# Machine Learning: Assignment #3

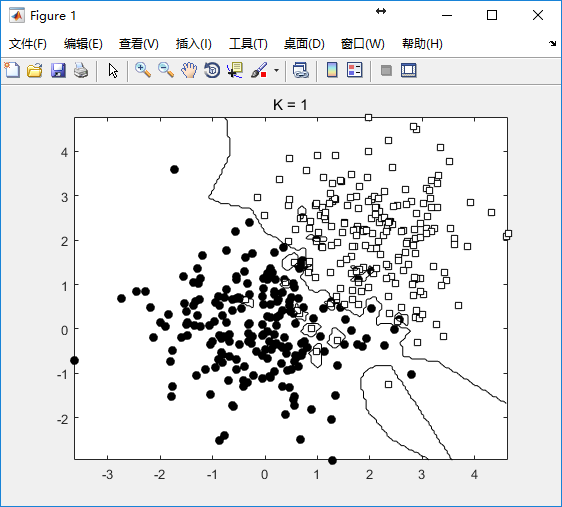
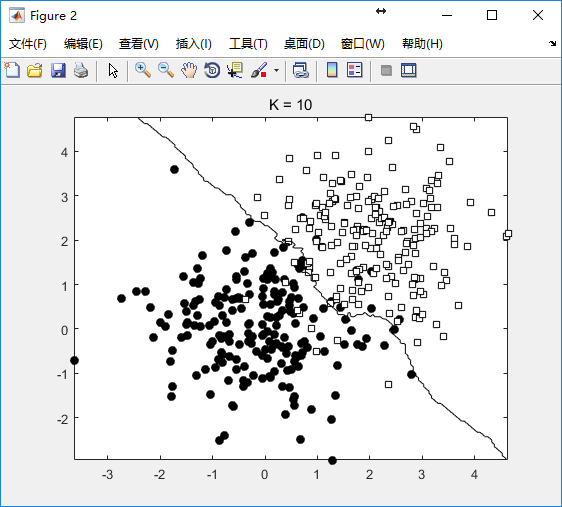
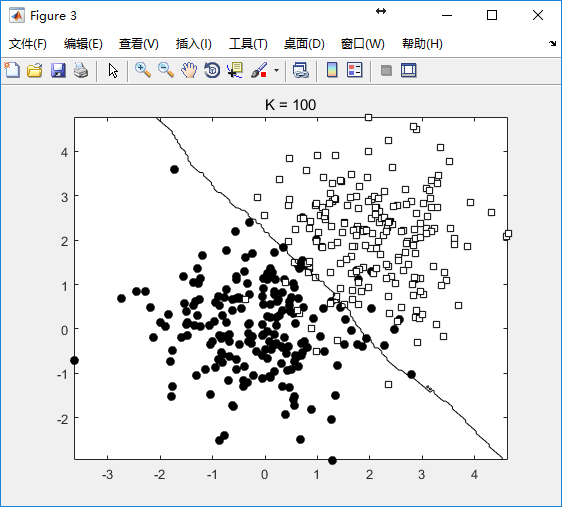
# 1. Neural Networks

The code is in zip, we can train three layer neural networks by my code. And my test accuracy is 92.700%

# K-Nearest Neighbor

(a)

I have completed the knn.m, and the decision boundary of K = 1, 10, and 100 is as follow:

(b)

The parameter K is difficult to choose, on real-world data, we can use Cross-Validation to choose proper K.

(c)

Firstly, we should get the training set. I have write a script to get the CAPTCHA images. I have put the script to github, so everybody can use my code to get the CAPTCHA images.

The address of code is <https://github.com/yysys/getDataFromURL>

Then we tag the label of training set, and complete hack.m to recognize the CAPTCHA image using KNN algorithm.

# Decision Tree and ID3

There are 540 samples, including 200 positive samples and 250 negative samples. So the entropy(S) is as follow:

Entropy(S) =  = 0.9911

记属性gender为T1，属性GPA为T2



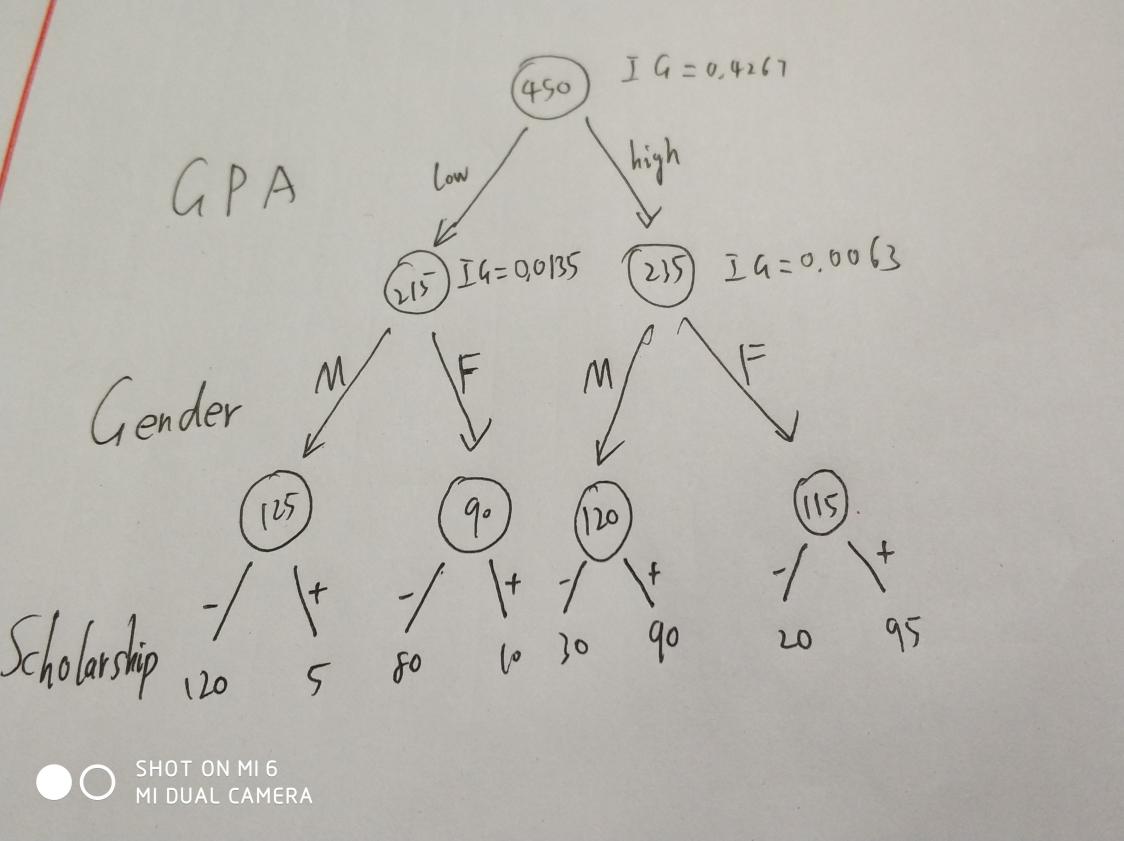






So we choose the GPA to divide set, after that we choose the gender to divide the set.

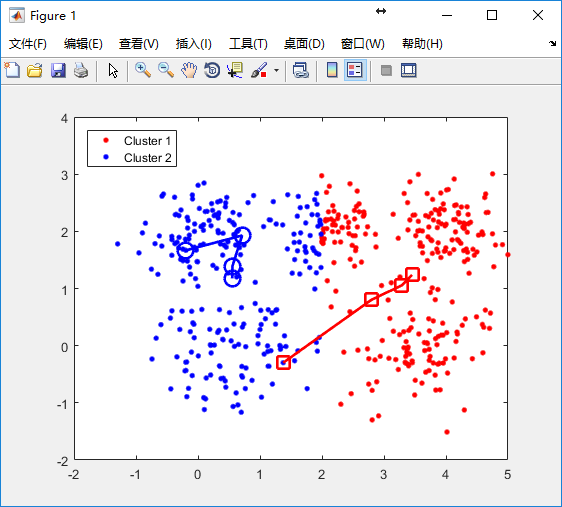
The decision tree is drew as follow:



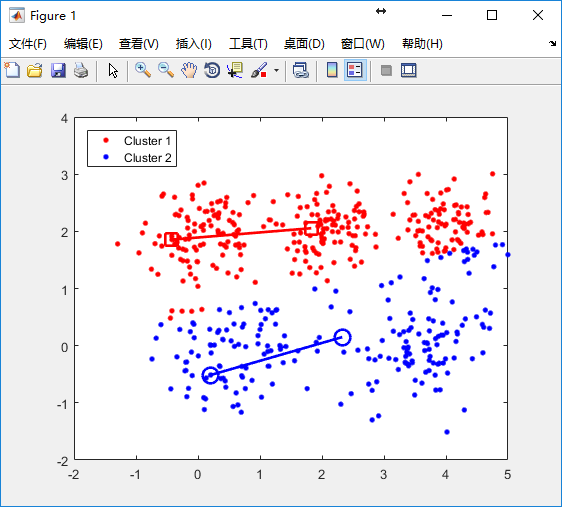
# 4. K-Means Clustering

(a)

The process of k-means algorithm for the two trials with smallest SD is as follow:



The process of k-means algorithm for the two trials with largest SD is as follow:



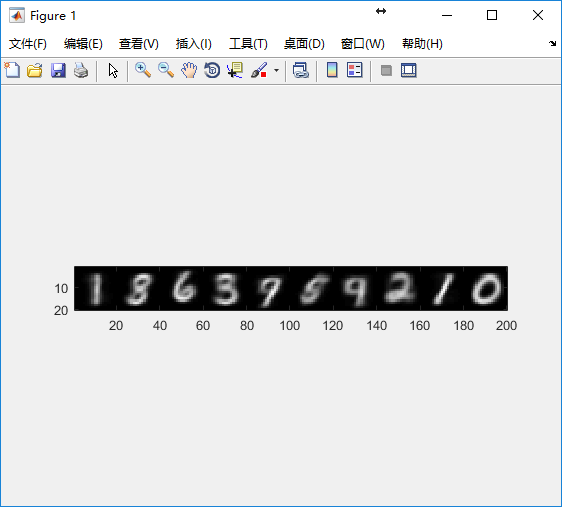
(b)

We can use k-means++ instead of k-means to reduce the influence of cluster centroids initialization.

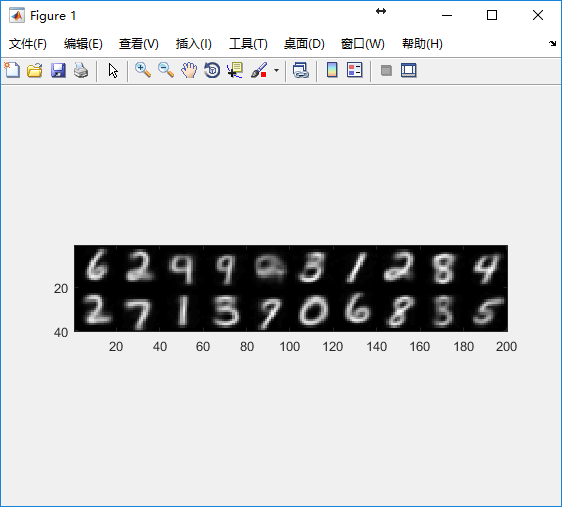
(c)

K-means algorithm can discover the patterns in dataset without any label information.

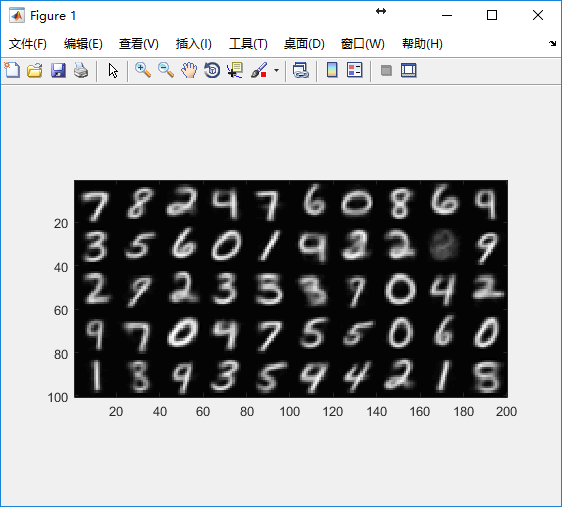
When K = 10, the picture is as follow:



When K = 20, the picture is as follow:



When K = 50, the picture is as follow:



(d)

We use k-means to vector quantization the three picture which is gave by TA.

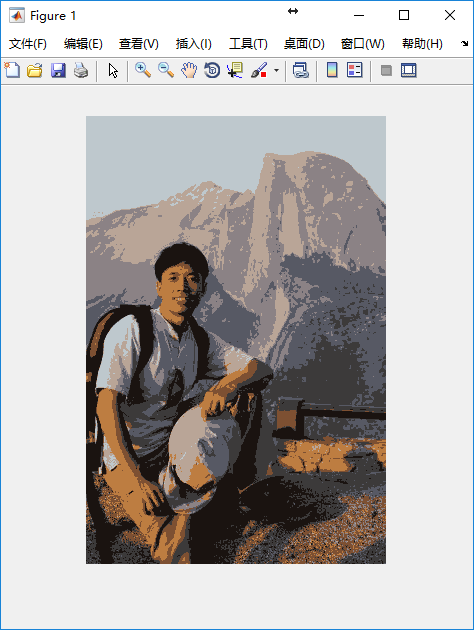
In “sample0.jpg”, the file size before vector quantization is 80619. After vector quantization , the file size is 26565. The compress ratio of “sample1.jpg” if we set K to 64 is 32.95%

In “sample1.jpg”, the file size before vector quantization is 481536. After vector quantization , the file size is 92627. The compress ratio of “sample1.jpg” if we set K to 64 is 19.23%

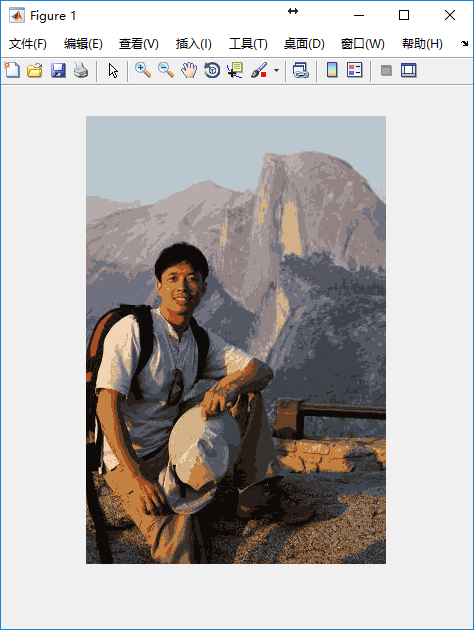
In “sample2.jpg”, the file size before vector quantization is 2760774. After vector quantization , the file size is 361478. The compress ratio of “sample1.jpg” if we set K to 64 is 13.09%

The picture when K is 8, 16, 32, 64 is as follow:

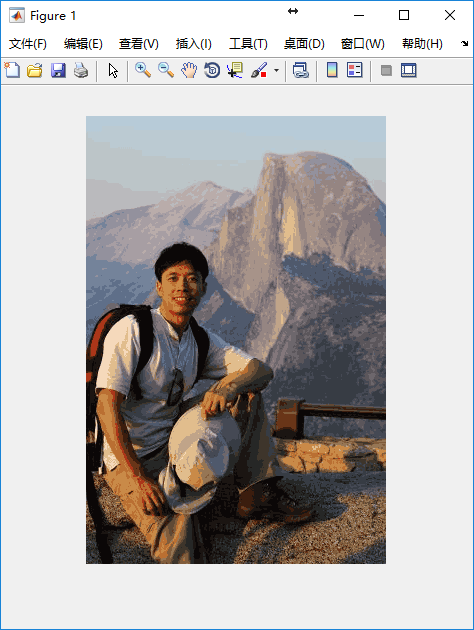
K = 8:



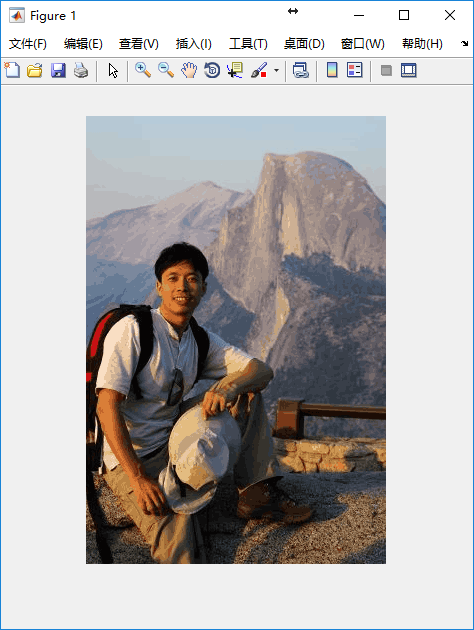
K = 16



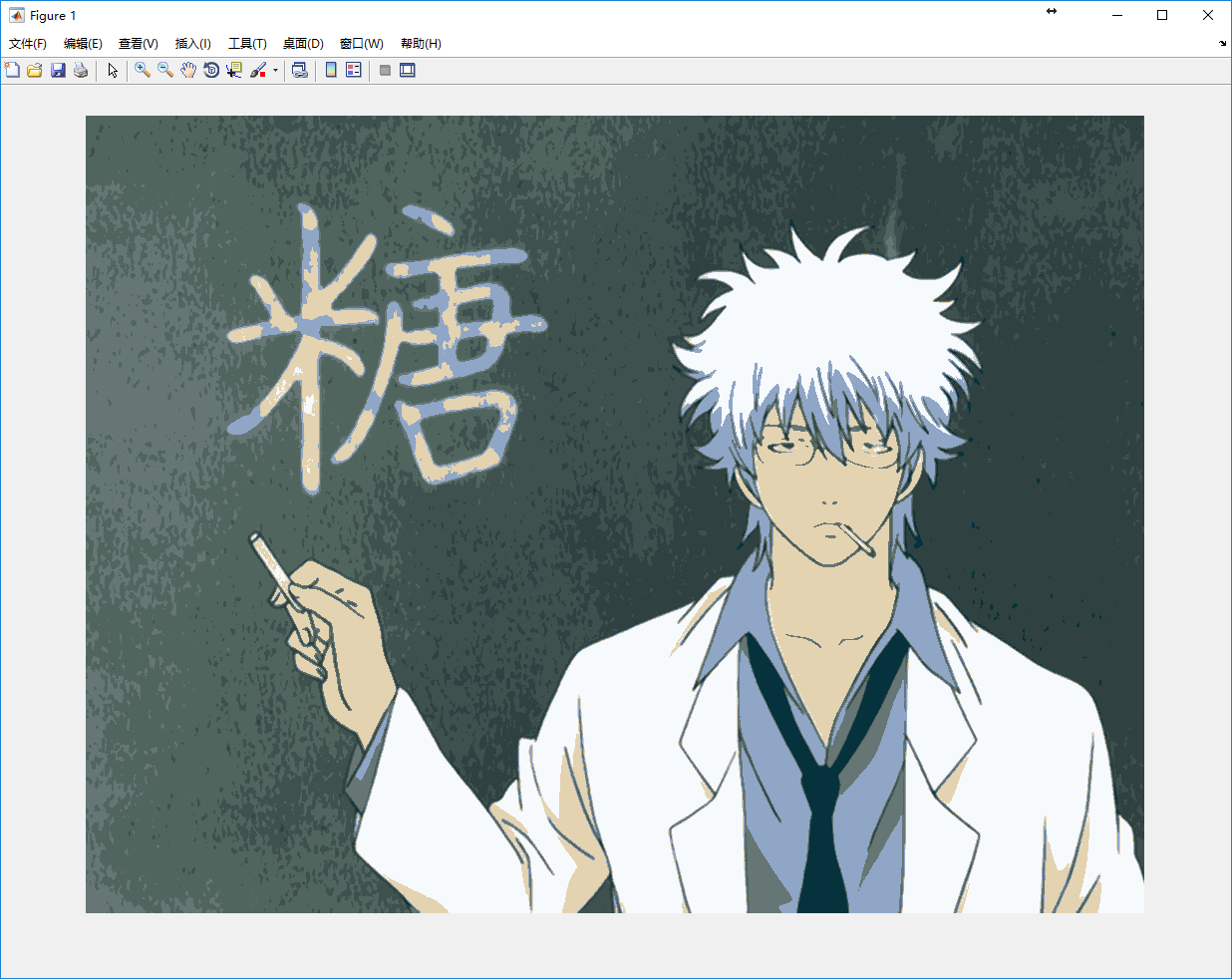
K = 32



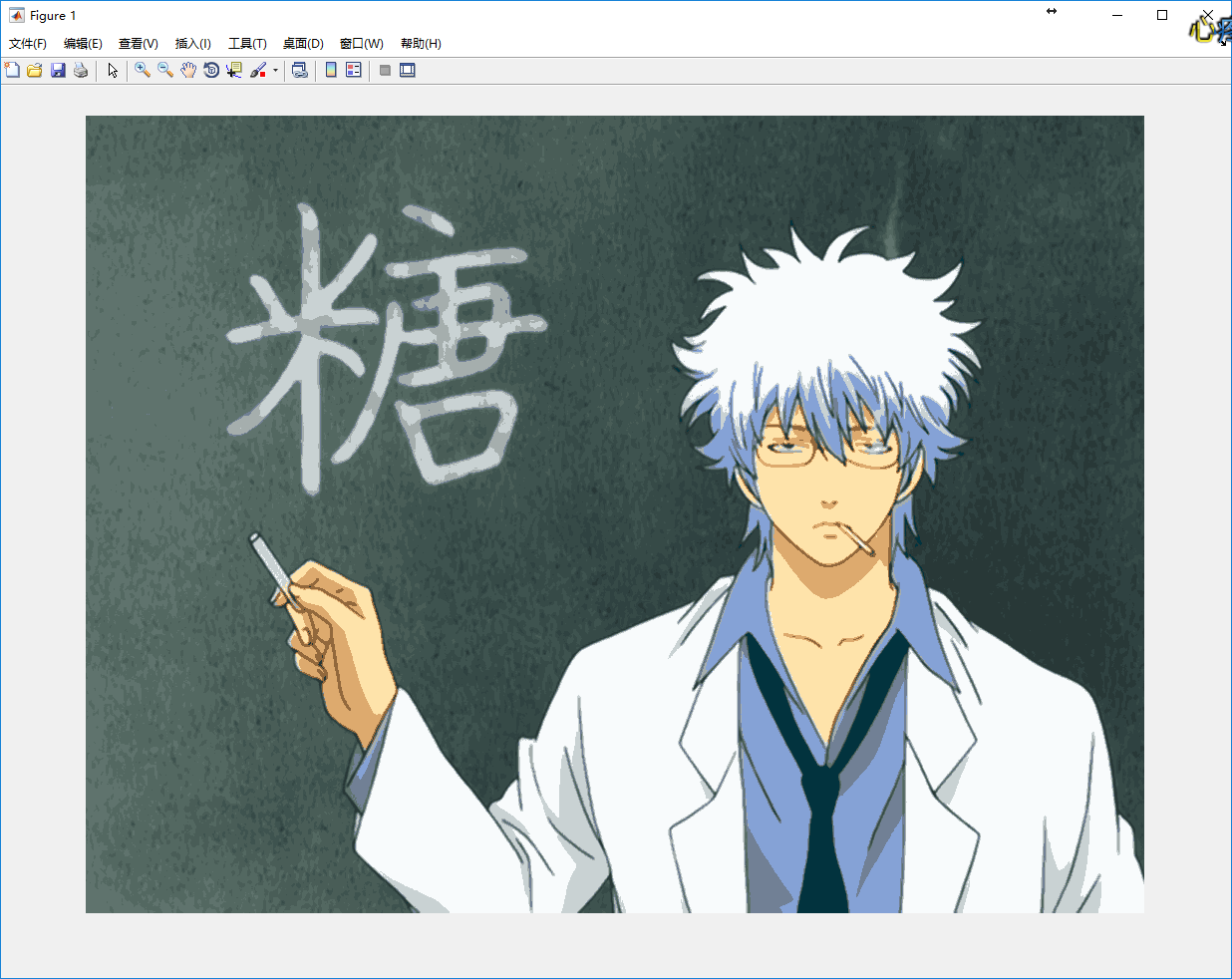
K = 64



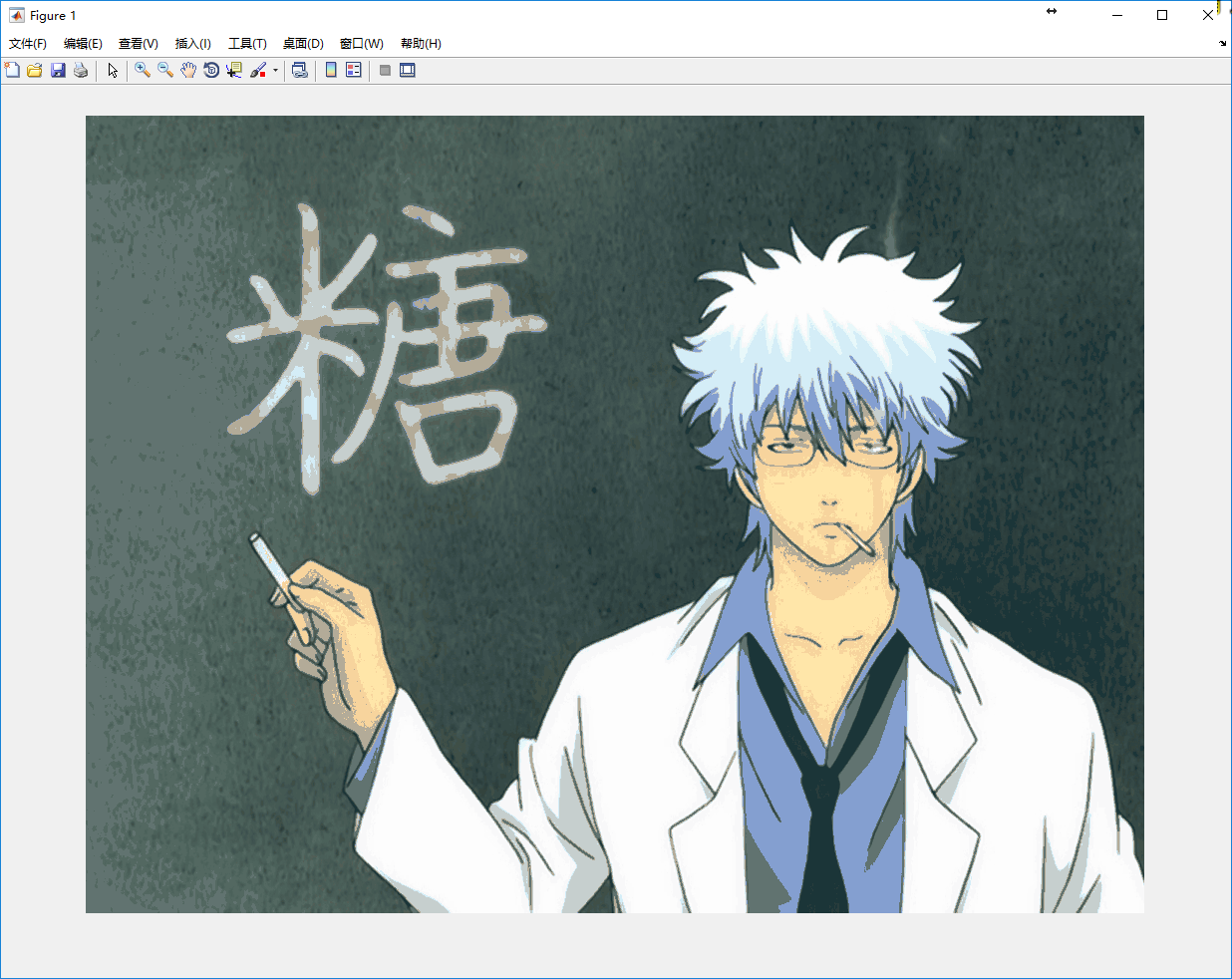
K = 8



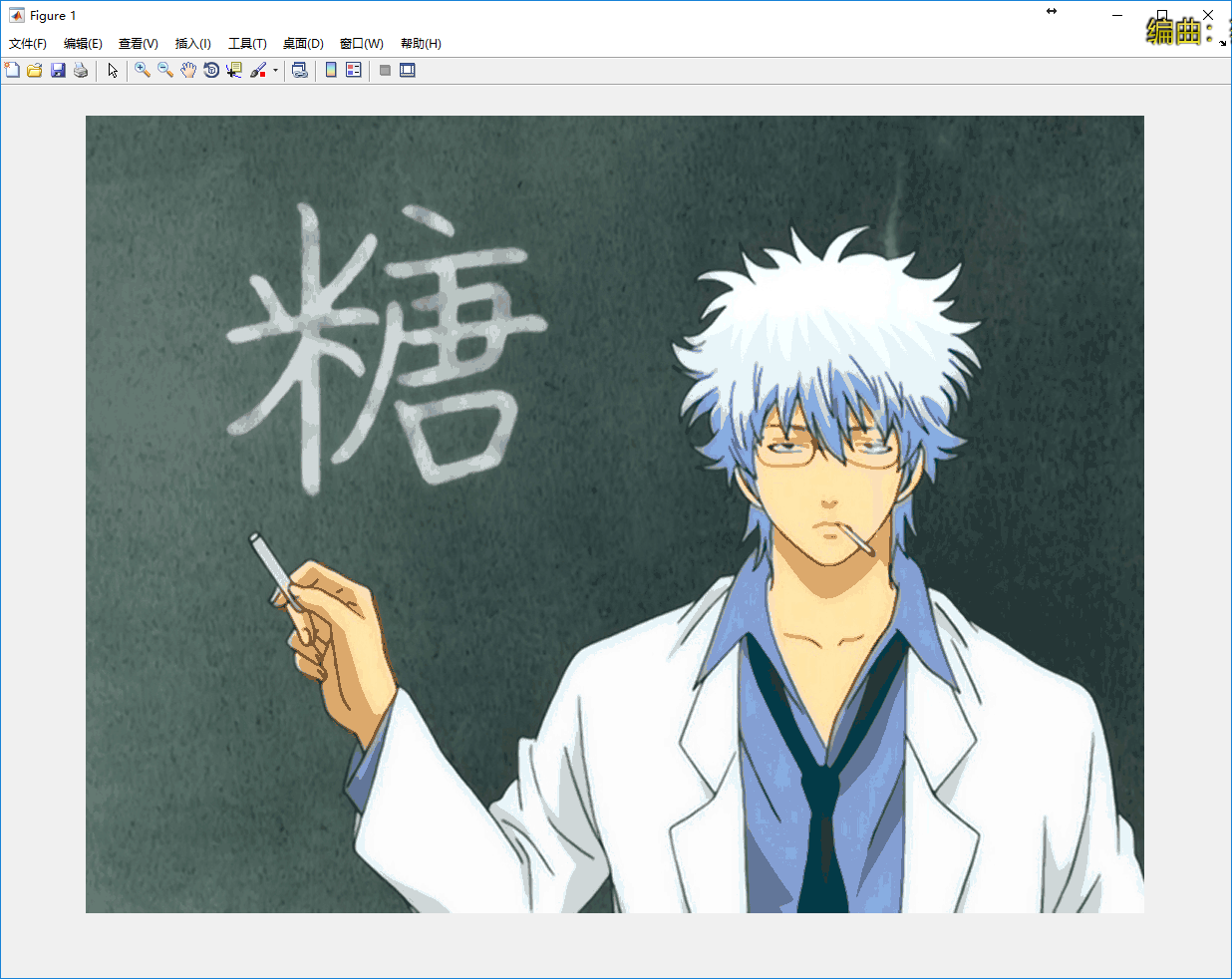
K = 16



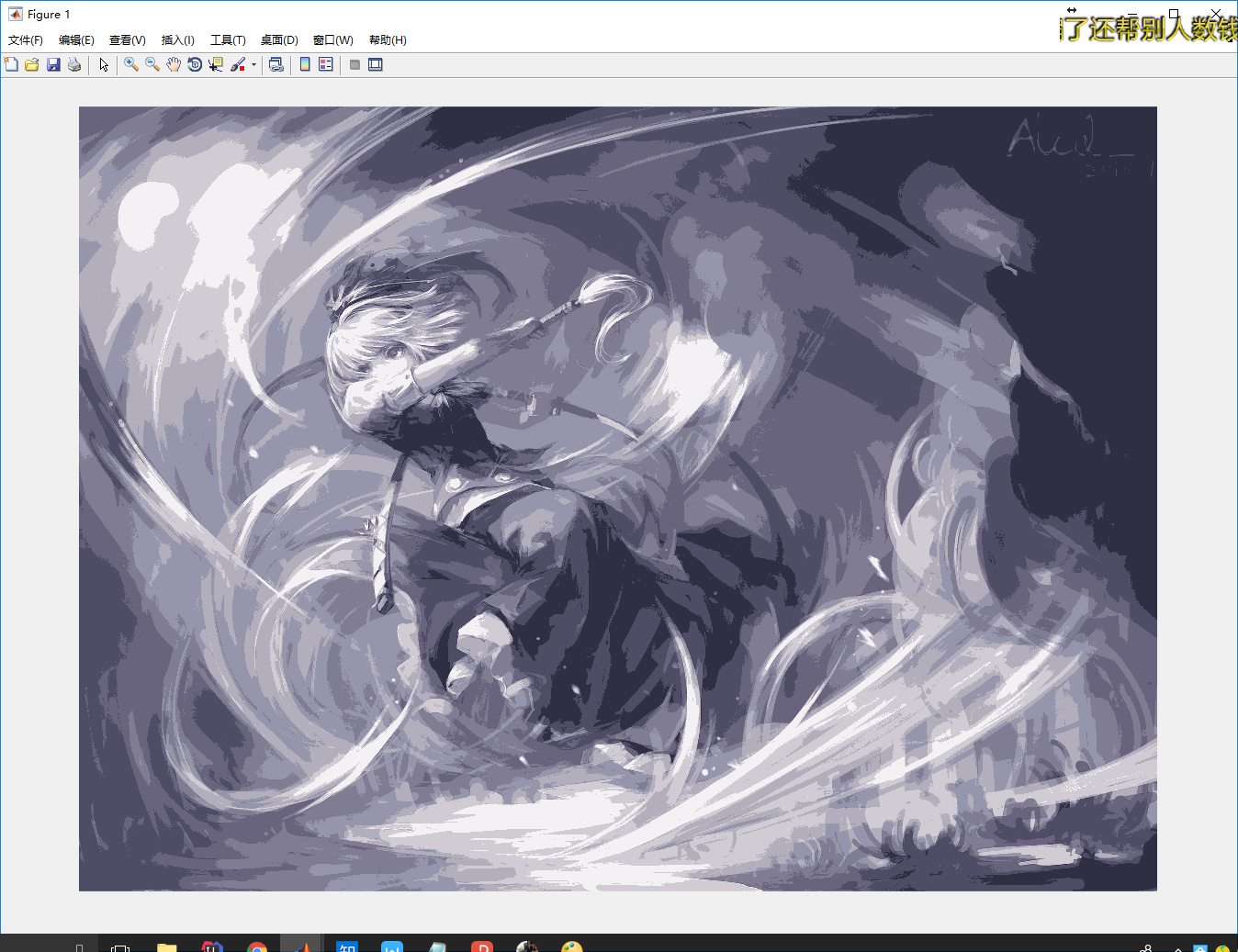
K = 32



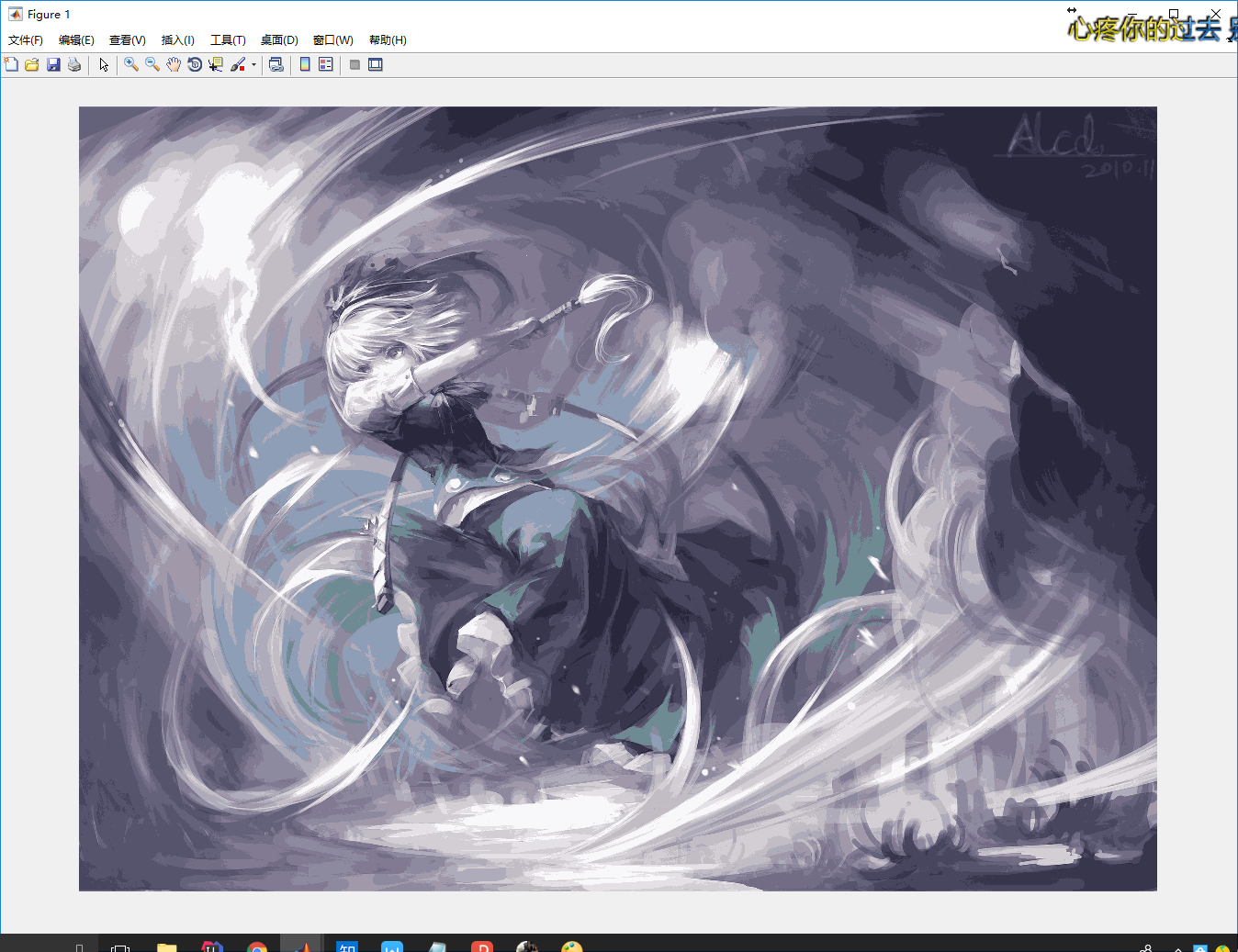
K = 64



K = 8



K = 16



K = 32



K = 64

