Homework 4

Collaborators:

Name: Jessie Peng

Student ID: xxxxxxxxxx

Problem 4-1.   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: In this experiments, k-means performs badly, while spectral clustering (with k_ in_ knn_ graph = 150 and threshold = 0.2) performs quite well.
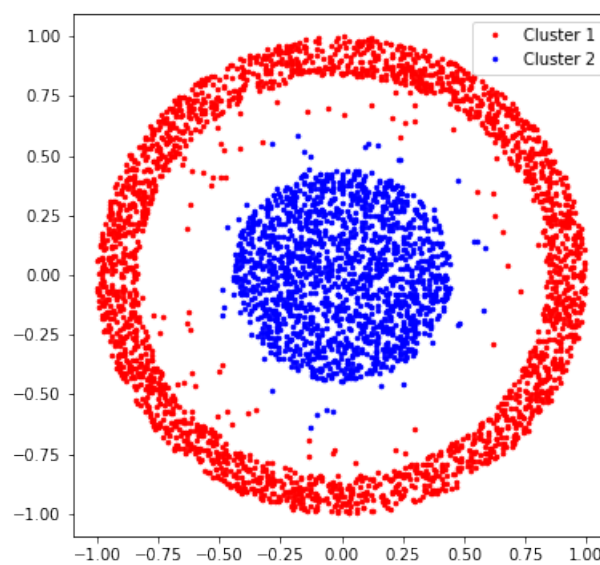
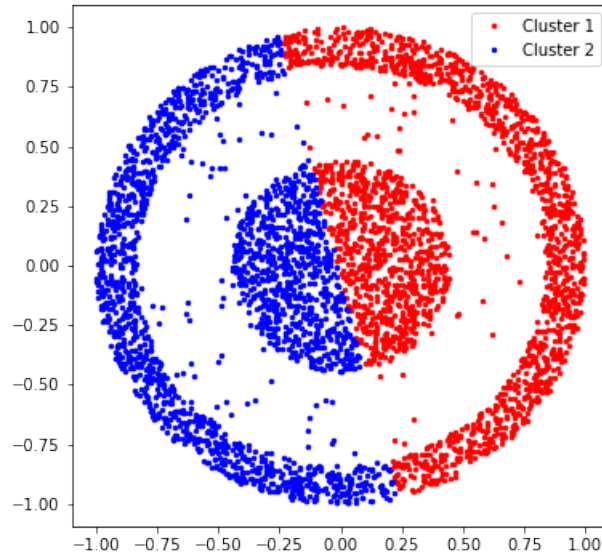

Figure 1: Spectral clustering

Figure 2: K-means clustering

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

Answer: The result of spectral clustering (with k_in_knn_graph = 10) has higher accuracy and normalized mutual information (with the ground truth clustering) than that of k-means clustering as shown below:

	acc	mutual info
spectral	0.7084	0.6243
k-means	0.4868	0.2148

Table 1: Acc and mutual info of spectral and k-means clustering

The higher accuracy means that spectral clustering assigns more data to the correct clusters, while the higher normalized mutual information describes the higher similarity between the resulting clustering and the ground truth. They both show that in this case, spectral clustering performs better than k-means clustering.

Problem 4-2. **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.

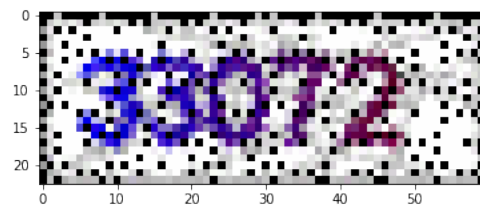**Answer:** The recovered CAPTCHA are shown below (I didn't do interpolation, so there exist some black pixels):
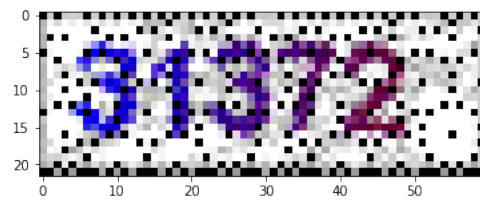
Figure 3: Recovered figure(1)
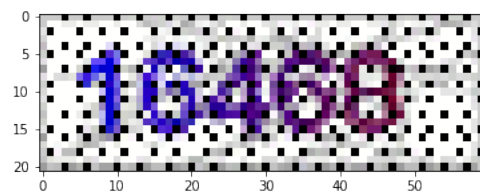

Figure 4: Recovered figure(2)

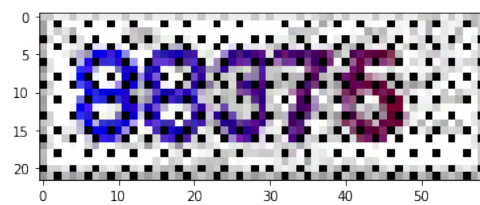**Figure 5**: Recovered figure(3)

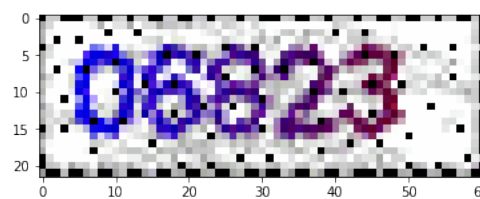**Figure 6**: Recovered figure(4)

Figure 7: Recovered figure(5)

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

   Answer:

Figure 8: Eigen faces

The testing error rate and recovered images after dimension reduction by PCA are shown below:

dimensionality	testing error rate
8	0.255
16	0.175
32	0.155
64	0.135
128	0.135

Table 2: Testing error rate of PCA

Figure 9: Recovered images with reduced dimensionality 8

Figure 10: Recovered images with reduced dimensionality 16

Figure 11: Recovered images with reduced dimensionality 32

Figure 12: Recovered images with reduced dimensionality 64

Figure 13: Recovered images with reduced dimensionality 128

Figure 14: Original images

The testing error rate after dimension reduction by LDA are shown below:

dimensionality	testing error rate
8	0.115
16	0.055
32	0.045
64	0.03
128	0.03

Table 3: Testing error rate of LDA