Homework 3

Collaborators:
Name:
Student ID: 3160104210

Problem 3-1. Neural Networks

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

(a) Answer: Test accuracy: 92%

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:

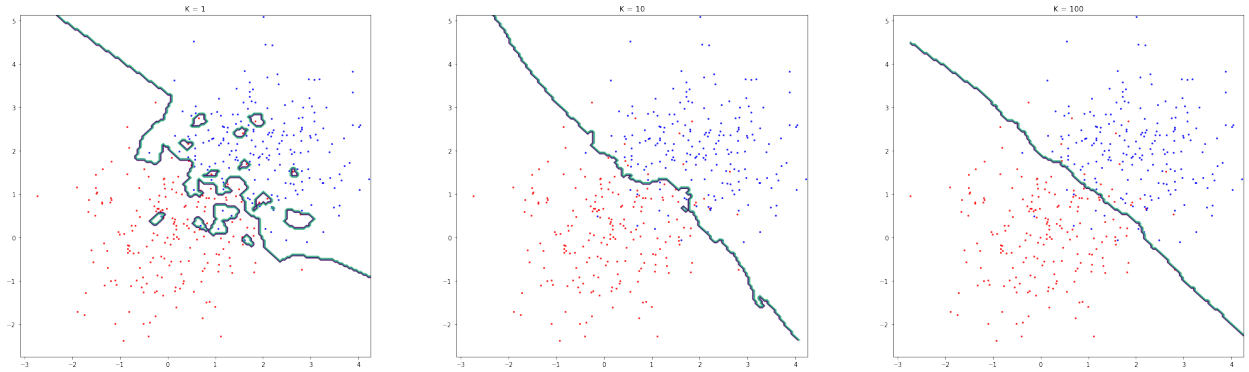

Figure 1: K = 1, 10, 100

(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: Finding the value of k is not easy. A small value of k means that noise will have a higher influence on the result and a large value make it computationally expensive. we can divide the entire data set into training set and test set. Apply the KNN algorithm into training set and cross validate it with test set.

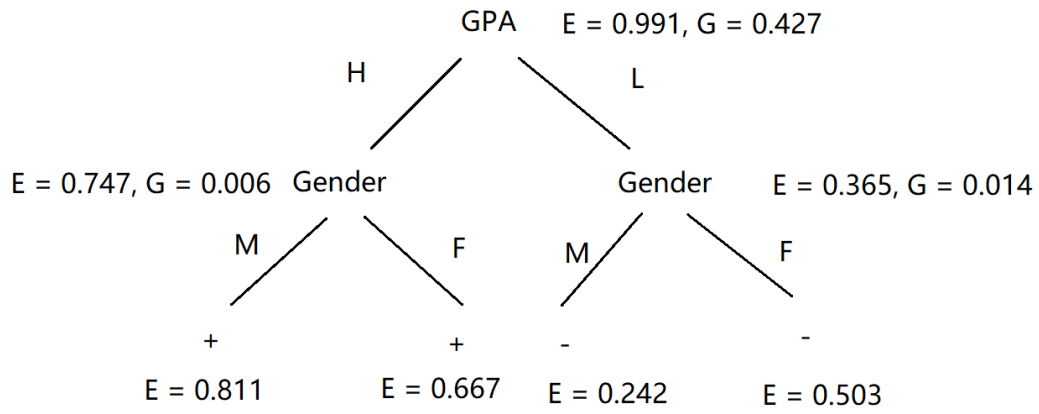

Figure 2: decision tree

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

Answer: First, i use scrawer.py to download images from the website.

Second, i use label.py to label the images one by one.

Third, i use generate_ npz.py to read both data and labels from file and store them as a dictionary in file hack_ data.npz.

Last, i use hack.py to get the six digits of testing image. I set K = 10 in hack.py.

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: See figure 2.

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: See figure 3.

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

Answer:

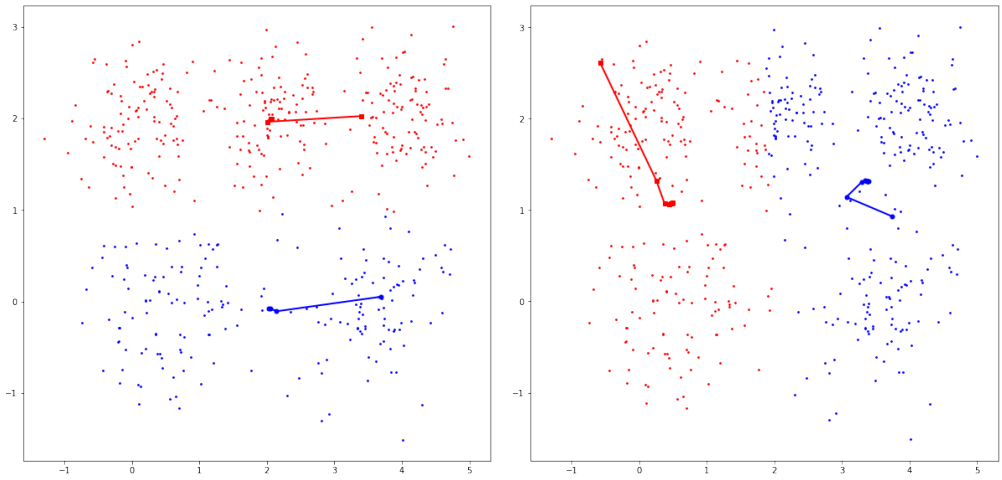

Figure 3: max_ SD, min_ SD

1 Repeat K-Means Clustering for many times, then choose the one with lowest SD.

2. Use validation set to find a proper K.

(c) Visualize the centroids.

Answer:

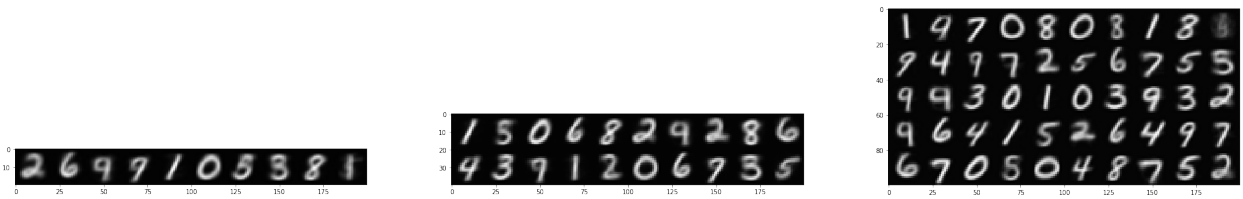

Figure 4: K = 10, 20, 50

(d) Vector quantization.

Answer:

If K is set to 64, then every pixel can be represented as log64 = 6 bits, which usually requires 24 bits, so compress ratio is 6 / 24 = 25%.

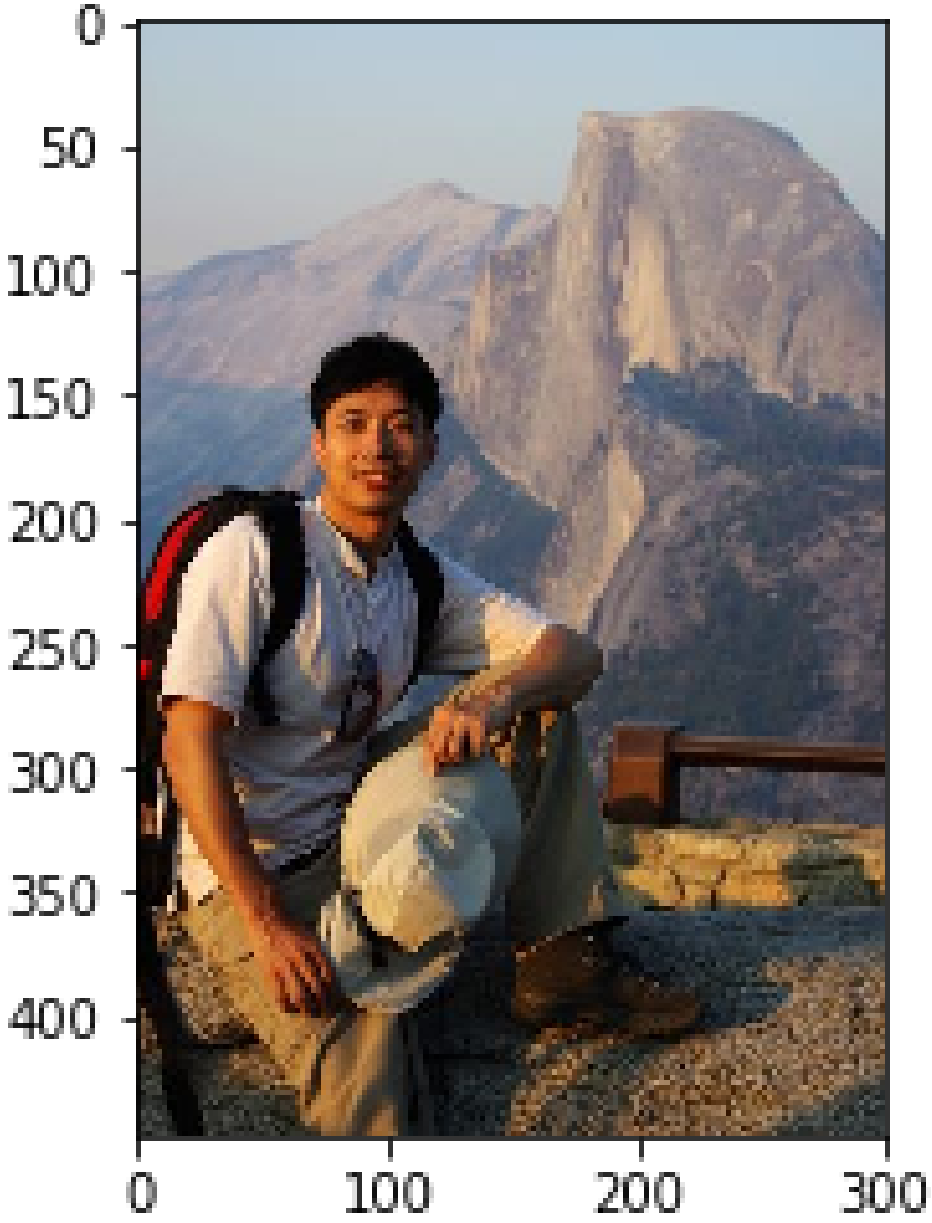

Figure 5: origin image

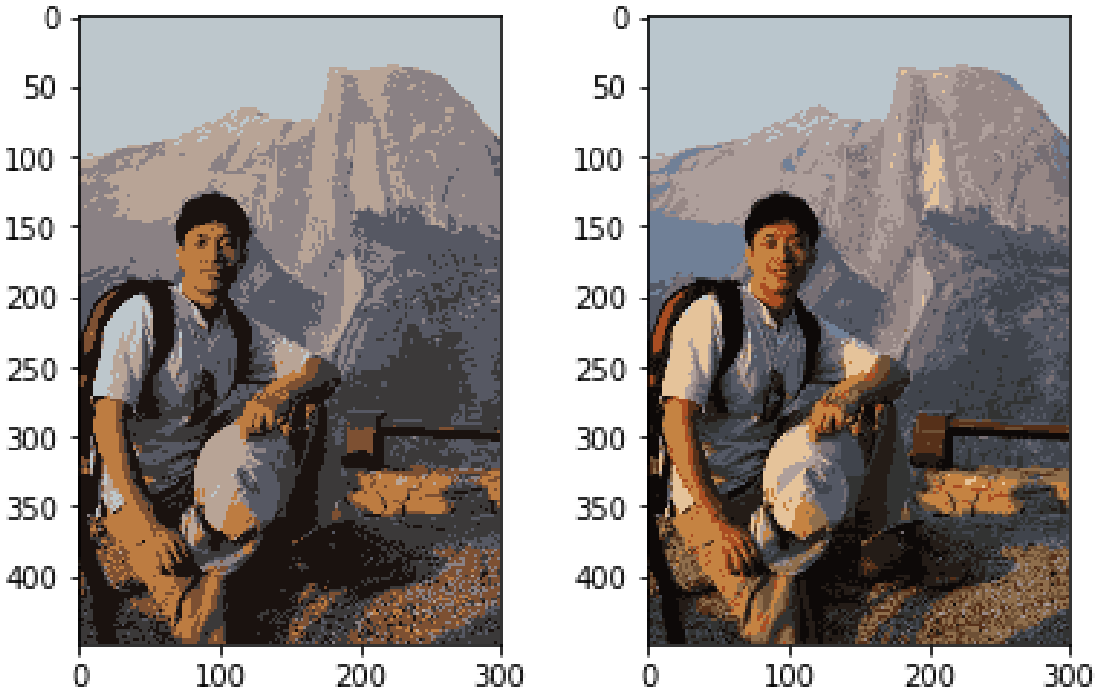

Figure 6: K = 8, 16

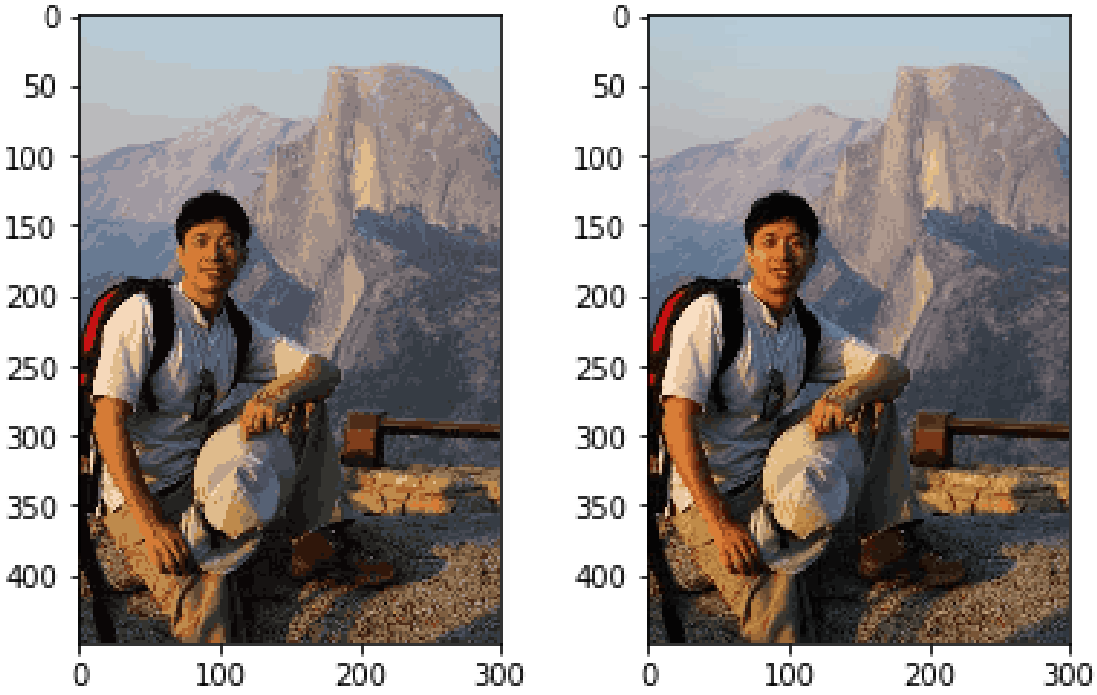

Figure 7: K = 32, 64