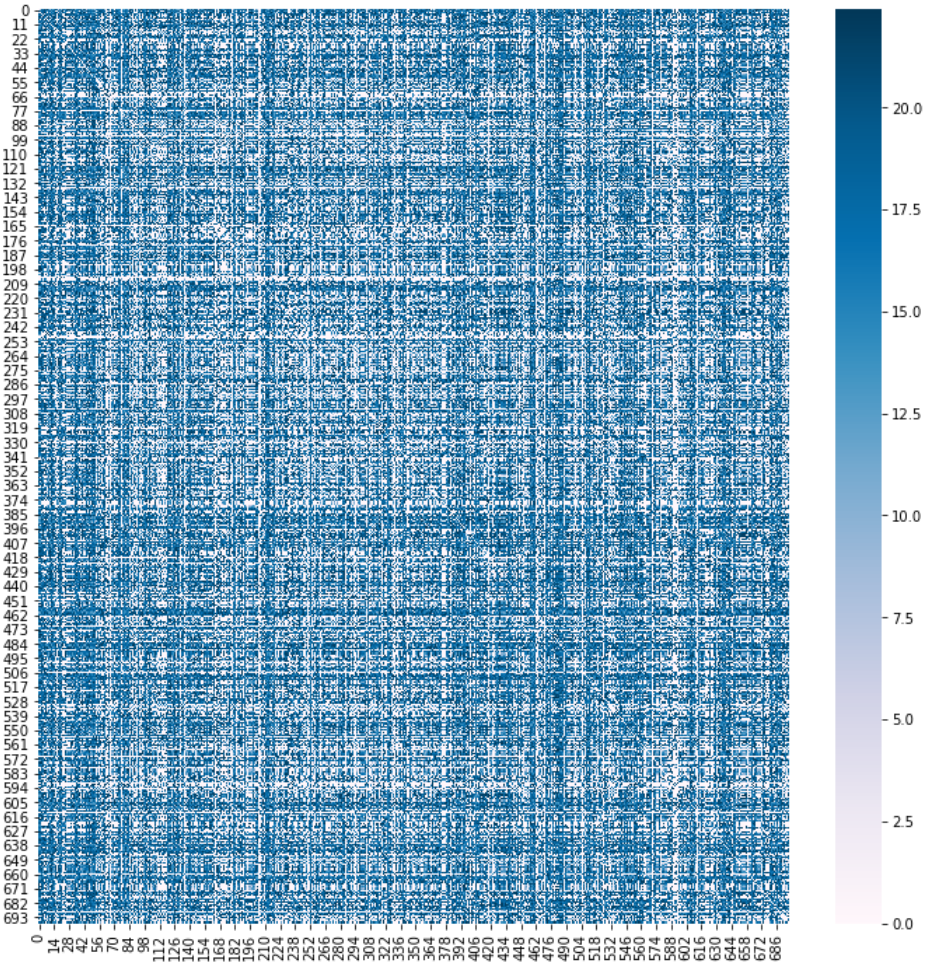
Homework 2

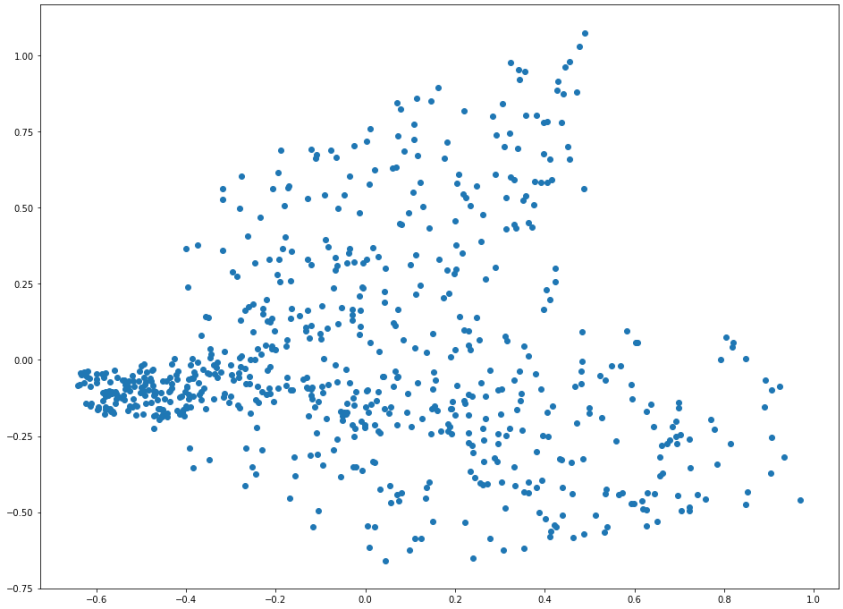
Jian Pang GT id: 903241601

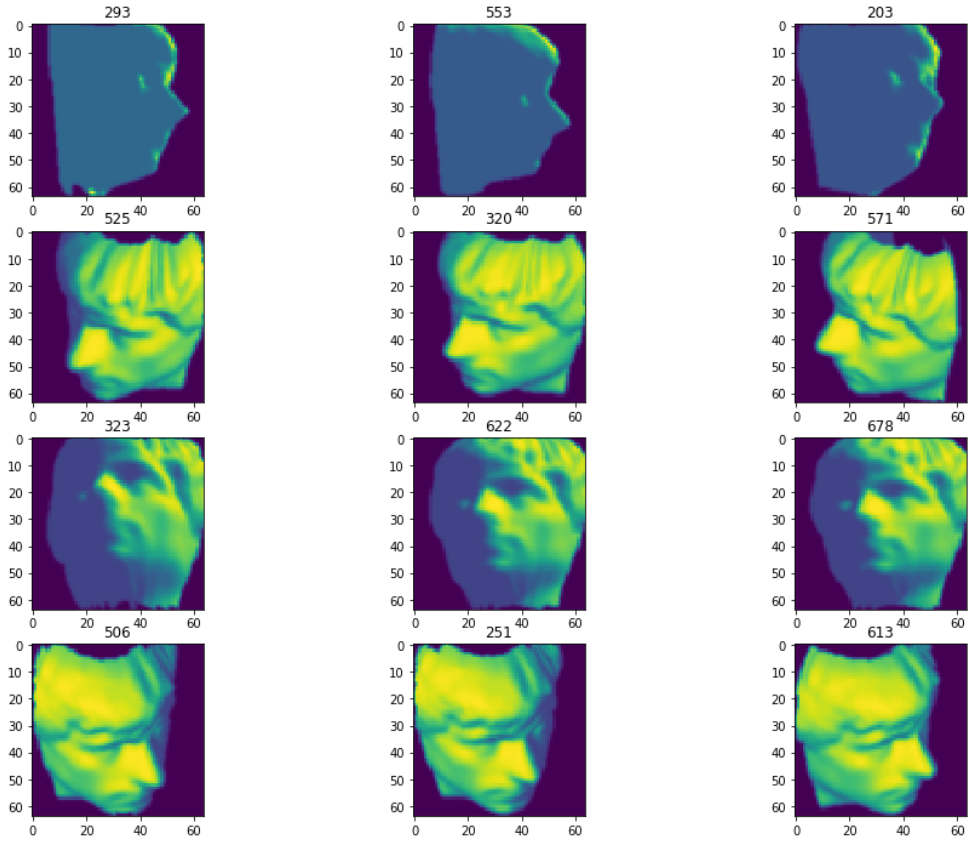
Problem 1.

Q1.



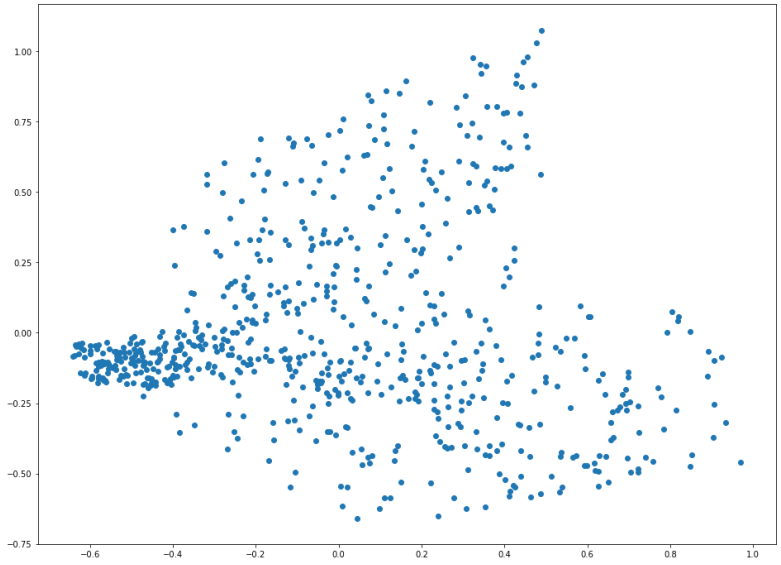
Q2.

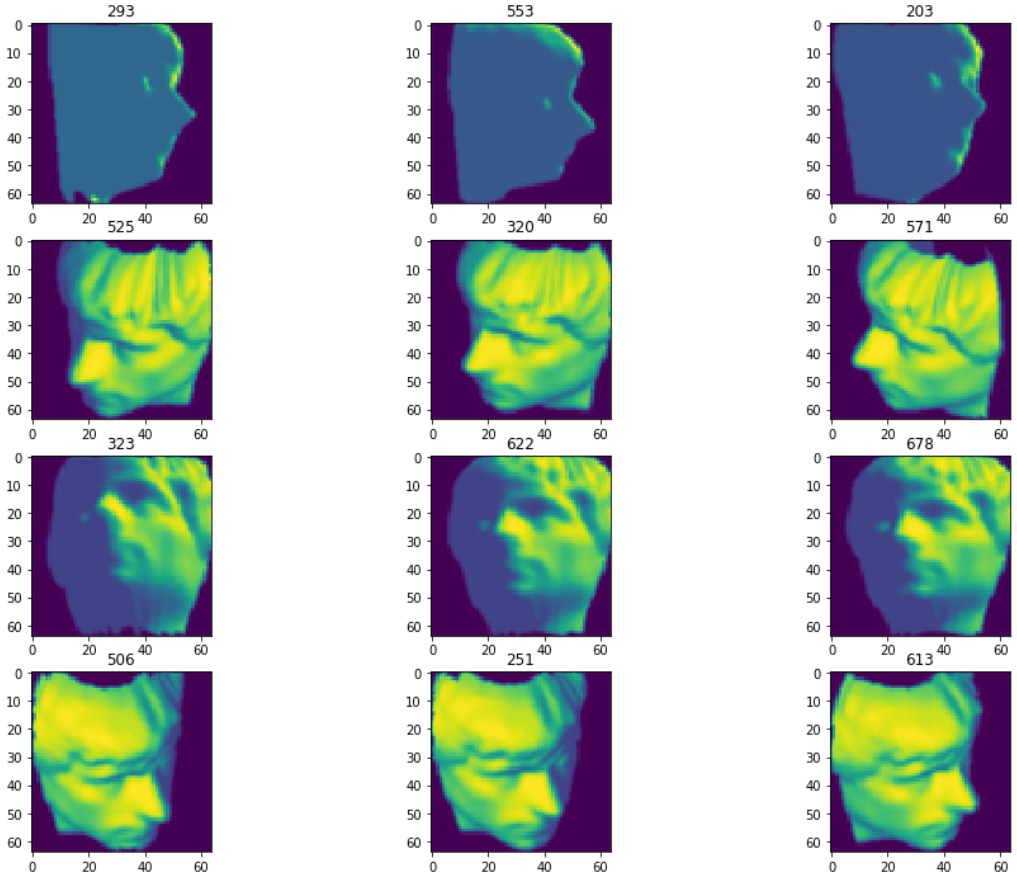




I find the three points that locate on either end of the two axis. So there are 12 images altogether. We can see that they all have similar face orientation.

Q3.

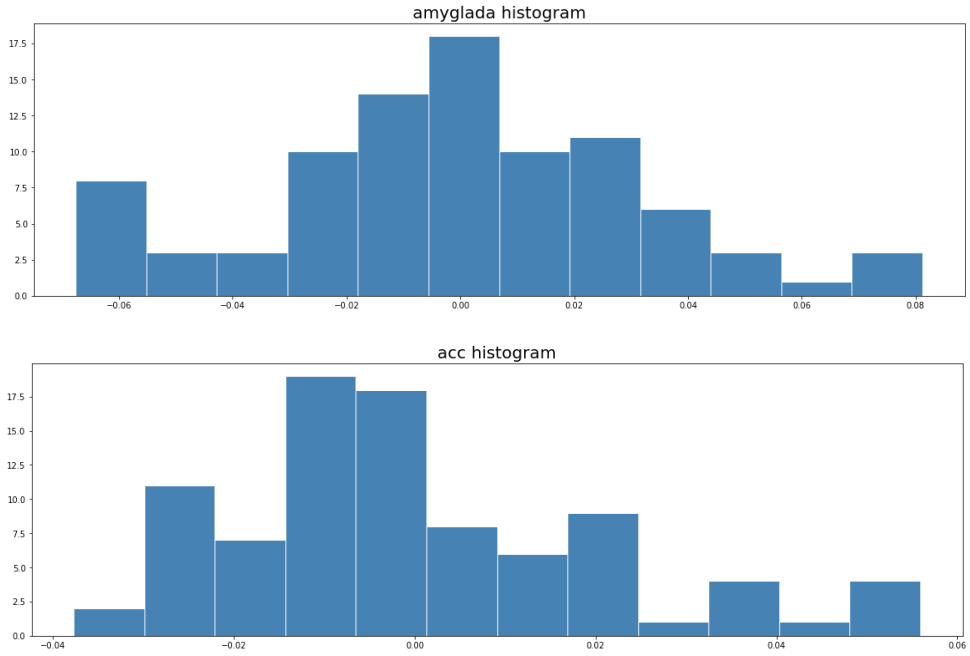




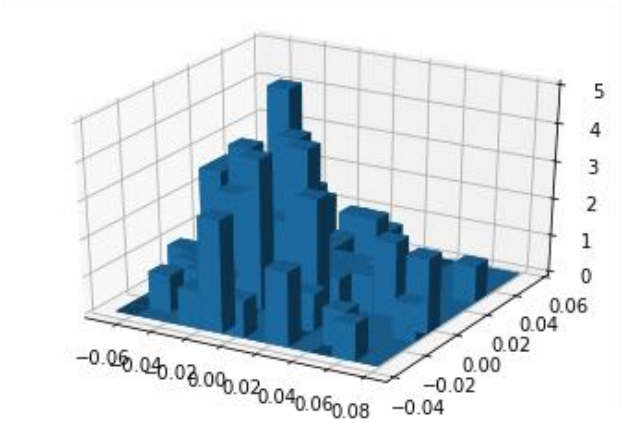
Using Manhattan distance, I don't see a significant difference in terms of the results compared to using Euclidean distance.

Problem 2.

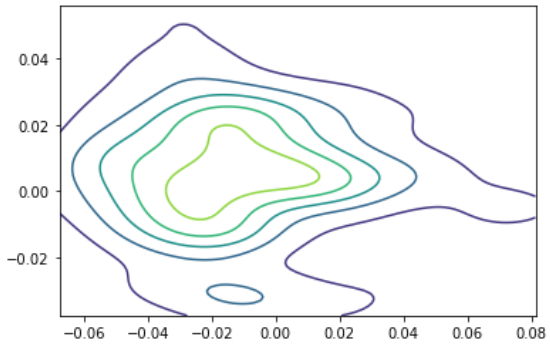
Q1.

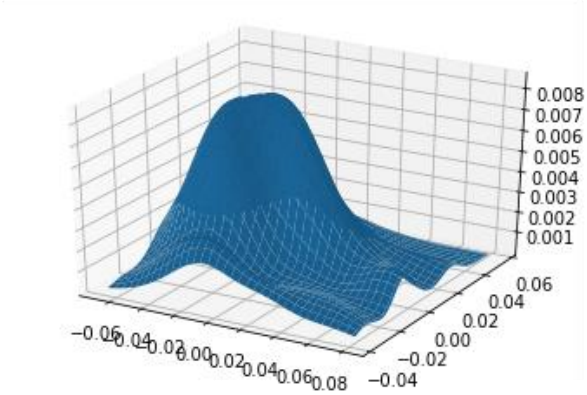


After comparison between different number of bins, I decided on using 12 bins. Below is the 3D plot for the dataset.

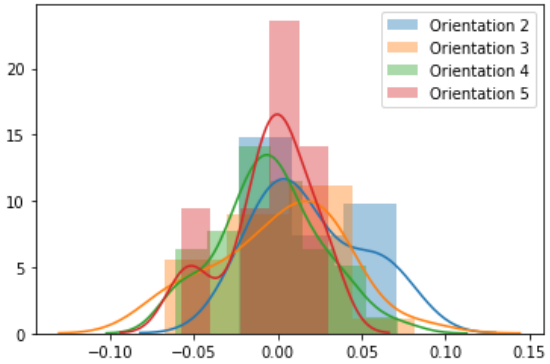


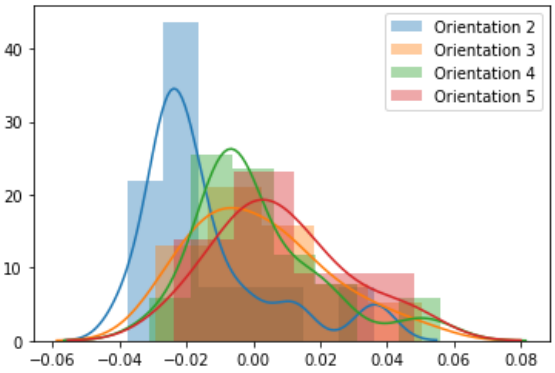
Q2.


0.04
0.02
0.00
−0.02
−0.06
−0.04
−0.02
0.00
0.02
0.04
0.06
0.08


0.008
0.007
0.006
0.005
0.004
0.003
0.002
0.001
0.06
0.04
0.02
0.00
−0.02
−0.04
−0.06
−0.04
−0.02
0.00
0.02
0.04
0.06
0.08

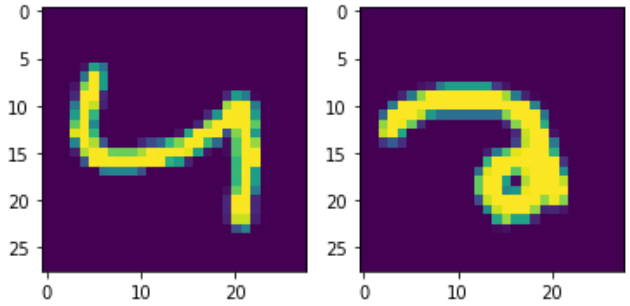
Q3.


Orientation 2
Orientation 3
Orientation 4
Orientation 5
20
15
10
5
0
−0.10
−0.05
0.00
0.05
0.10
0.15


Orientation 2
Orientation 3
Orientation 4
Orientation 5
40
30
20
10
0
−0.06
−0.04
−0.02
0.00
0.02
0.04
0.06
0.08

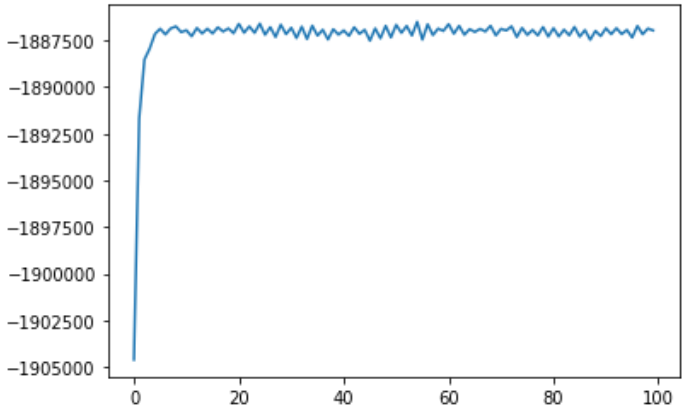
Problem 3.

Q1.



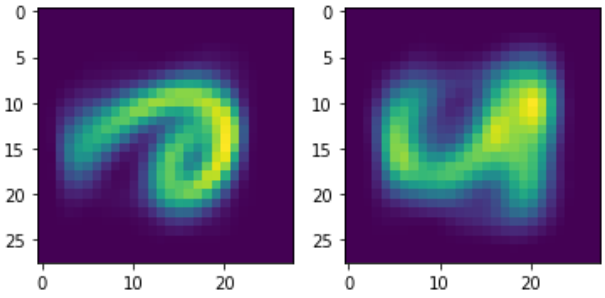
Q2.

After testing with different r values, r is set at 250 when doing low-rank approximation. Maximum iteration is 100 steps.



Q3.

Below is the converged mean vector visualized.



The gaussian mix ratio is 0.382 and 0.618. The accuracy rate is 84.5%. The “2”s being misclassified are 20% of the total “2”s. Also there are 6% of the “6”s being misclassified

Q4.

The accuracy for KMeans is 94%, 2% of the “2”s are misclassified, and 6% of the “6”s are misclassified.

Q5.

Because the data has few dimensions, there is no need for using low rank approximation, and it's possible to use the package multivariate normal pdf function to calculate probability density all the way. I generate a modified version of GMM function based on this question.

The result shows that the accuracy for KMeans is 94%, 2% of the “2”s are misclassified, and 6% of the “6”s are misclassified.