Homework 3

Collaborators:

Name: Shen Kai

Student ID: 21921071

Problem 3-1. Neural Networks

In this problem, we will implement the feedforward and backpropagation process of the neural networks.

(a) Answer: The test accuracy are shown as following:

```
loss:0.239, accuracy:0.929
```

Figure 1: The accuracy on test dataset.

Problem 3-2. K-Nearest Neighbor

In this problem, we will play with K-Nearest Neighbor (KNN) algorithm and try it on real-world data. Implement KNN algorithm (in *knn.m*/*knn.py*), then answer the following questions.

(a) Try KNN with different K and plot the decision boundary.

Answer:

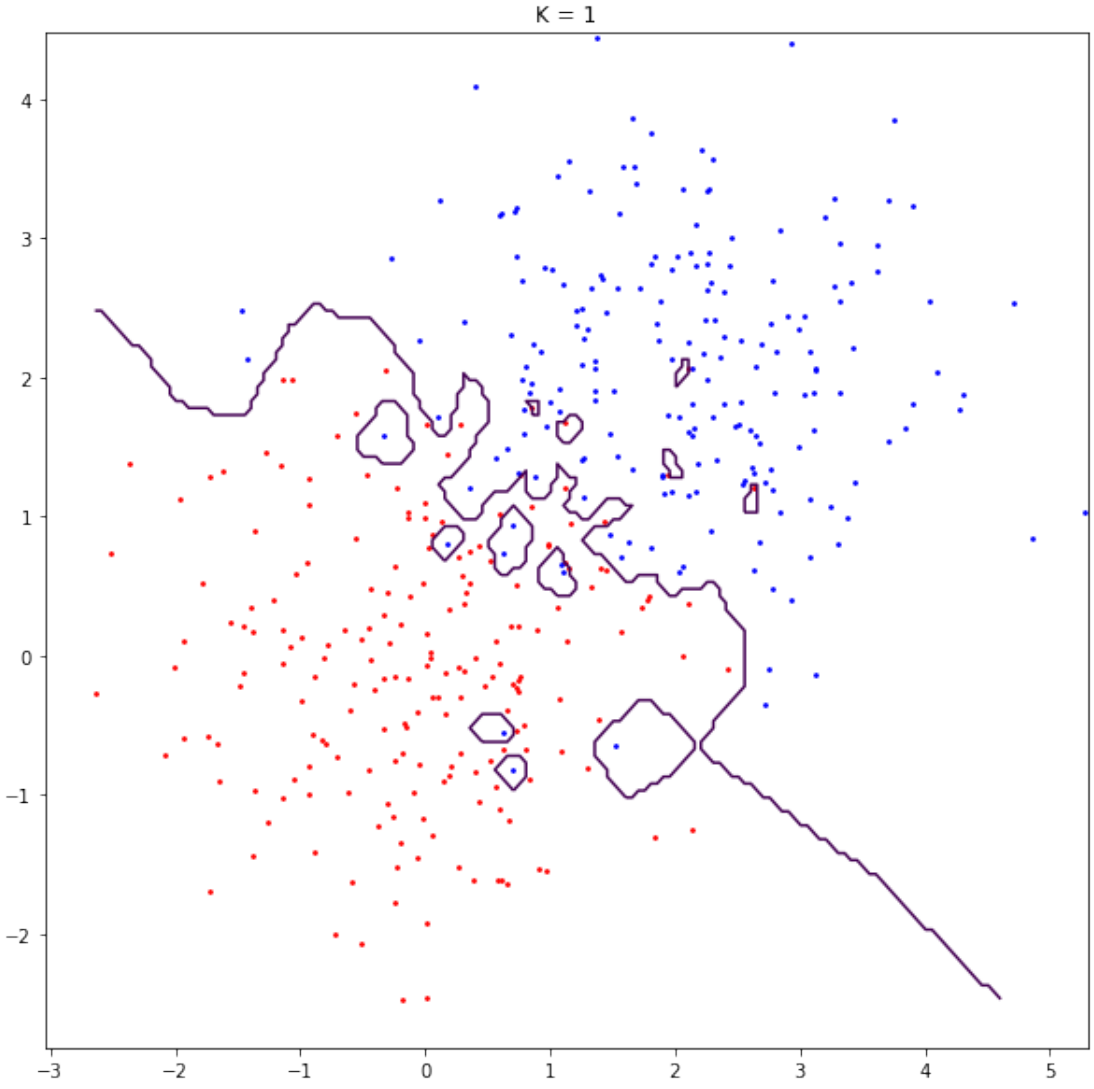

Figure 2: k=1.

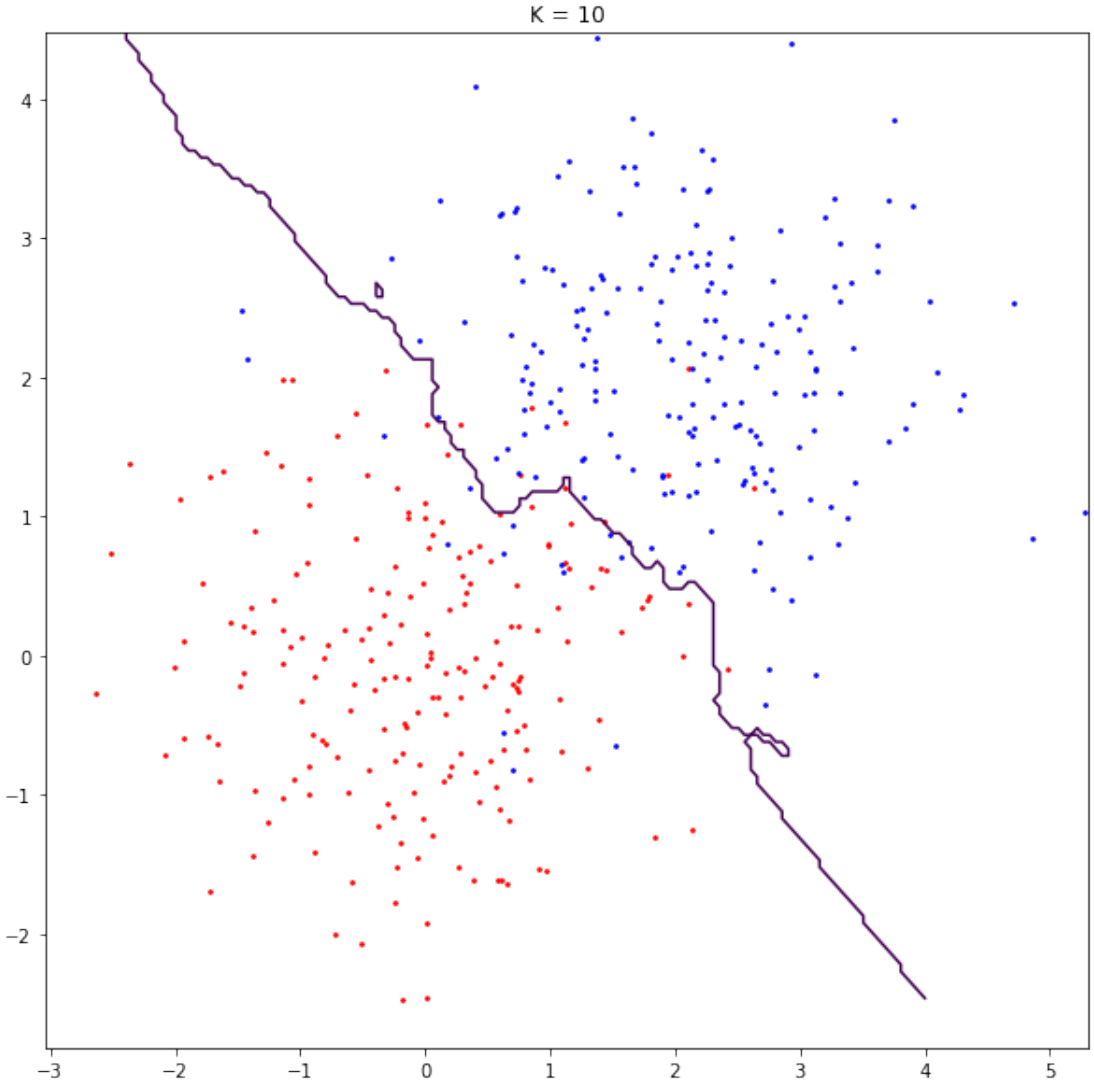

Figure 3: k=10.

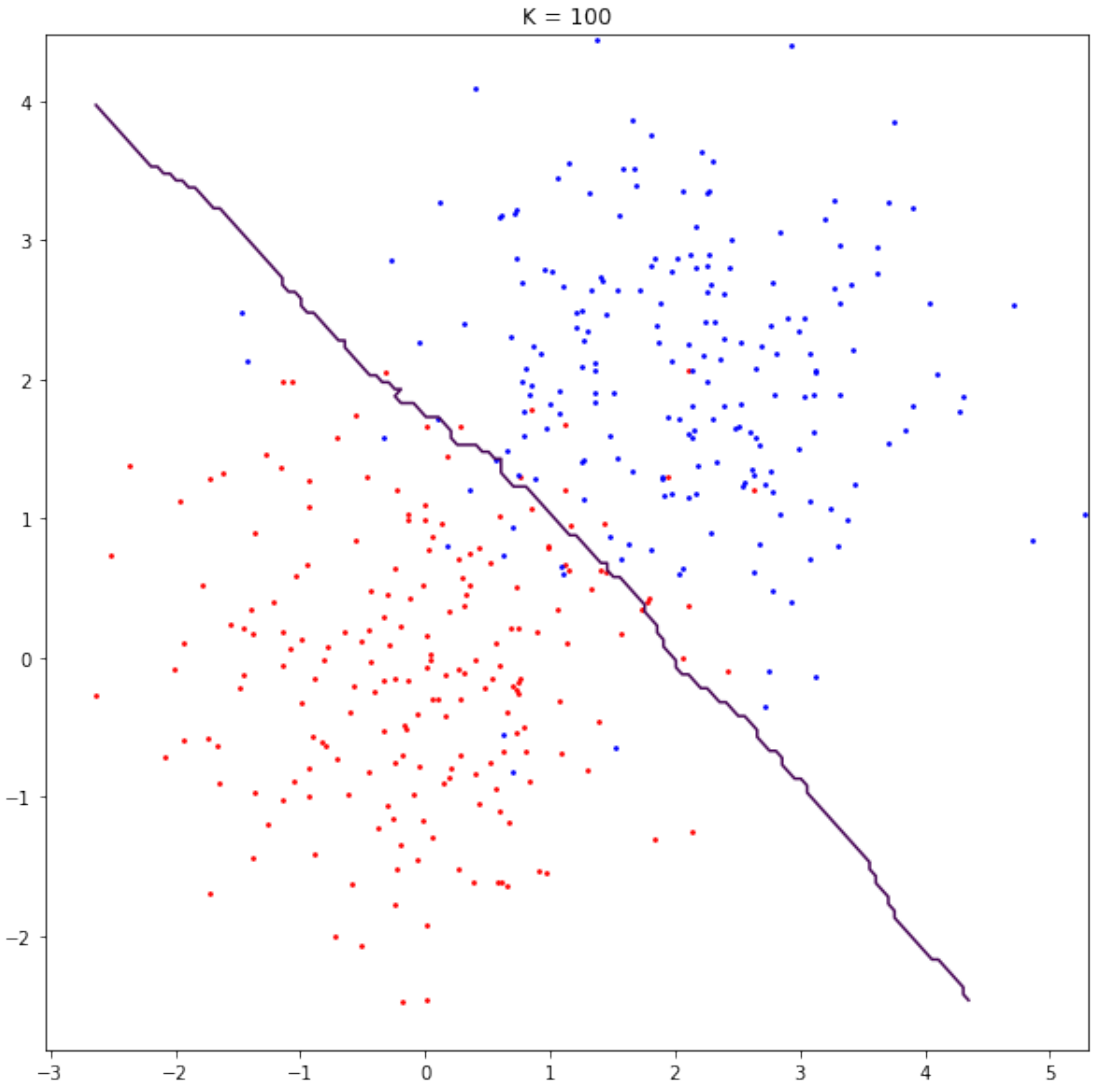

Figure 4: k=100.

(b) We have seen the effects of different choices of K. How can you choose a proper K when dealing with real-world data ?

Answer: We can use cross validation to choose a proper one.

(c) Finish *hack.m*/*hack.py* to recognize the CAPTCHA image using KNN algorithm.

Answer:

```
Out[43]: array([3., 7., 3., 6.])
```

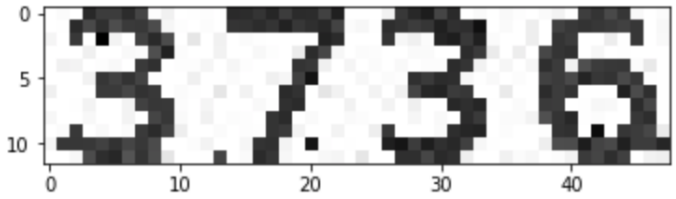

Figure 5: .

Problem 3-3. Decision Tree and ID3

Consider the scholarship evaluation problem: selecting scholarship recipients based on gender and GPA. Given the following training data:

Answer:

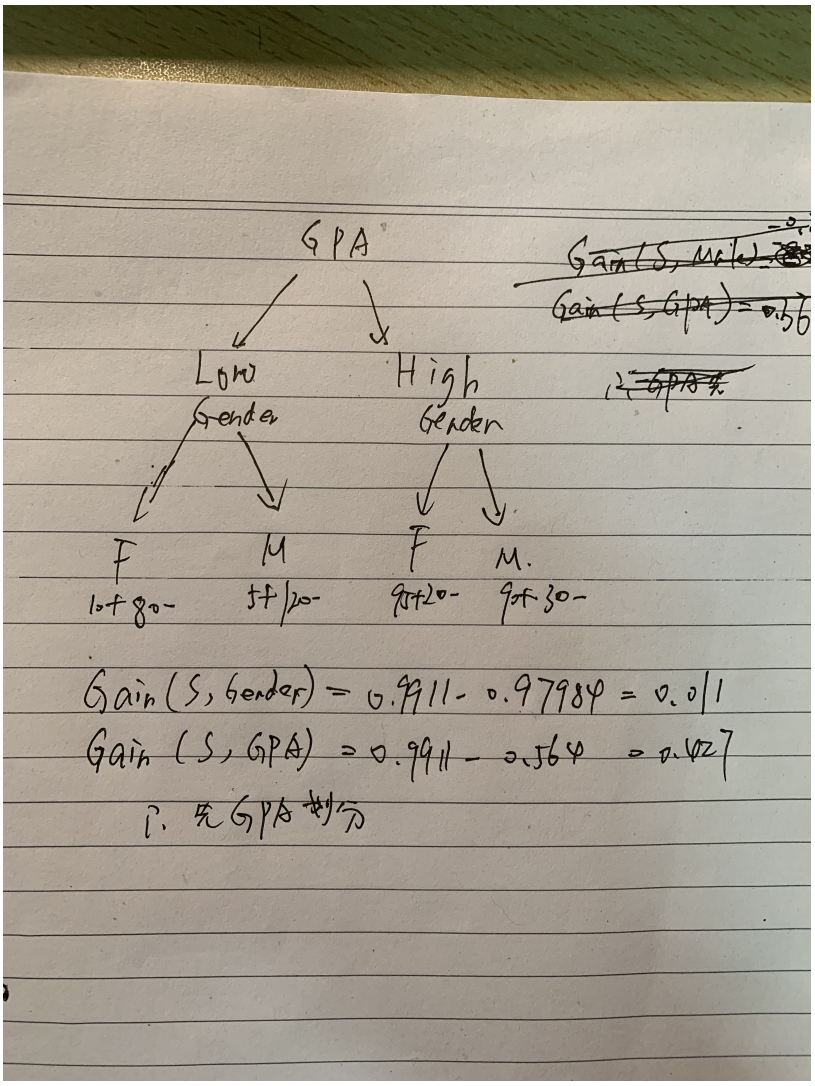

Figure 6: .

Problem 3-4. 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:

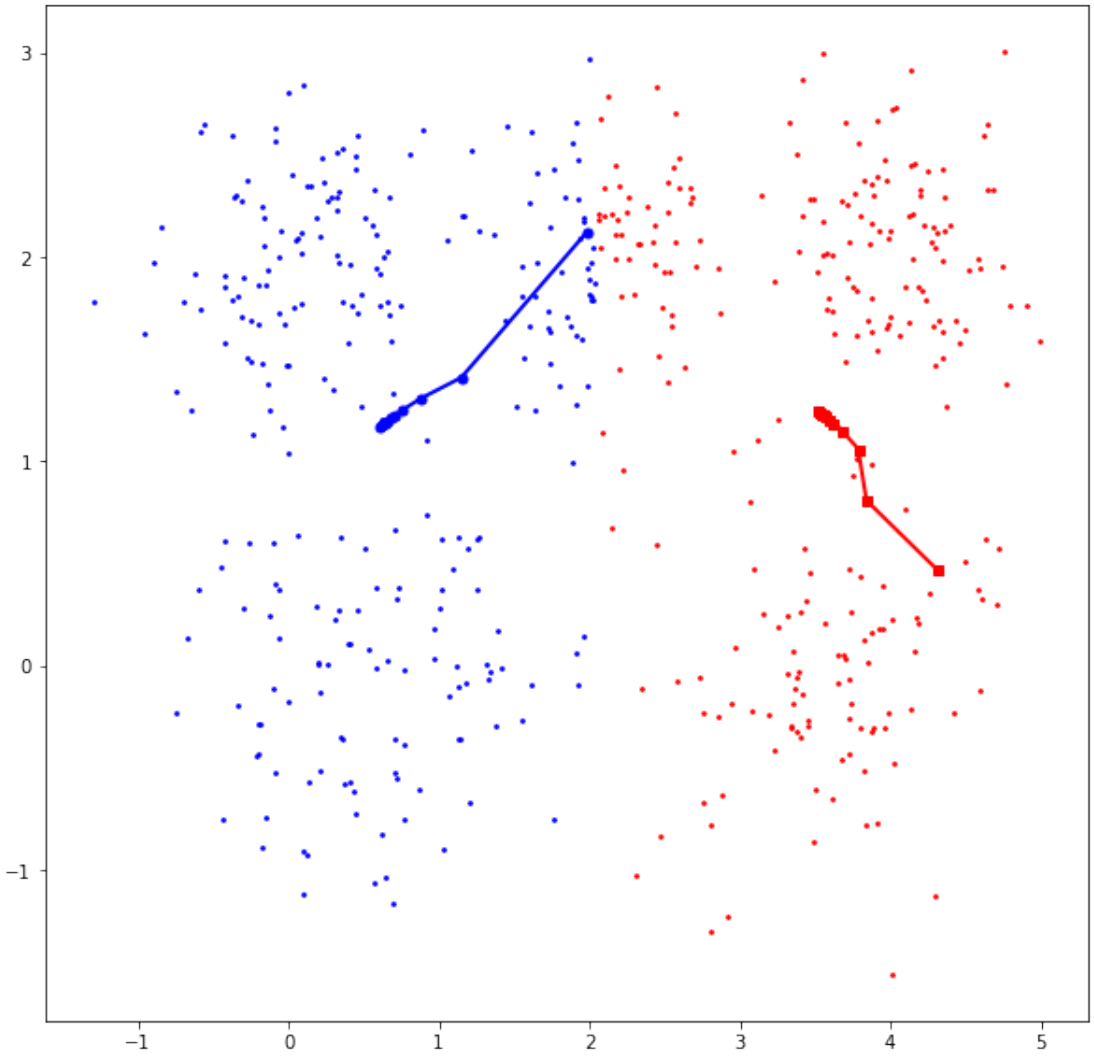

Figure 7: min SD.

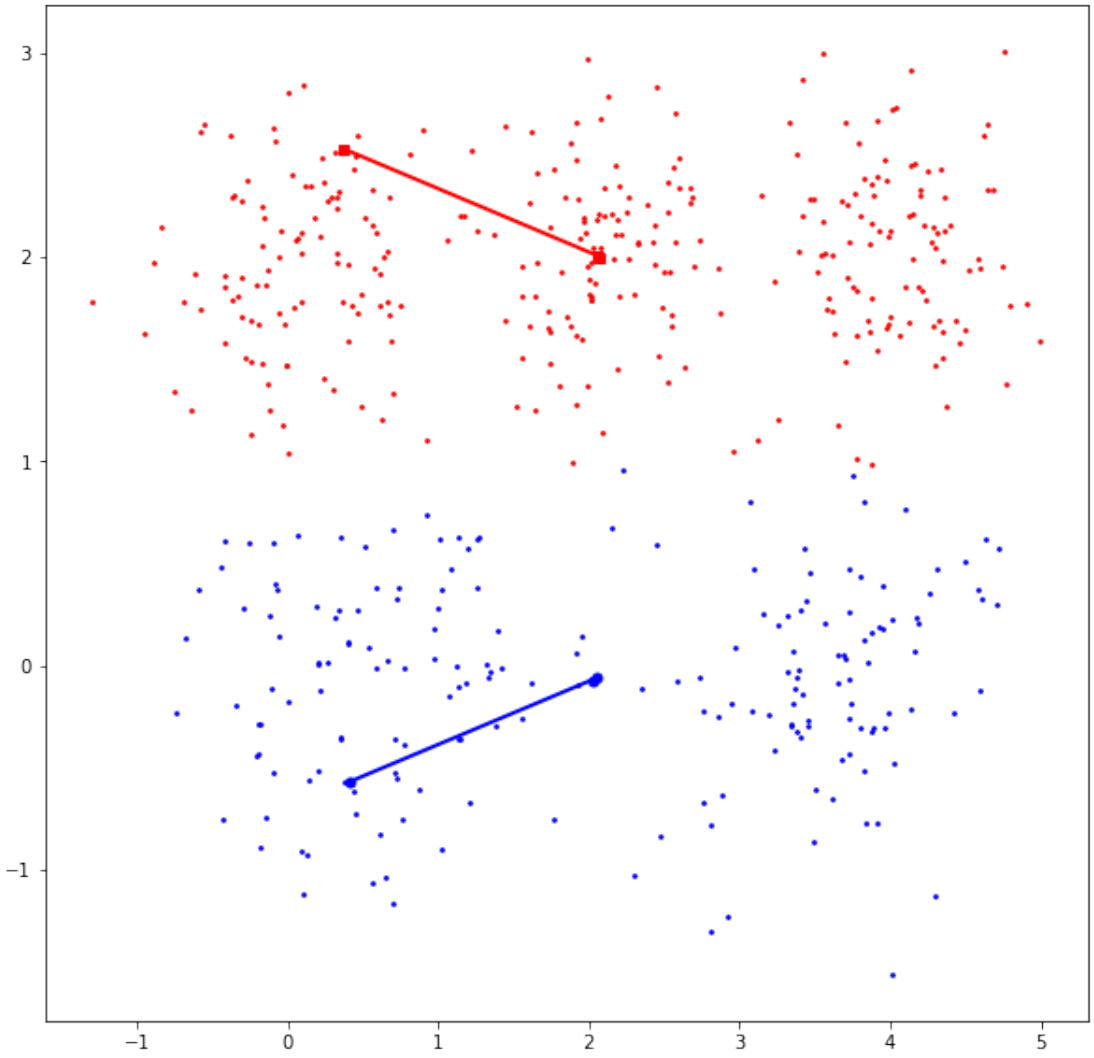

Figure 8: max SD.

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

Answer: We should choose a proper K. And then, we should run K-means many times to choose a best MSE results.

(c) Visualize the centroids.

Answer:

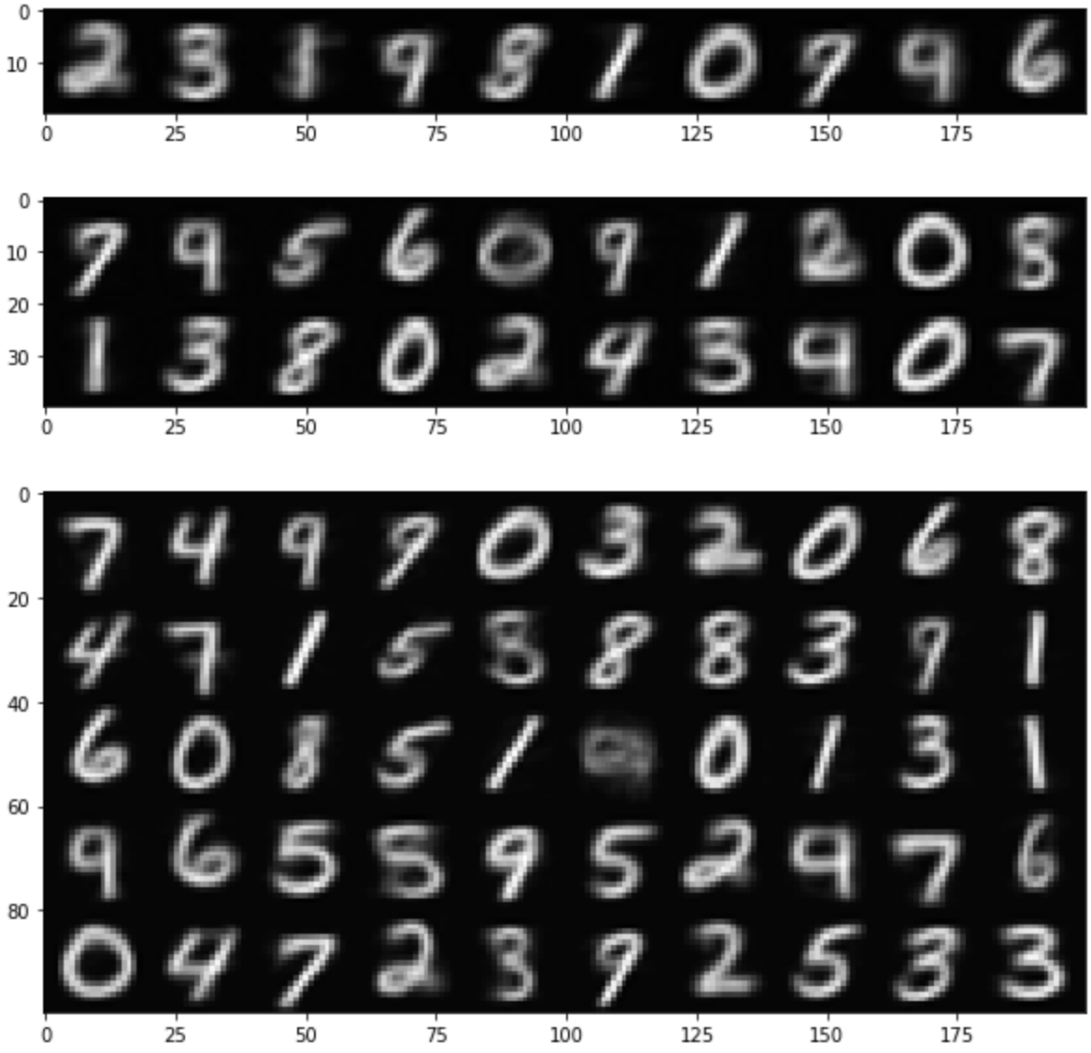

Figure 9: The results when K=10, 20, 50.

(d) Vector quantization.

Answer:

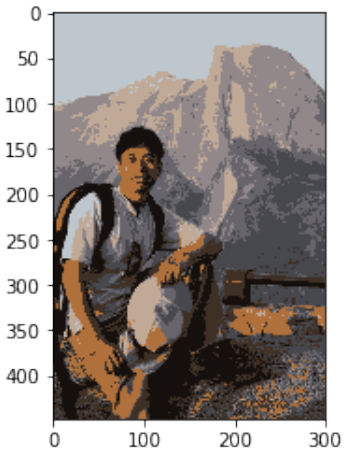

Figure 10: The results when K=8.

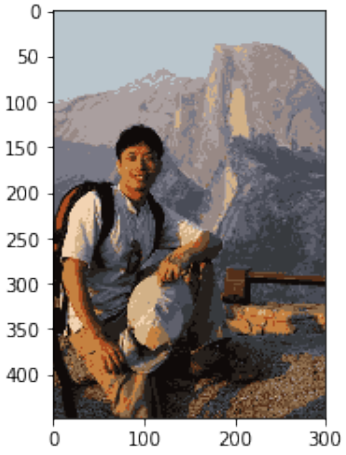

Figure 11: The results when K=16.

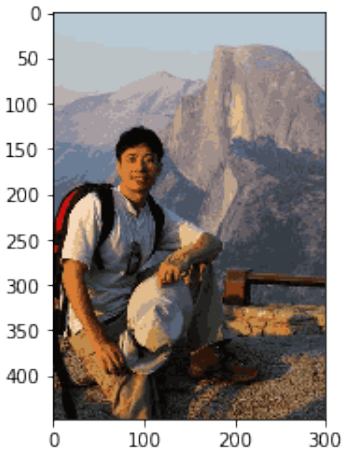

Figure 12: The results when K=32.

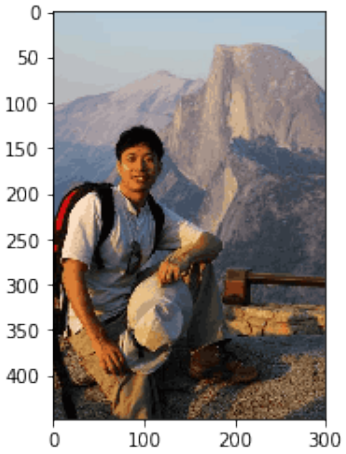

Figure 13: The results when K=64.

$compressratio = log_2^K/24 = 25\%$