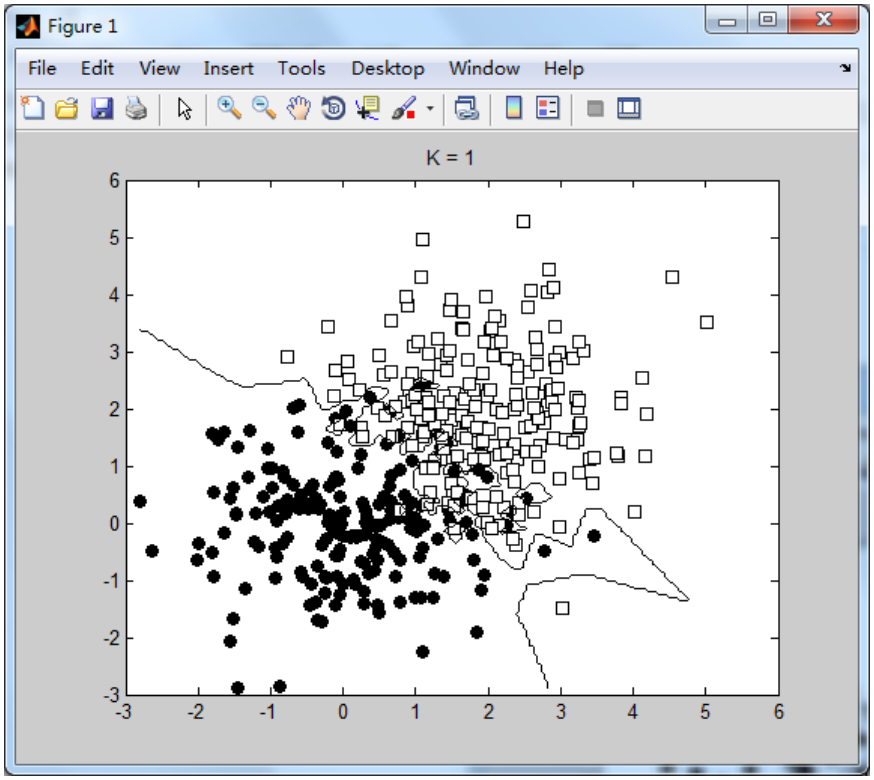
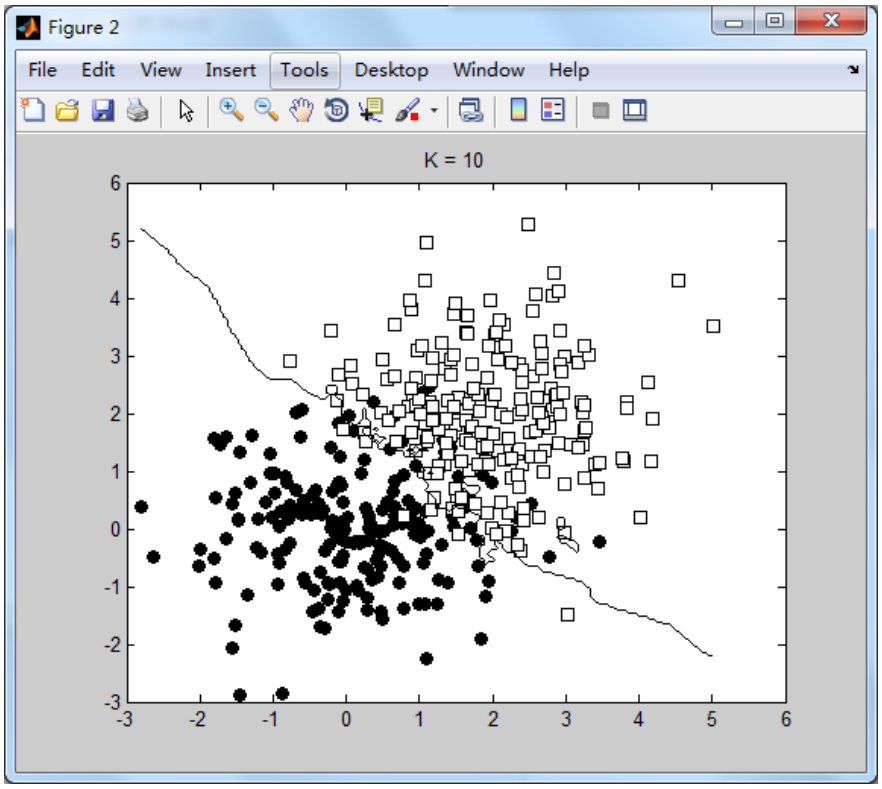
HW3_Answer

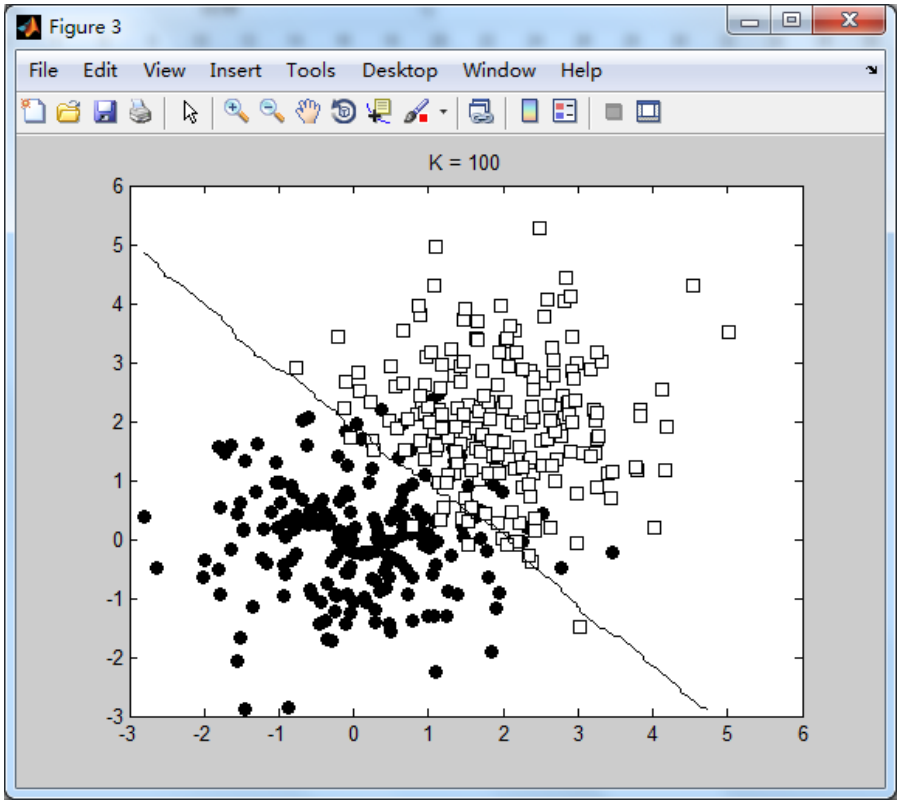
3120000060 秦昇

1. K-Nearest Neighbor

(a)







(b)

Dealing with real world data, I can run several tests using different values of k to see the

results and choose one k from them with the best performance and consider the time costs that we can afford.

(c)

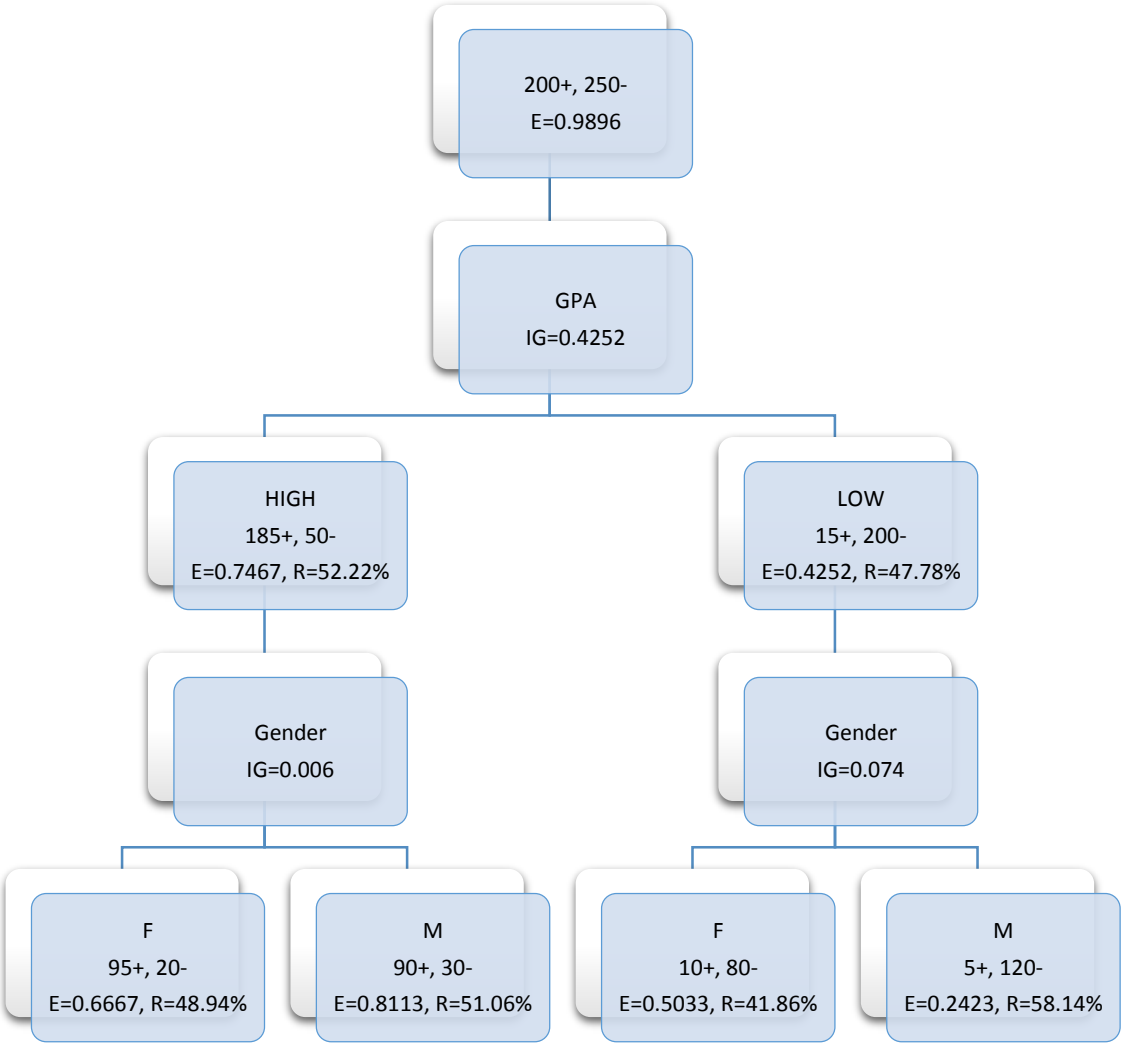
```
66      % test and calculate test error
67      % load test data and test
68 -    result=zeros(1,0);
69 -    for testI=1:20
70 -        filePath=['train_photo/#', num2str(testI+41), '.bmp'];
71 -        result=[result, hack(filePath)];
72 -    end
73
74      % calculate error rate
75 -    error=0;
76 -    for i=1:100
77 -        if(result(i)~=testY(i))
78 -            error=error+1;
79 -        end
80 -    end
81 -    errorRate=error/100
```

```
errorRate =

     0
```

During testing, I found that when k is 2 or 4, the error rate is 0.

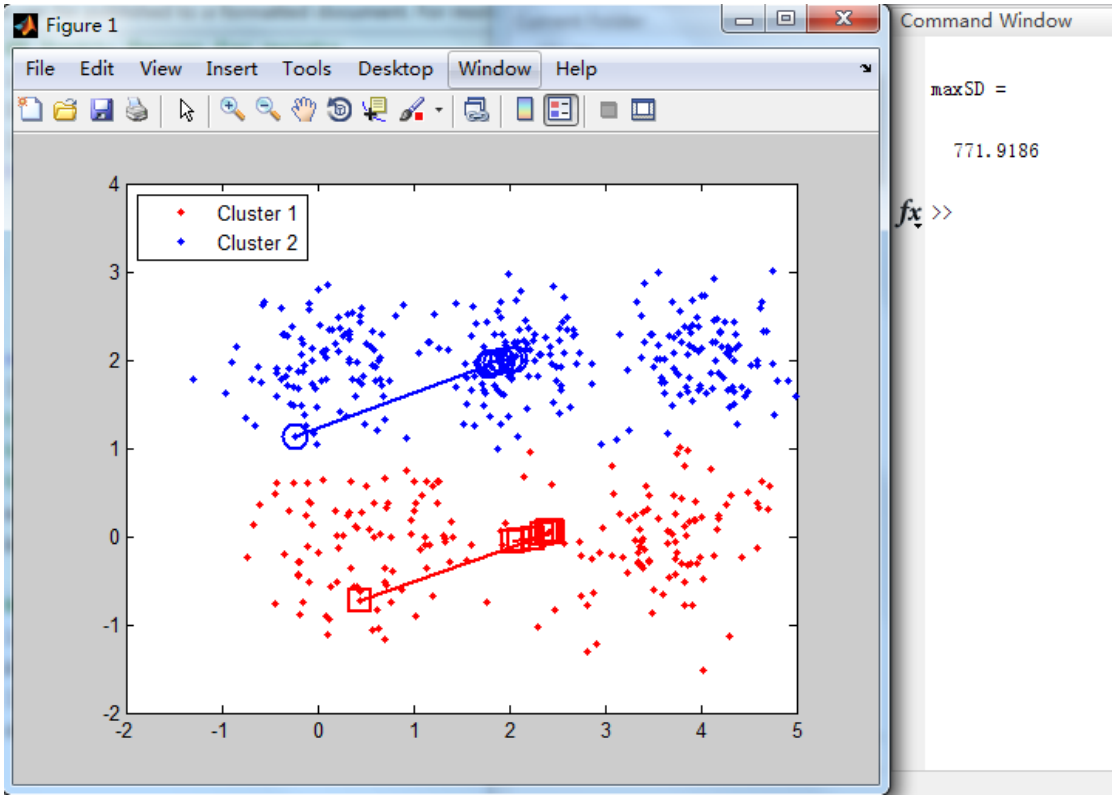
2. Decision Tree and ID3



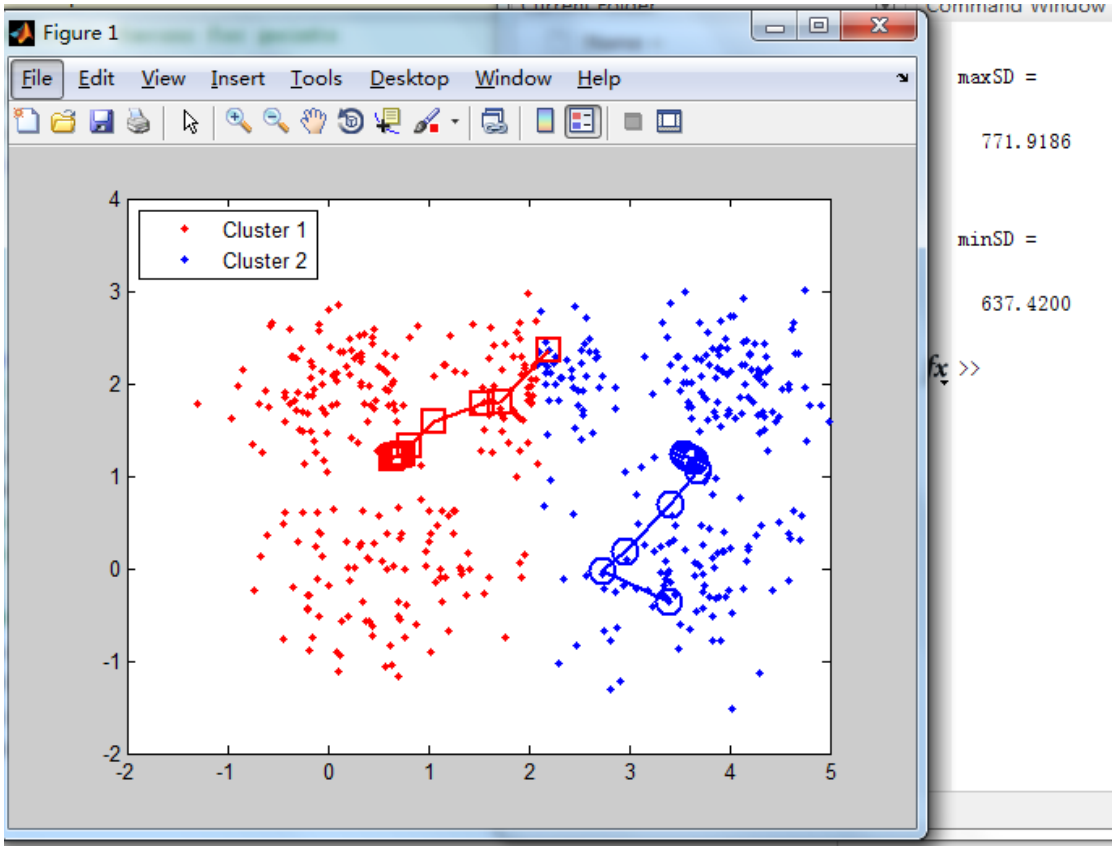
3. K-Means Clustering

(a)

Trials with largest SD:



Trials with smallest SD:



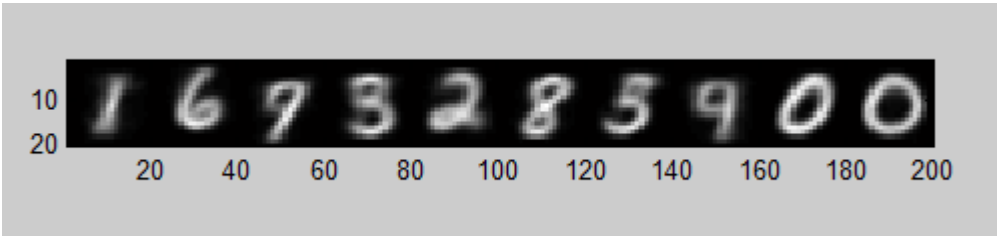
(b)

We can do K-Means algorithm many times on the data set to get the optimal clustering result by choosing one with the smallest SD or by human mind.

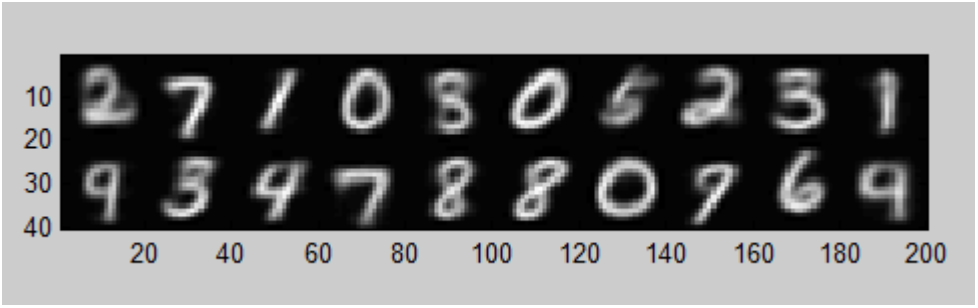
Or we can let the initial center points distribute more separately to get the stable result.

(c)

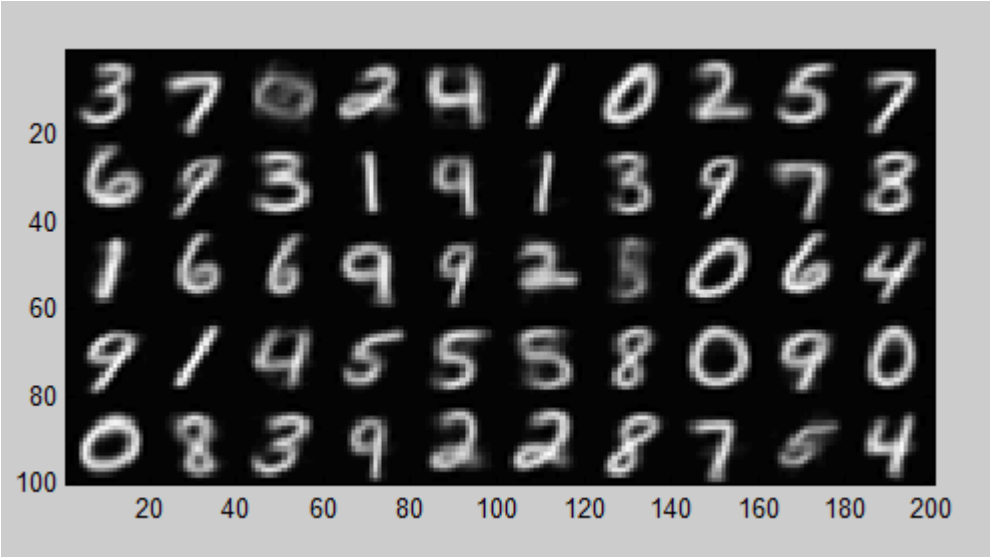
K=10:



K=20:



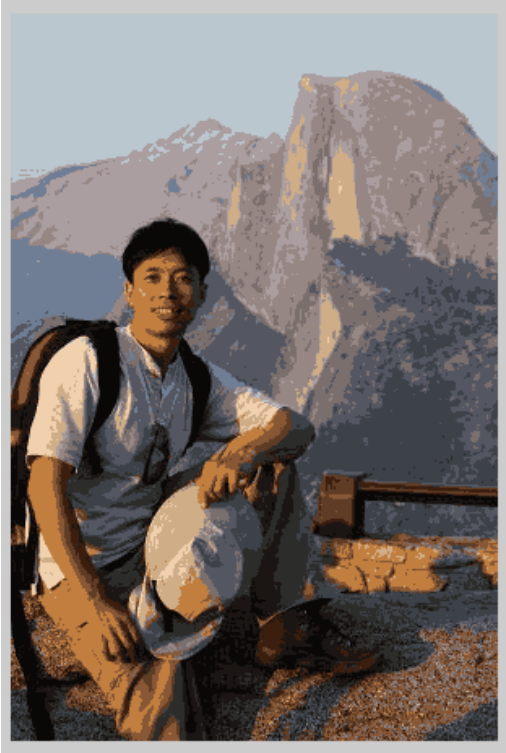
K=50:



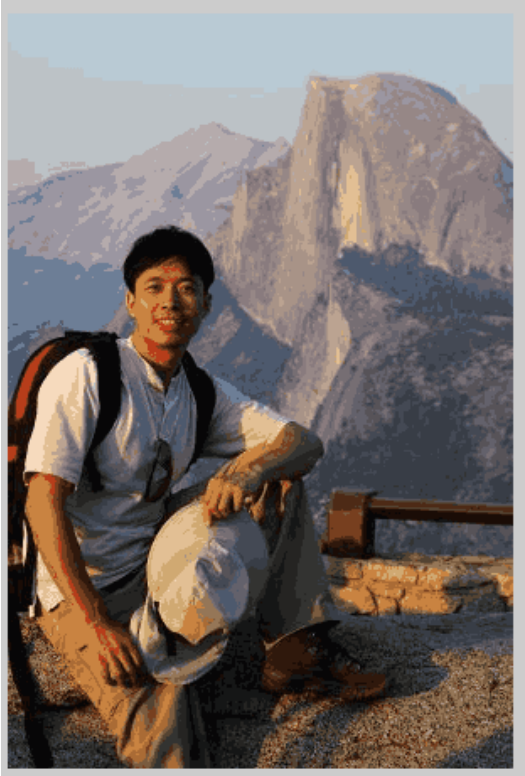
(d)

K=8:

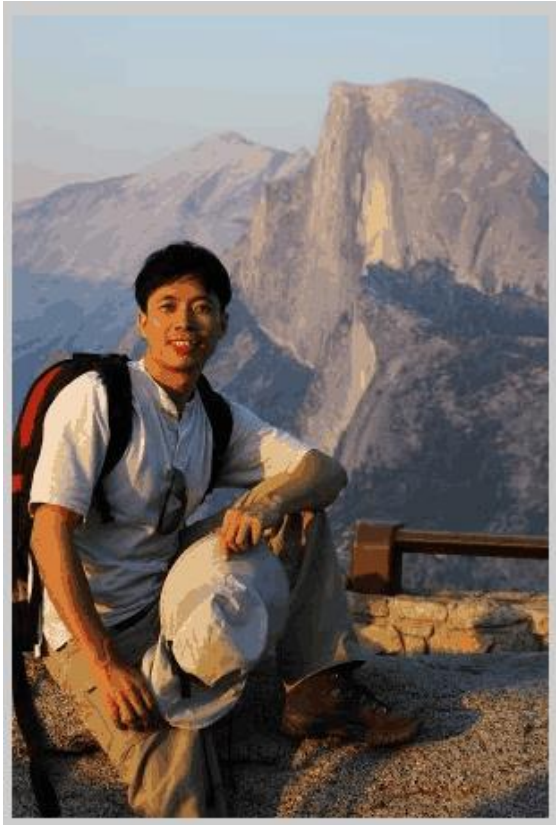
K=16:



K=32:



K=64:

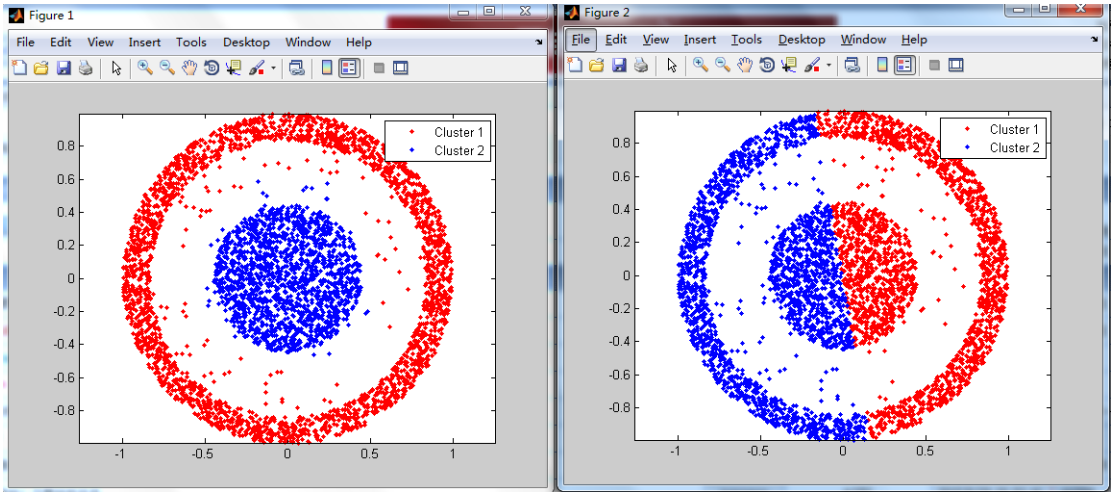


When we set K to 64, we can represent each pixel with $\log_2(64) = 6$ bits rather than 24 bits, so the compress ratio is roughly 75%.

4. Spectral Clustering

(a)

K=100, threshold=0.1



(b)

```
>> spectral_exp2

average_spectral_AC =

    0.8675


average_spectral_MIhat =

    0.6384


average_kmeans_AC =

    0.5365


average_kmeans_MIhat =

    0.3272
```