracy

Accuracy rate for classifying digit ‘1’ = 91.0204081632653 %

Accuracy rate for classifying digit ‘0’ = 91.22448979591837 %