Homework 5

Collaborators:
    Name: Youchao Zhang
    Student ID: 3170100125

Problem 4-1.   K-Means Clustering

Finally, we will run our first unsupervised algorithm – k-means clustering.

(a) Visualize the process of k-means algorithm for the two trials.
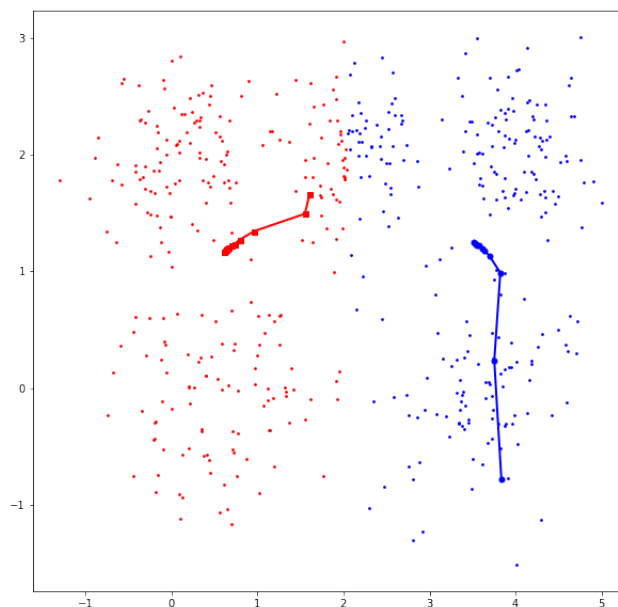
Answer:

Figure 1: k-means algorithm for the two trials

(b) How can we get a stable result using k-means?

Answer:
The issue that the outcome of k-means algorithm is very sensitive to cluster centroids initialization from the above experiment.
Prior knowledge of the data, or simple analysis of the data can be obtained
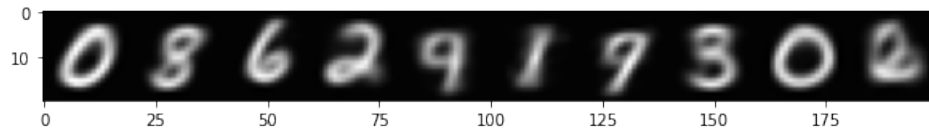
(c) Visualize the centroids.

Answer:

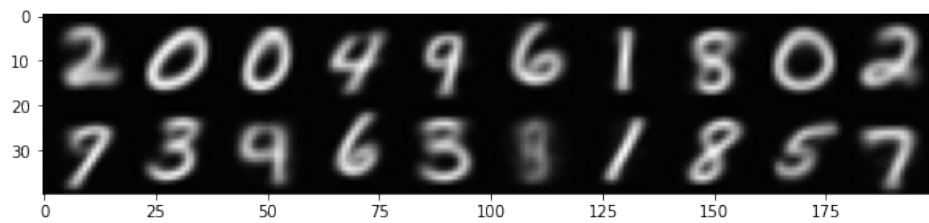**Figure 2**: k-means algorithm for 10 clusters

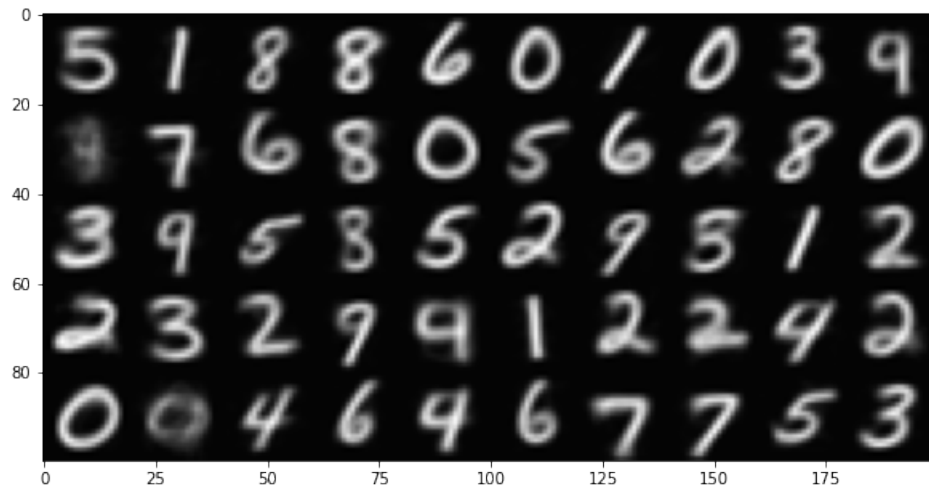**Figure 3**: k-means algorithm for 20 clusters

Figure 4: k-means algorithm for 50 clusters

(d) Vector quantization.

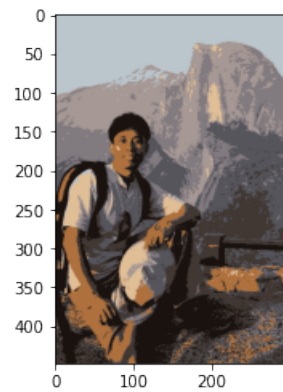**Answer:**

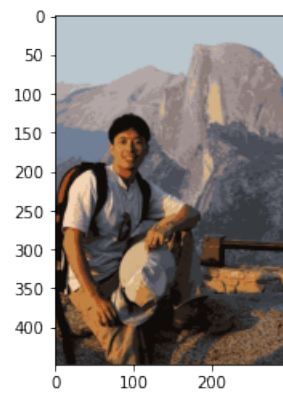**Figure 5**: k-means algorithm for 8 clusters

Figure 6: k-means algorithm for 16 clusters
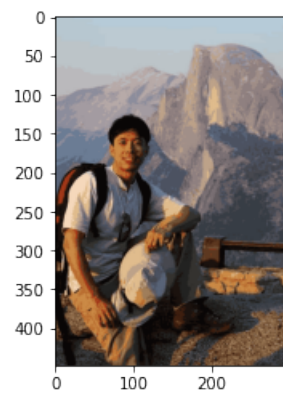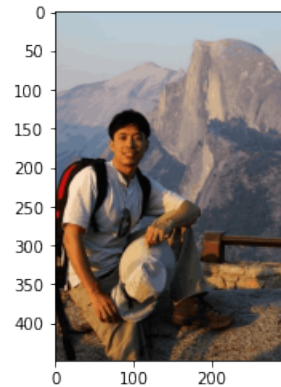

Figure 7: k-means algorithm for 32 clusters

Figure 8: k-means algorithm for 64 clusters

$$If\ we\ set\ K\ to\ 64\ , compress\ ratio\ =\ log_2(64)/24 = 0.25$$

Problem 4-2. Spectral Clustering

In this problem, we will try a dimensionality reduction based clustering algorithm – Spectral Clustering.

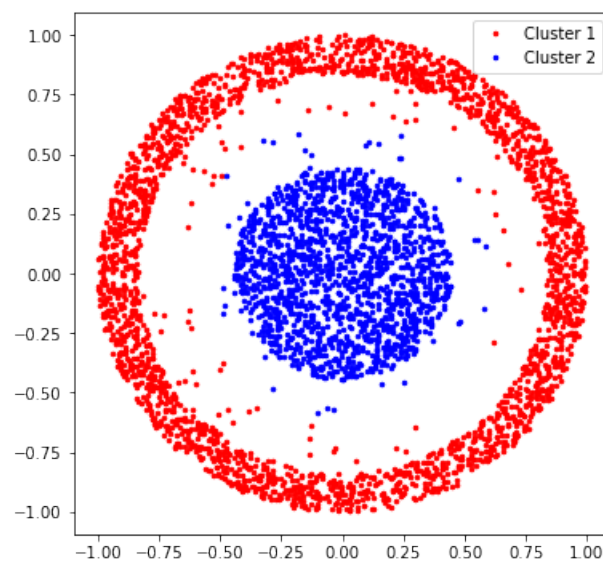 **(a)** We will first experiment Spectral Clustering on synthesis data

      Answer:
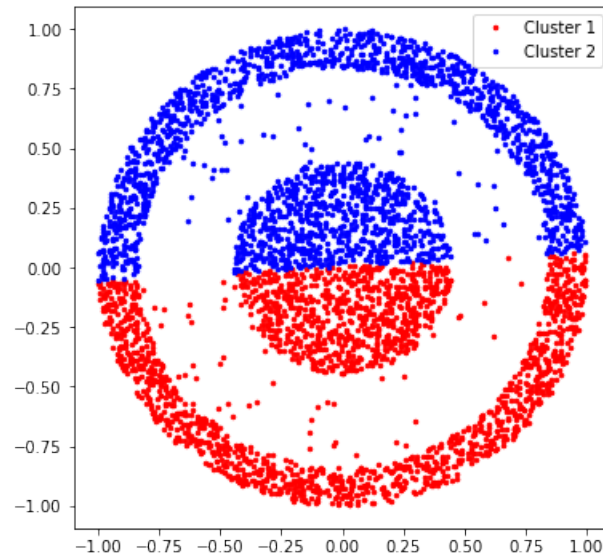

Figure 9: Spectral Clustering

Figure 10: Kmeans

(b) Now let us try Spectral Clustering on real-world data.

Answer:
spectral clustering accuracy: 0.716 normalized mutual information: 0.638
K-means clustering accuracy: 0.4685 normalized mutual information: 0.382

Problem 4-3.   Principal Component Analysis Let us deepen our understanding of PCA by the following problems.

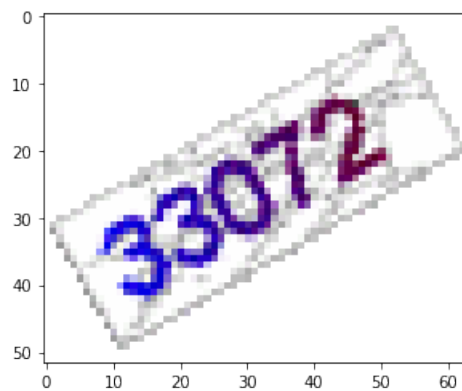**(a)** Your task is to implement *hack_pca.m* to recover the rotated CAPTCHA image using PCA.

Answer:

Figure 11: rotated CAPTCHA
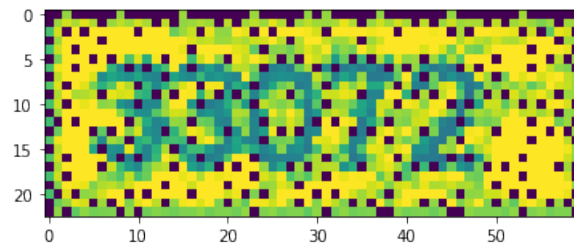
Figure 12: recovered CAPTCHA

(b) Now let us apply PCA to a face image dataset.

Answer:

(I)

Figure 13: Eigenface

(II)PCA algorithm
PCs=8,error rate = 0.495
PCs=16,error rate = 0.455
PCs=32,error rate = 0.38
PCs=128,error rate = 0.37
(III)

Figure 14: 8PCs

Figure 15: 16PCs

Figure 16: 32PCs

Figure 17: 64PCs

(III)LDA algorithm
PCs=8,error rate = 0.105
PCs=16,error rate = 0.075
PCs=32,error rate = 0.055
PCs=128,error rate = 0.05