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
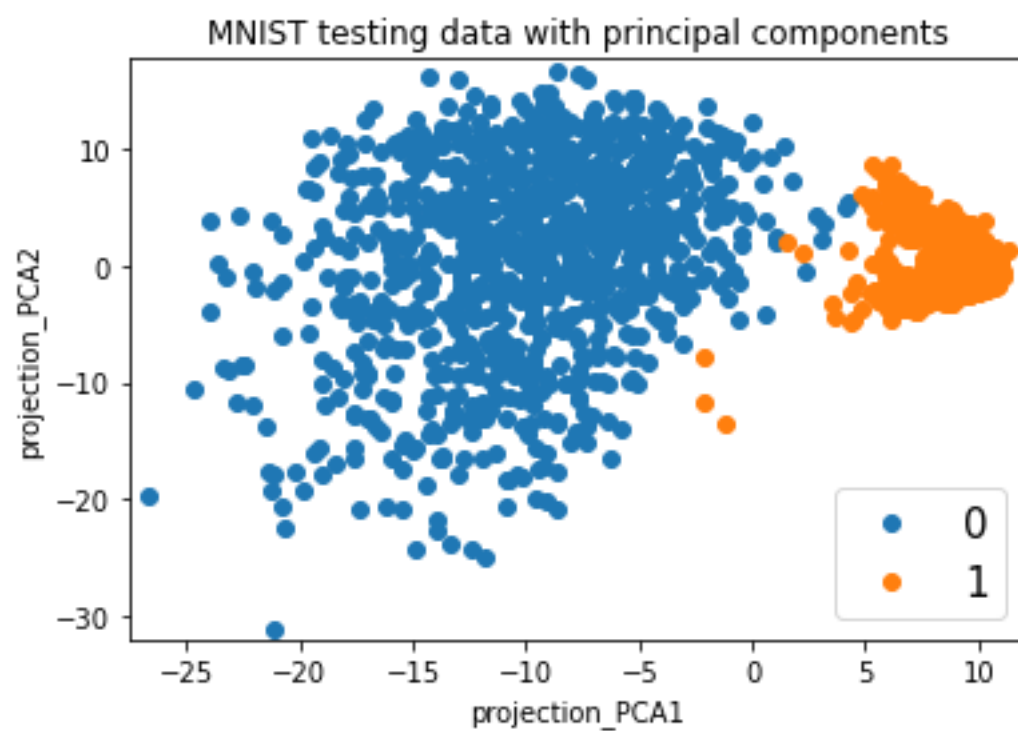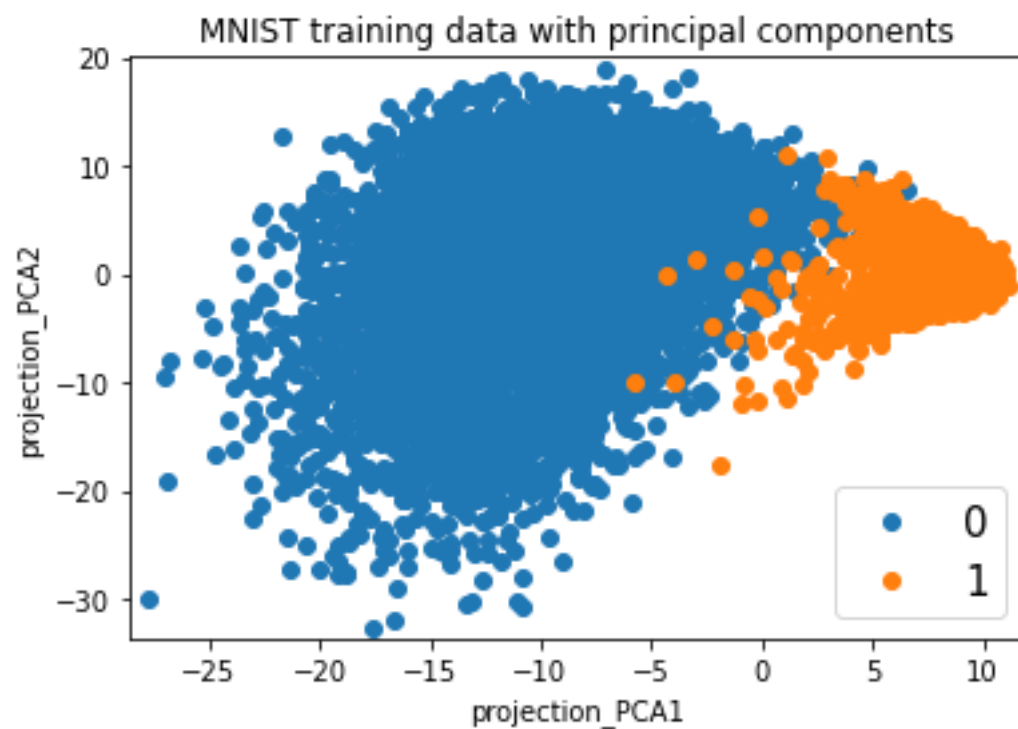PCA 3: 3.34% of the variance
PCA 4: 2.98% of the variance

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

MNIST training data with principal components

MNIST testing data with principal components

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:

	projection_PCA1	projection_PCA2	Class	P(x|class_0)	P(x|class_1)	Bayesian Classification
0	-8.287180	6.037707	0	3.183861e-03	1.074319e-34	0
1	-9.143768	6.154939	0	3.178604e-03	6.548683e-38	0
2	-3.583723	11.405663	0	1.258452e-03	1.403940e-25	0
3	-6.751113	11.714661	0	1.764875e-03	3.529958e-35	0
4	-16.196288	2.401657	0	1.406669e-03	1.181605e-67	0
...	...	...	...	...	...	...
12660	9.015963	-0.470301	1	9.033887e-07	5.310652e-02	1
12661	6.946586	2.426879	1	7.662481e-06	7.571179e-03	1
12662	8.999804	-0.354868	1	9.364756e-07	5.272101e-02	1
12663	7.203885	1.913189	1	5.884022e-06	1.335509e-02	1
12664	7.967265	-0.418538	1	2.210310e-06	4.720107e-02	1

12665 rows × 6 columns

Testing data:

	projection_PCA1	projection_PCA2	Class	P(x|class_0)	P(x|class_1)	Bayesian Classification
0	-9.648301	5.741488	0	3.233664e-03	1.532383e-39	0
1	-9.817616	-5.068244	0	2.934851e-03	3.592732e-39	0
2	-8.885307	-1.681431	0	3.460567e-03	1.295335e-34	0
3	-18.884301	2.406377	0	5.366515e-04	1.764954e-82	0
4	-8.805008	2.226388	0	3.699139e-03	1.149352e-34	0
...	...	...	...	...	...	...
2110	8.456875	1.061682	1	1.905347e-06	3.613122e-02	1
2111	9.343843	0.505359	1	8.137544e-07	4.103075e-02	1
2112	8.127253	2.340307	1	3.032248e-06	1.573443e-02	1
2113	10.058899	-0.973656	1	3.216798e-07	3.526599e-02	1
2114	10.773843	-0.250510	1	1.916689e-07	1.906558e-02	1

2115 rows × 6 columns

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