In this part of the project, major goal is to find density estimation and use the Naïve Bayes model for the Classification under the supervision learning method.

In this Project I have used two Python libraries namely scipy.io and numpy.np. scipy.io is mainly used to load the dataset and NumPy library used for the matrix and other mathematical calculation purpose.

1. I have extracted the following two features from the given training and testing data set of each image we used the mean and variance formula to calculate the average of the both features.

   Feature 1: Average brightness of each image

   The feature1 has been extracted by using the mean formula and considering each image size as 28*28 pixel which helps to find total number of pixels 784 of each image.

   Feature 2: Average of the variance of each rows of the image

   The feature 2 is also extracted by using the variance formula for each row and then averaged over all the rows of the image.

2. I have estimated the parameters for the normal distribution for label 0 and label 1 using the training data.

   In order to estimate the parameters of the normal distribution, we are using Maximum Likelihood estimation of mean and standard deviation. I have applied this to both the features.

   The probability density function for the normal distribution is defined by two parameters (mean and standard deviation)

   The following formula is used to calculate the MLE Density Estimation parameters

   Maximum Likelihood estimation Mean:

$$\mu = \frac{\sum_i^n x_i}{n}$$

   Maximum Likelihood estimation Standard deviation:

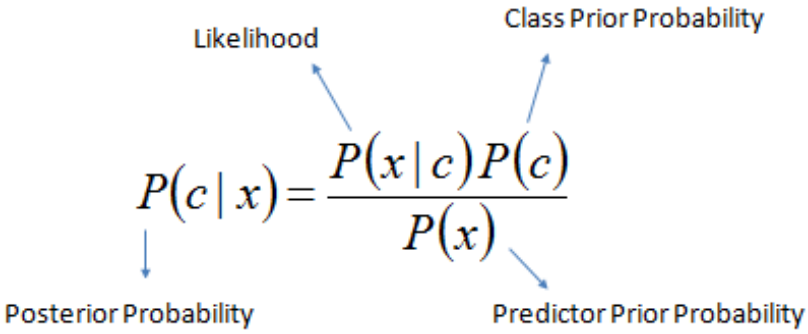
$$\sigma^2 = \frac{\sum_i^n (x_i - \mu)^2}{n}$$

The following are the parameter values of both the features estimated by using above MLE Density Estimation formula.

Mean parameter of the feature 1 for label 0 : **44.2168279054**
Mean parameter of the feature 1 for label 1 : **19.3796538528**
Variance parameter of the feature 1 for label 0 : **115.265788272**
Variance parameter of the feature 1 for label 1 : **31.4468597862**

Mean parameter of the feature 2 for label 0 : **6241.22125504**
Mean parameter of the feature 2 for label 1. : **2515.08624236**
Variance parameter of the feature 2 for label 0 : **1879799.15144**
Variance parameter of the feature 2 for label 1 : **1126270.93468**

3. We assume that the both features are independent, these feature values form normal distribution. This implies that we can choose Naïve Bayes algorithm for classification.



$$P(c \mid \mathrm{X}) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

I have used the given prior probability as 0.5 for both the classes - label 0 and label 1.

We have two features for each image of the test data. So, we calculate likelihood of both features and use them in the above Bayesian formula. And then we calculate the posteriors for both one for class label 0 and one for class label 1.

posterior (label 0) = P (label 0) * p (feature1 | label 0) * p (feature2 | label0)
posterior (label 1) = P (label 1) * p (feature1 | label 1) * p (feature2 | label1)

where:
P(label 0) or P(label 1) is called the prior probability.
P (feature1 | label 0) p (feature2 | label0) is called the likelihood probability

Likelihood is calculated using the probability density function of the normal distribution formula

$$p(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{(2\sigma^2)}}$$

By replacing these values in Bayes formula and calculated posterior probability for label 0 and label 1 for all the images in test data. If the posterior probability for label 0 is greater than that of label 1, we recognize the image as the label 0. Similarly, If the posterior probability for label 1 is greater than that of label 0, we recognize the image as the label 1.

4. The final classification accuracy for both the label 0 and label 1 for the testing data has been calculated.

    Accuracy of predicting label 0: **92.7551020408**
    Accuracy of predicting label 1: **94.5374449339**