Homework 3

Collaborators:

Name: Wang Yanwei

Student ID: 11821049

Problem 3-1. Neural Networks

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

(a) Answer: Test accuracy is 0.924.

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:

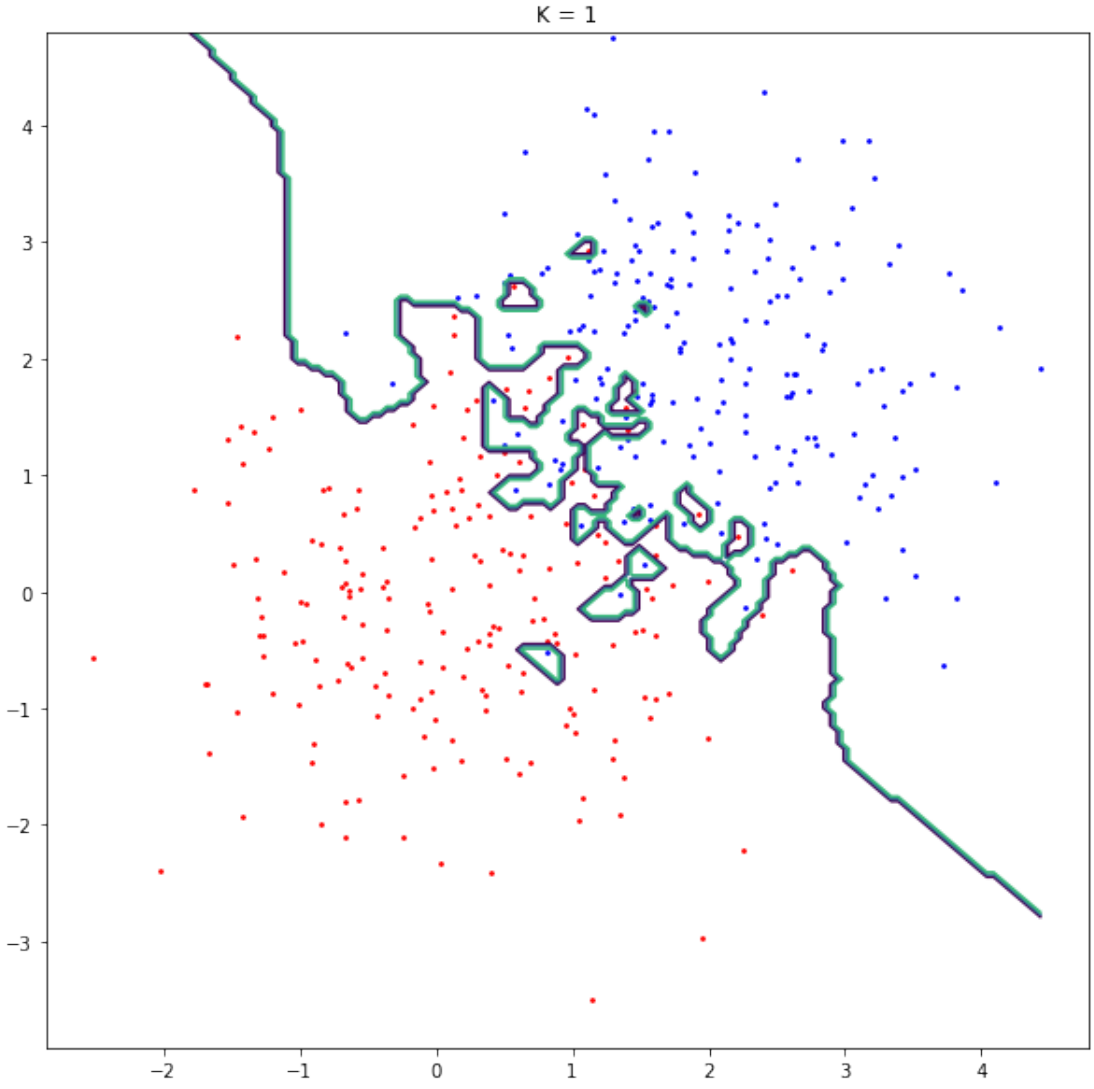

Figure 1: K = 1.

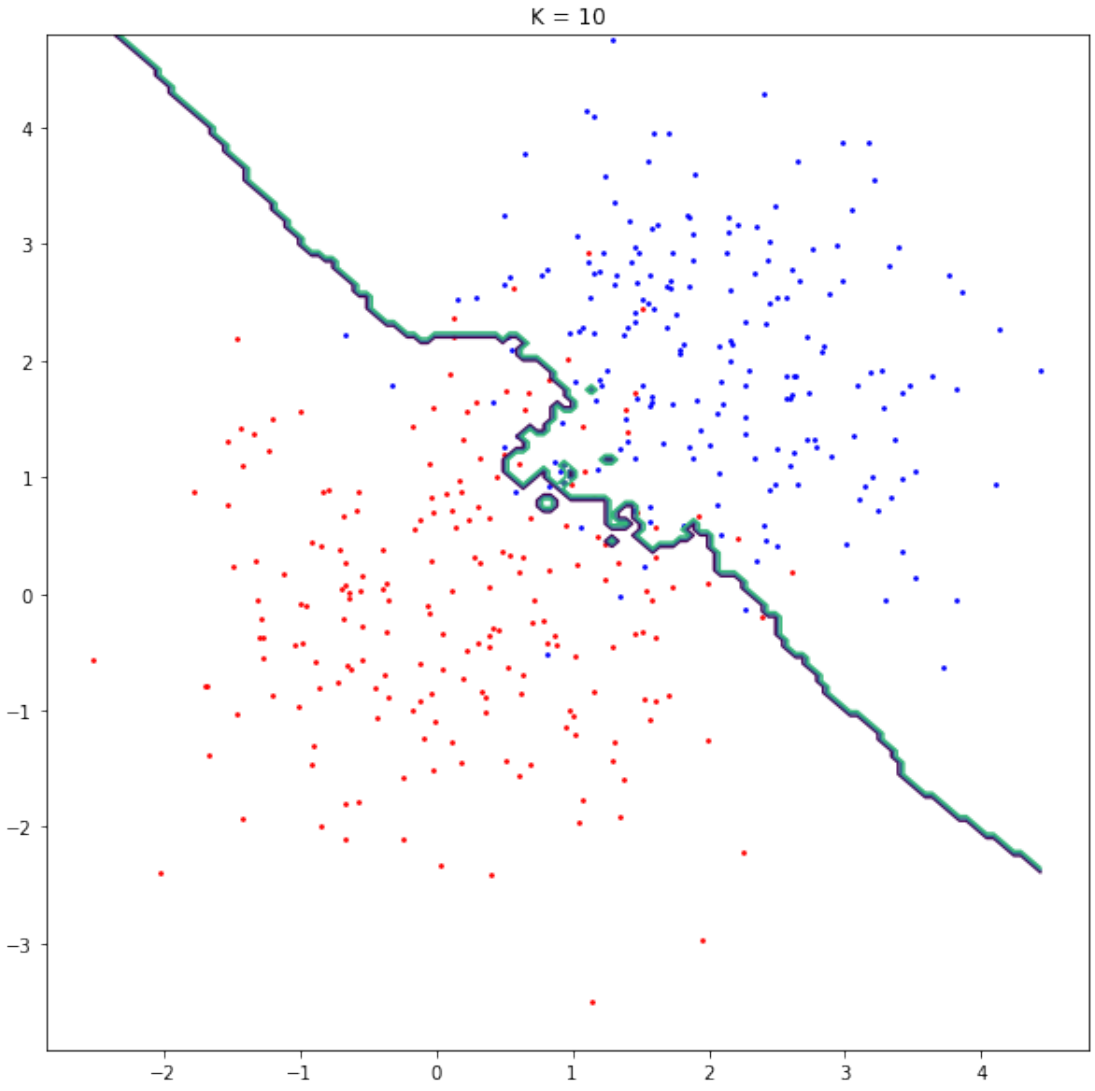

Figure 2: K = 10.

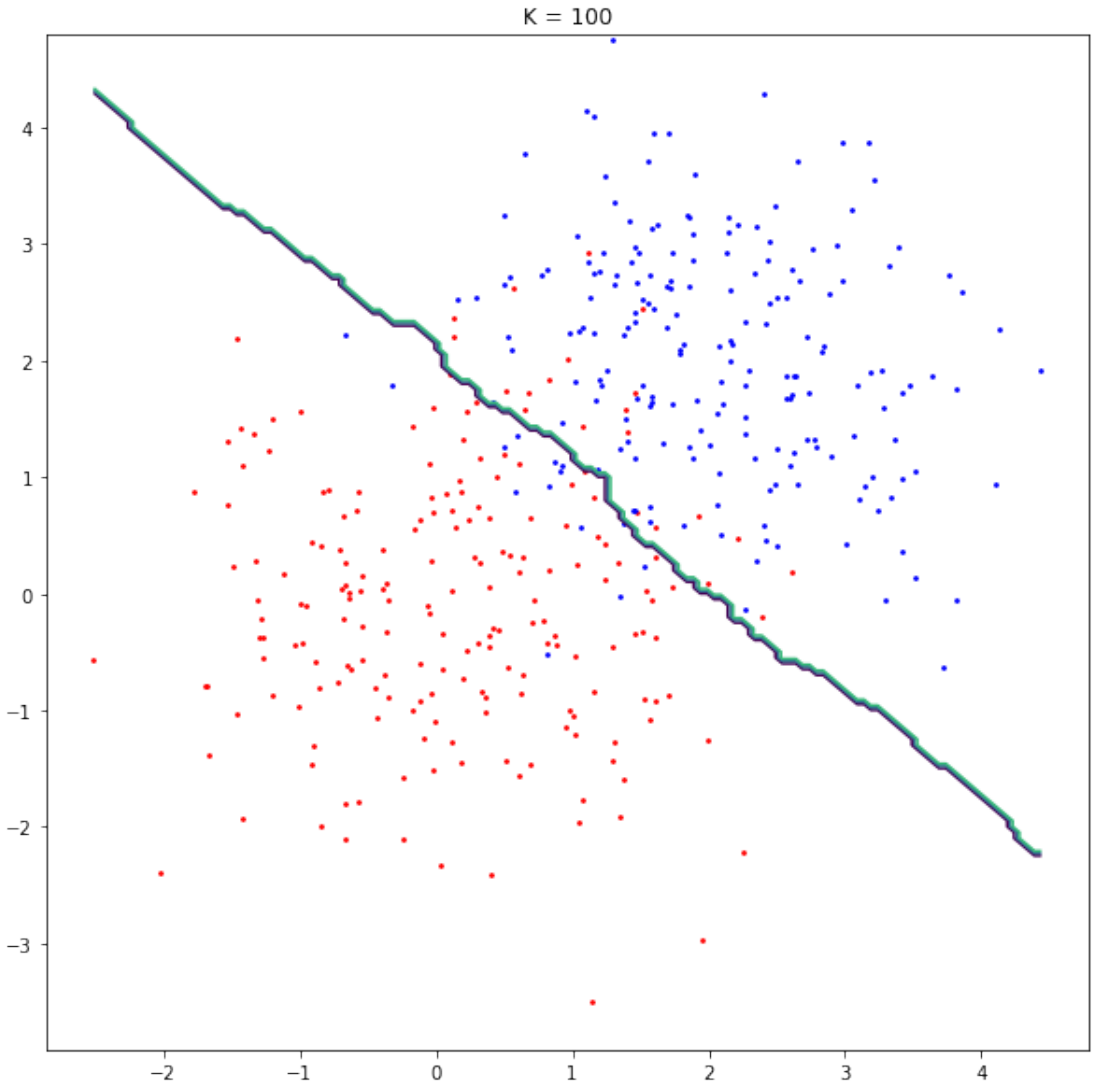

Figure 3: K = 100.

(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: Use dimension reduction method to view data in low dimension, choose a proper K according to the data struct.

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

Answer: See the code.

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:

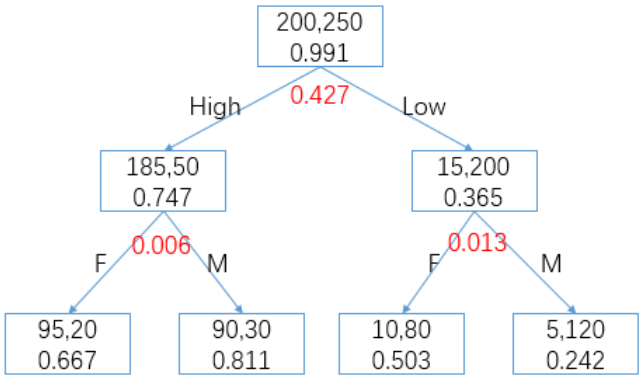

Figure 4: Desion Tree.

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:

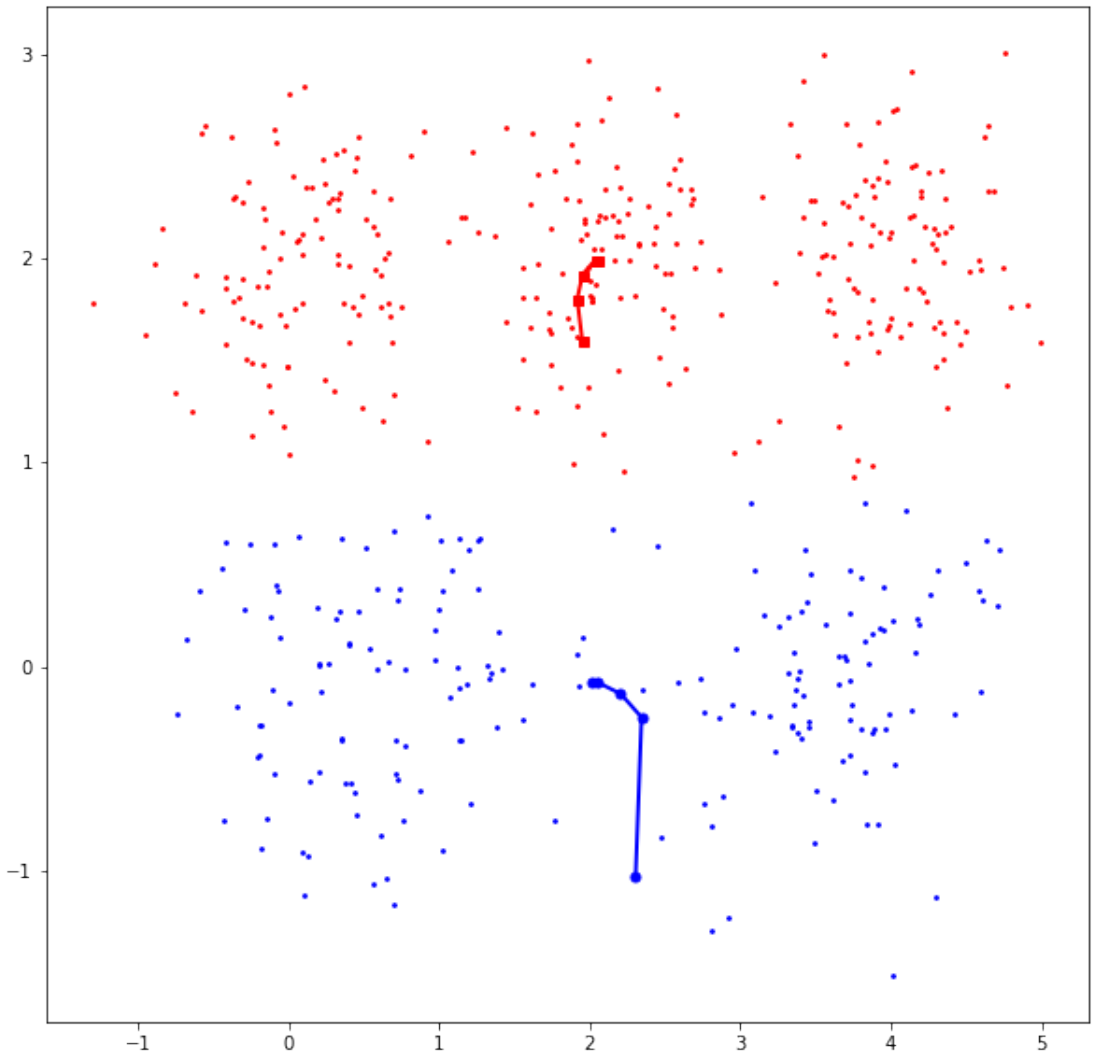

Figure 5: Kmeans Min SD.

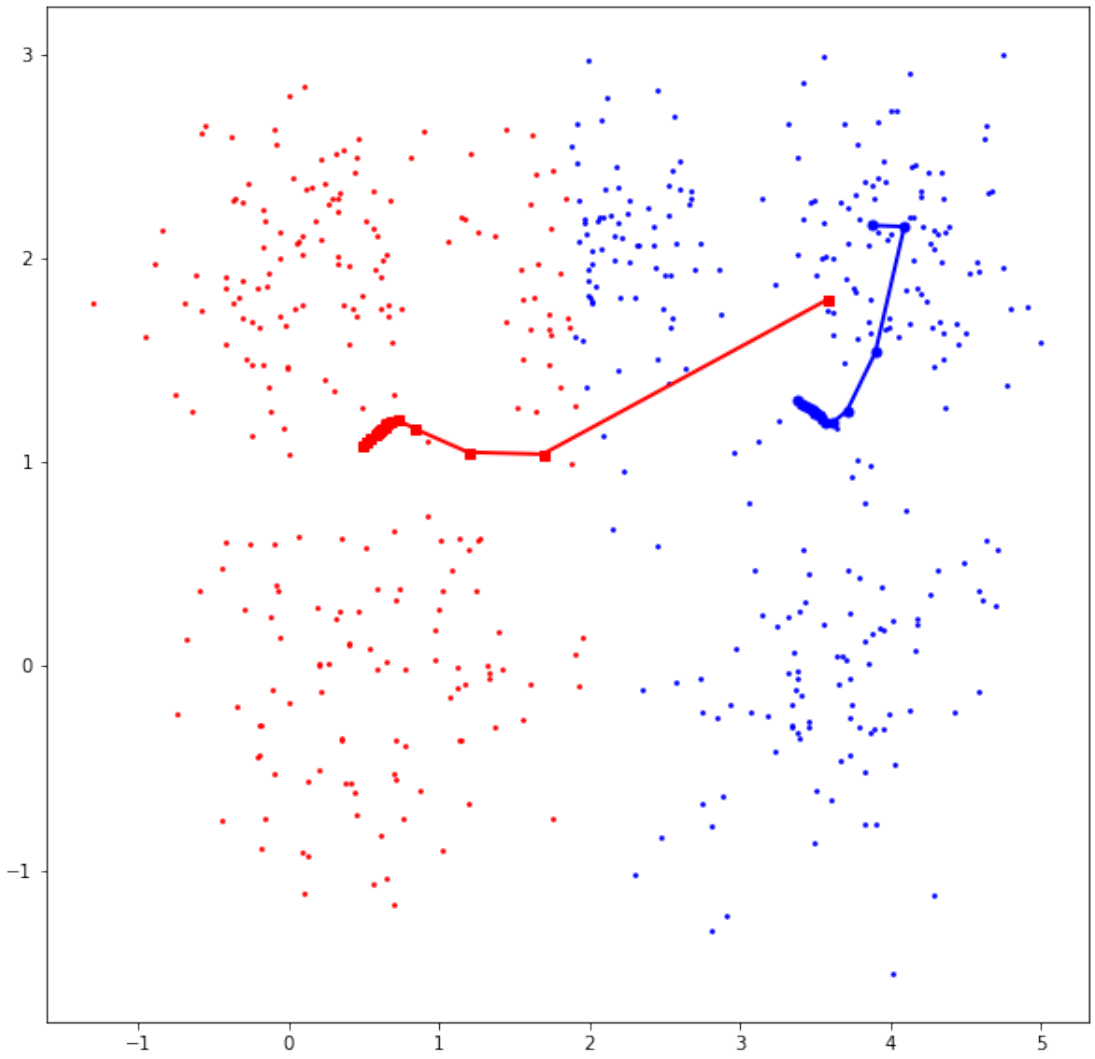

Figure 6: Kmeans Max SD.

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

Answer: Run code several times, choose the best model which has the smallest SD.

(c) Visualize the centroids.

Answer:

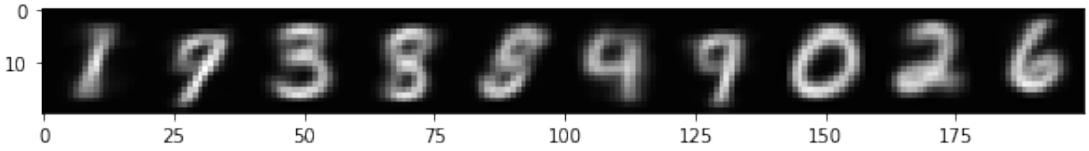

Figure 7: K = 10.

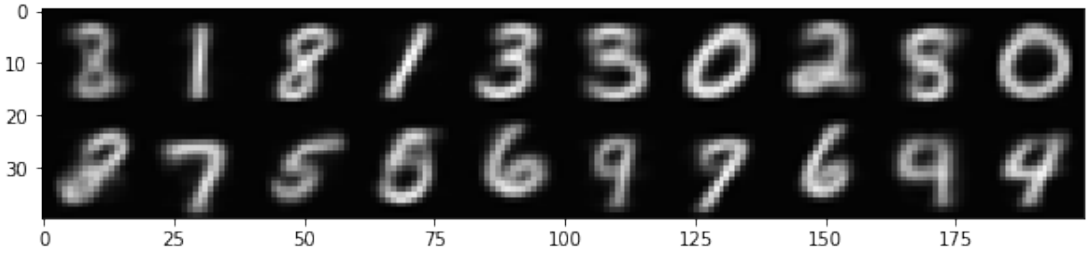

Figure 8: K = 20.

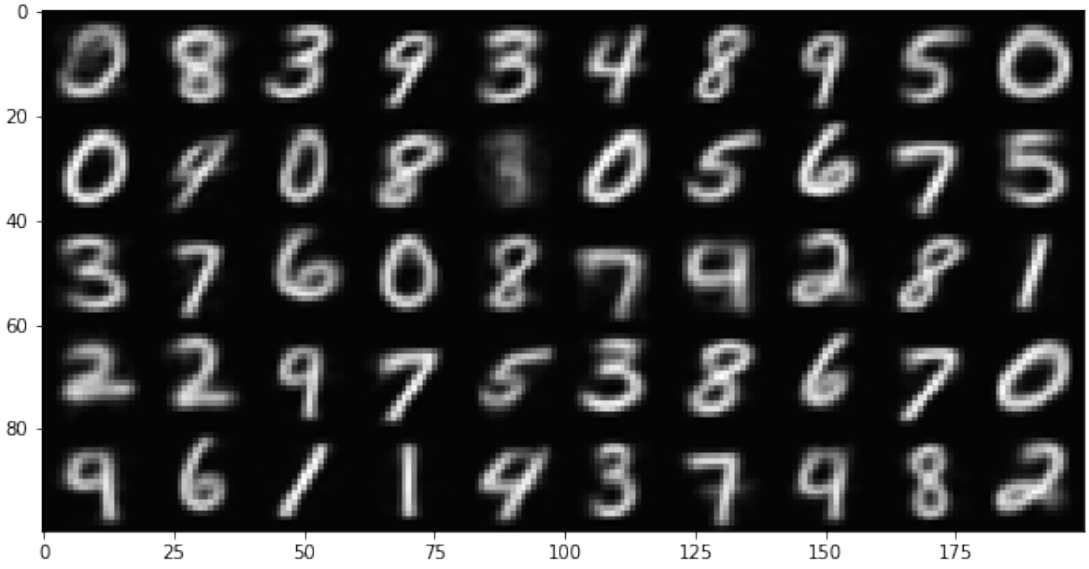

Figure 9: K = 50.

(d) Vector quantization.

Answer:

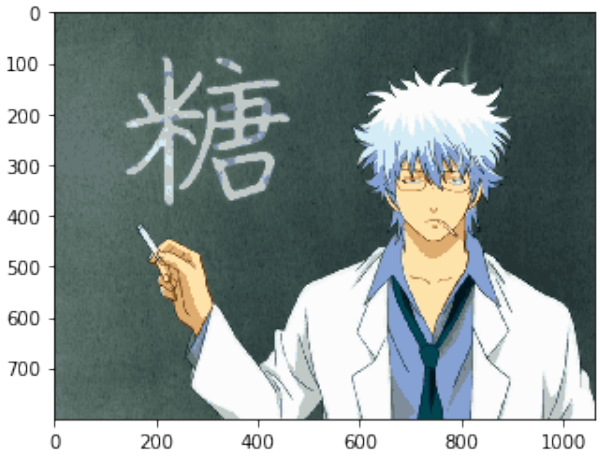

Figure 10: vector quantization.

The compress ration is 0.75 , I ignore 64 $\times$24 bits data, beacuse it is rather small to the $H \times W \times 24$ image data.