**PROJECT PART 1 REPORT**

**Task 1 results:**

Normalized training data samples: Please refer the code output (784-D data).

Normalized testing data samples: Please refer the code output (784-D data).

**Task 2 results:**

Variance on PCA 1 (eigen value): 99.46082192723038

Variance on PCA 2 (eigen value): 39.80304184115661

PCA 1 (eigen vector): Please refer the code output (784-D data).

PCA 2 (eigen vector): Please refer the code output (784-D data).

Variance covered by the first 4 PCA components as percentage of total variance:

PCA 1: 12.69% of the variance

PCA 2: 5.08% of the variance

PCA 3: 3.34% of the variance

PCA 4: 2.98% of the variance

**Task 3 results:**

Projected training data samples:

projection\_PCA1 projection\_PCA2 Class

0 -8.287180 6.037707 0

1 -9.143768 6.154939 0

2 -3.583723 11.405663 0

3 -6.751113 11.714661 0

4 -16.196288 2.401657 0

Projected testing data samples:

projection\_PCA1 projection\_PCA2 Class

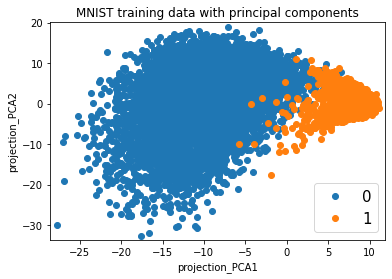
0 -9.648301 5.741488 0

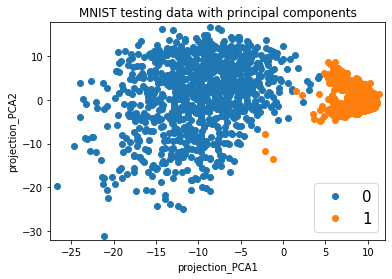
1 -9.817616 -5.068244 0

2 -8.885307 -1.681431 0

3 -18.884301 2.406377 0

4 -8.805008 2.226388 0





From the above visualization of both training and testing data in the 2-D space, we observe that each class indeed looks like a generalized gaussian distribution (with generic covariance matrix). The graph is dense at the center(mean of gaussian) indicating high probability to fall in that area, and the data points reduce in number as the distance from the mean/center of gaussian distribution increase. The shape of the spread of data points resemble to one of a gaussian distribution.

**Task 4 results:**

Parameters estimation for Class-0:

Mean : [-9.923453900798169, 0.8514234891159469]

Covariance matrix : [[25.3226061 15.8988892 ][15.8988892 79.10687303]]

Parameters estimation for Class-1:

Mean : [8.717979450374884, -0.7479948570207307]

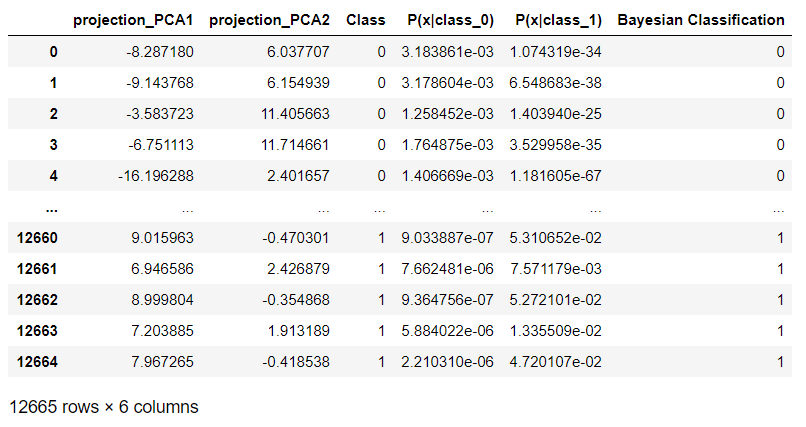
Covariance matrix: [[ 2.06660417 -0.02148369][-0.02148369 4.08381293]]

**Task 5 results:**

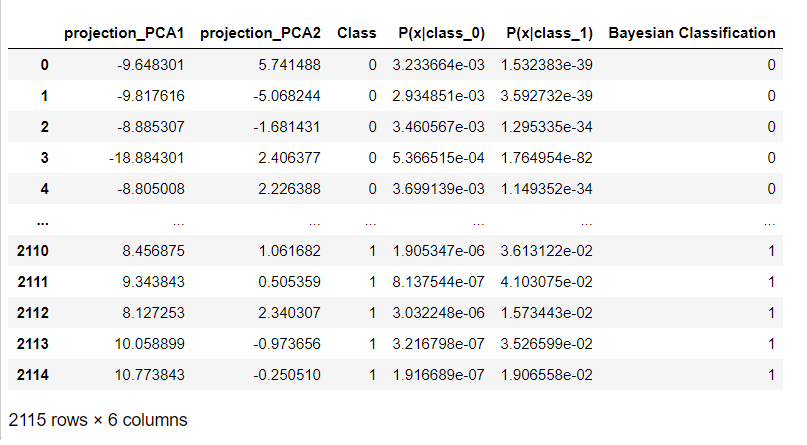
Bayesian Decision Theory for optimal classification

Since prior probabilities are the same, we only compare the likelihood of a data point given the class so as to classify it. If P(x|class\_0)>P(x|class\_1), we classify x as class 0, otherwise class 1.

**Training data:**



**Testing data:**



In the above tables, Bayesian Classification column gives the assigned class label after Bayesian optimal classification for the training and testing data points. Comparing this label to the actual/given Class label(which is Class column), we get the respective accuracy as shown below. P(x|class) is the likelihood value evaluated from the estimated parameters for the given class.

Accuracy of training dataset: 98.87879984208449

Accuracy of testing dataset : 99.19621749408984