

Homework 5

Collaborators:

Name: Chen Zhibo

Student ID: 3190100923

Problem 5-1. 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:

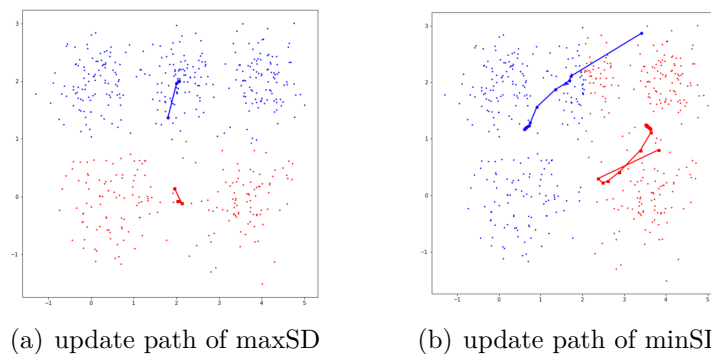


Figure 1: image of trials with largest and smallest SD

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

Answer: One method is to use Partitioning Around Medoids. Instead of using the average as the central, use "medoid" point that is most close to the average and exactly exists in data sets as the central points.

Another way is to use Kmeans++. When picking the startup k points, it doesn't randomly choose from all possible points, it picks one by one. When picking the next point, it tries to choose a point that is as far from points that have been picked and use different weights to evaluate which one to pick. It's a good trade-off method.

(c) Visualize the centroids.

Answer:

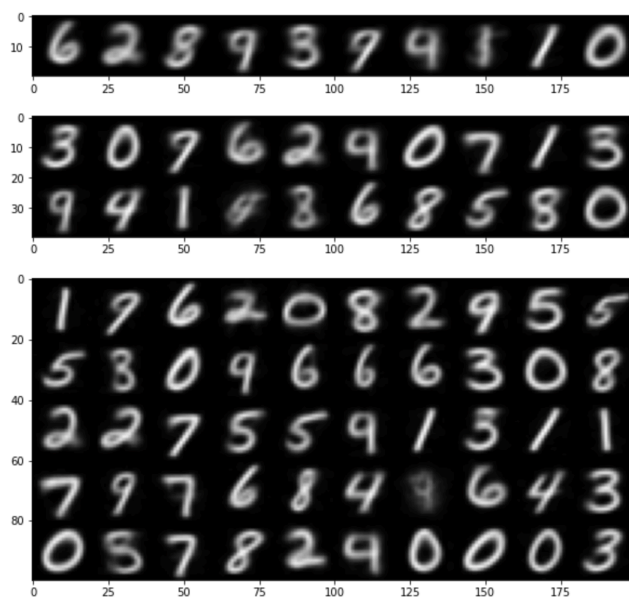


Figure 2: Centroids Visualization

(d) Vector quantization.

Answer:

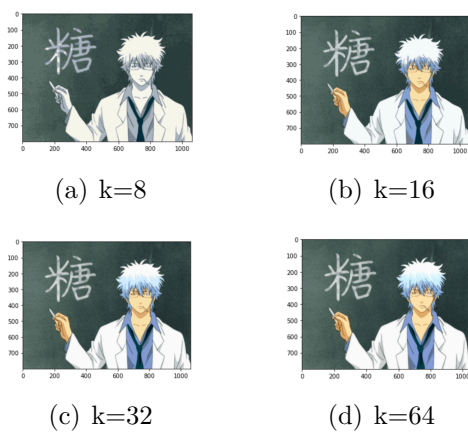


Figure 3: image of KNN

Problem 5-2. 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:

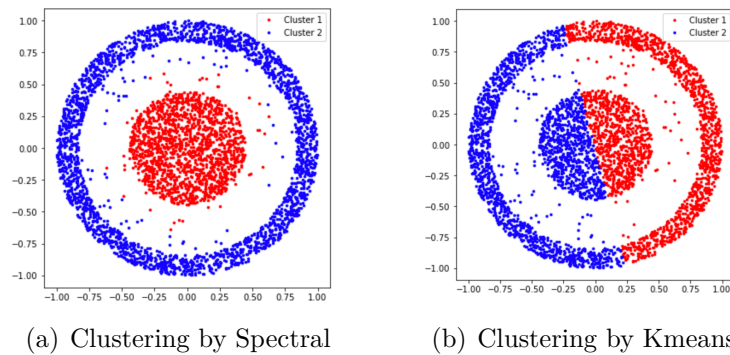


Figure 4: imag of trials with largest and smallest SD

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

Answer: After attempting several parameters, I finally get accuracy of 74.54% and normalized mutual information of 61.5% on the average of 50 times.

0.7454131918119787 0.6152773912542459

Figure 5: Centroids Visualization

Problem 5-3. 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.

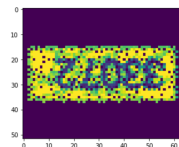


Figure 6: PCA CAPTCHA

Answer:

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

Answer:



Figure 7: Eigenface

```
k=8 error:0.245
k=16 error:0.2
k=32 error:0.18
k=64 error:0.15
```

Figure 8: Error Rate of Dimension Reduction



Figure 9: Face Recovering