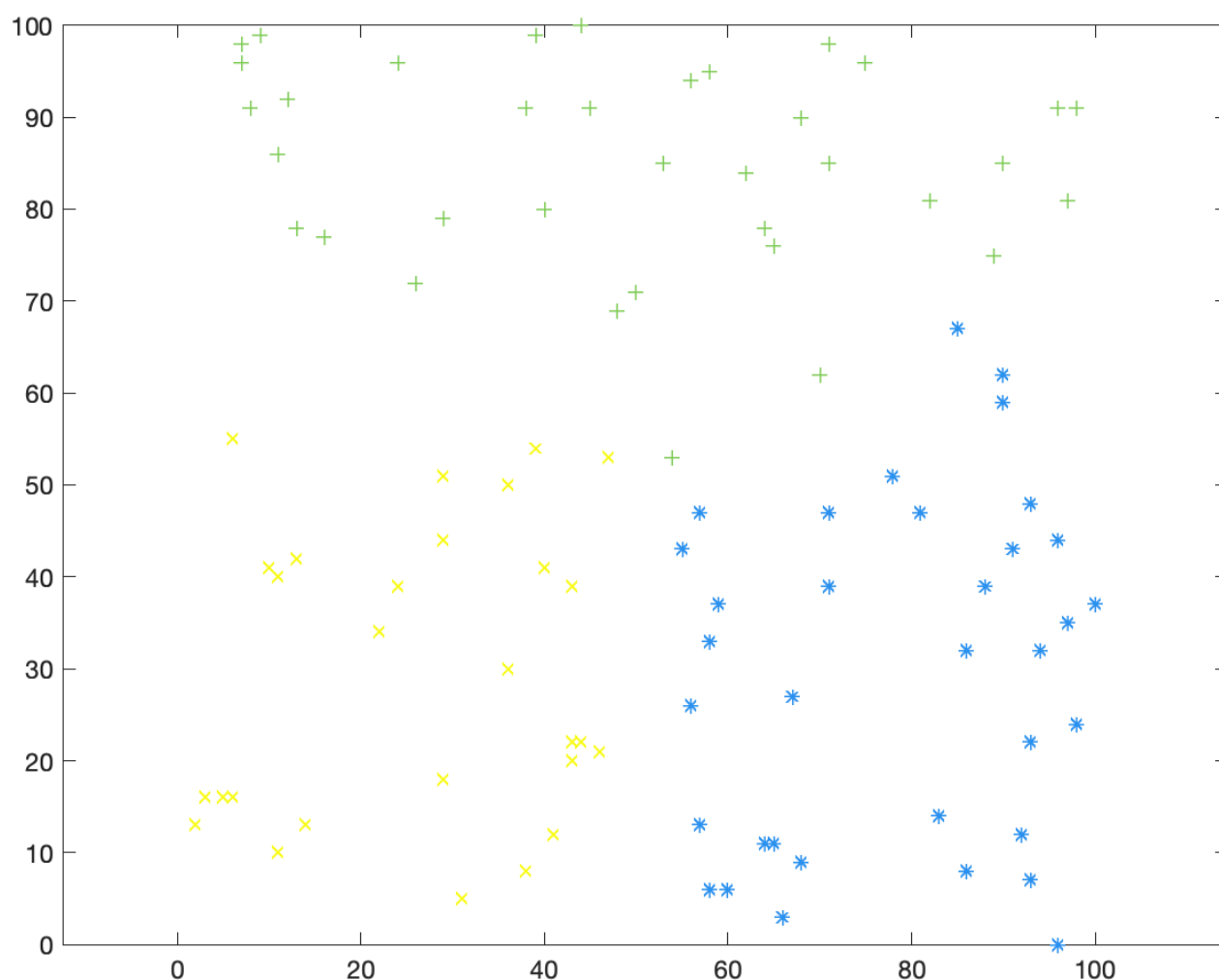# 作业5--KMeans 171870632 史健均

设置K=3 迭代次数=5时：

```
read a line:1,35,78.34285714285714,29.742857142857144
read a line:2,37,48.54054054054054,85.16216216216216
read a line:3,28,26.464285714285715,29.464285714285715
finished!

Process finished with exit code 0
```
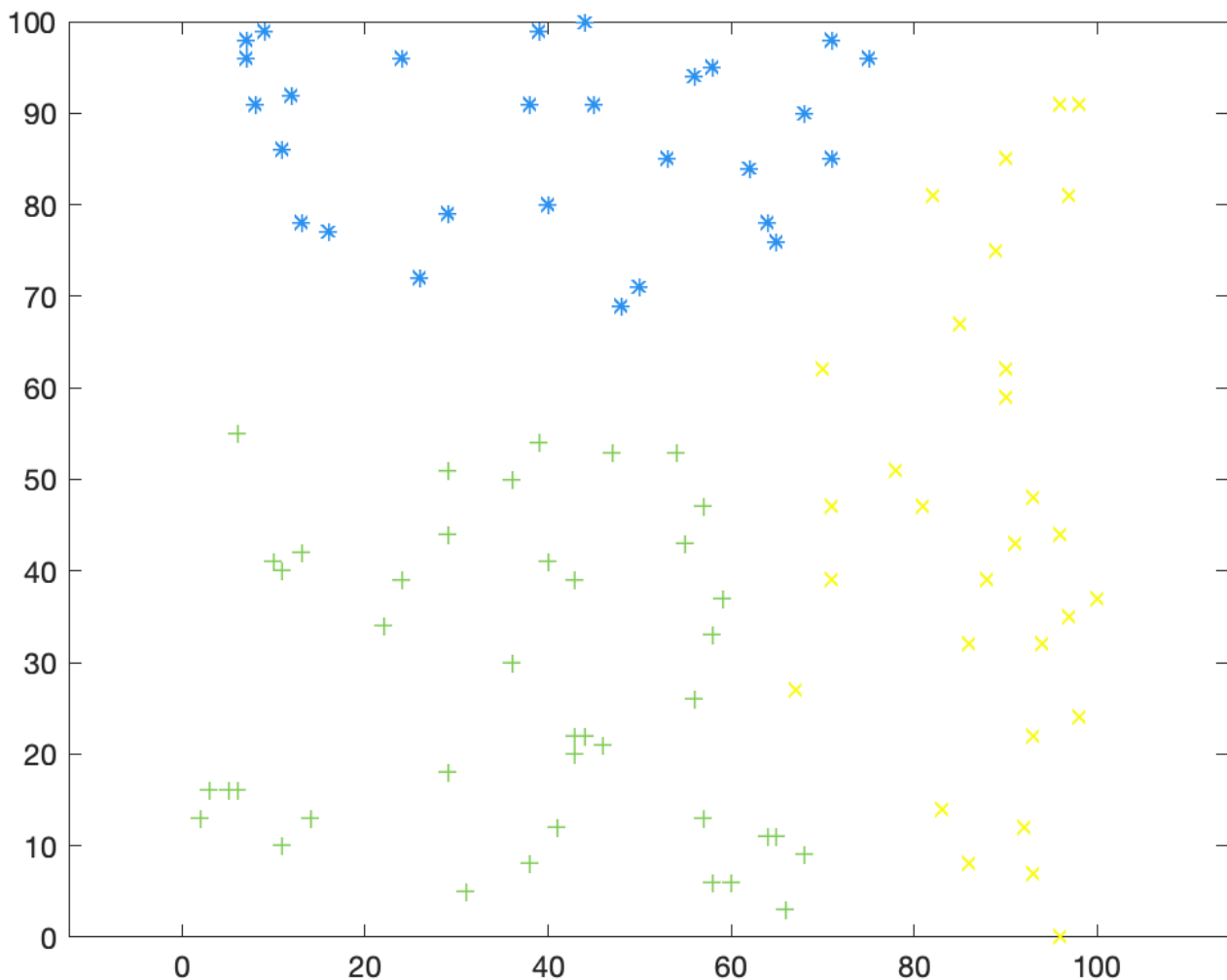
```
shijianjun@shijianjundeMacBook-Pro  ~/hadoop_installs/hadoop-3.2.1/bin    ./hadoop fs -ls /output
Found 7 items
drwxr-xr-x   - shijianjun supergroup          0 2019-11-04 14:30 /output/cluster-0
drwxr-xr-x   - shijianjun supergroup          0 2019-11-04 14:30 /output/cluster-1
drwxr-xr-x   - shijianjun supergroup          0 2019-11-04 14:30 /output/cluster-2
drwxr-xr-x   - shijianjun supergroup          0 2019-11-04 14:30 /output/cluster-3
drwxr-xr-x   - shijianjun supergroup          0 2019-11-04 14:30 /output/cluster-4
drwxr-xr-x   - shijianjun supergroup          0 2019-11-04 14:30 /output/cluster-5
drwxr-xr-x   - shijianjun supergroup          0 2019-11-04 14:30 /output/clusteredInstances
```


设置k=3 迭代次数=10时：

```
read a line:1,29,38.62068965517241,87.3103448275862
read a line:2,41,37.0243902439024,27.390243902439025
read a line:3,30,88.03333333333,45.4
finished!

Process finished with exit code 0
```
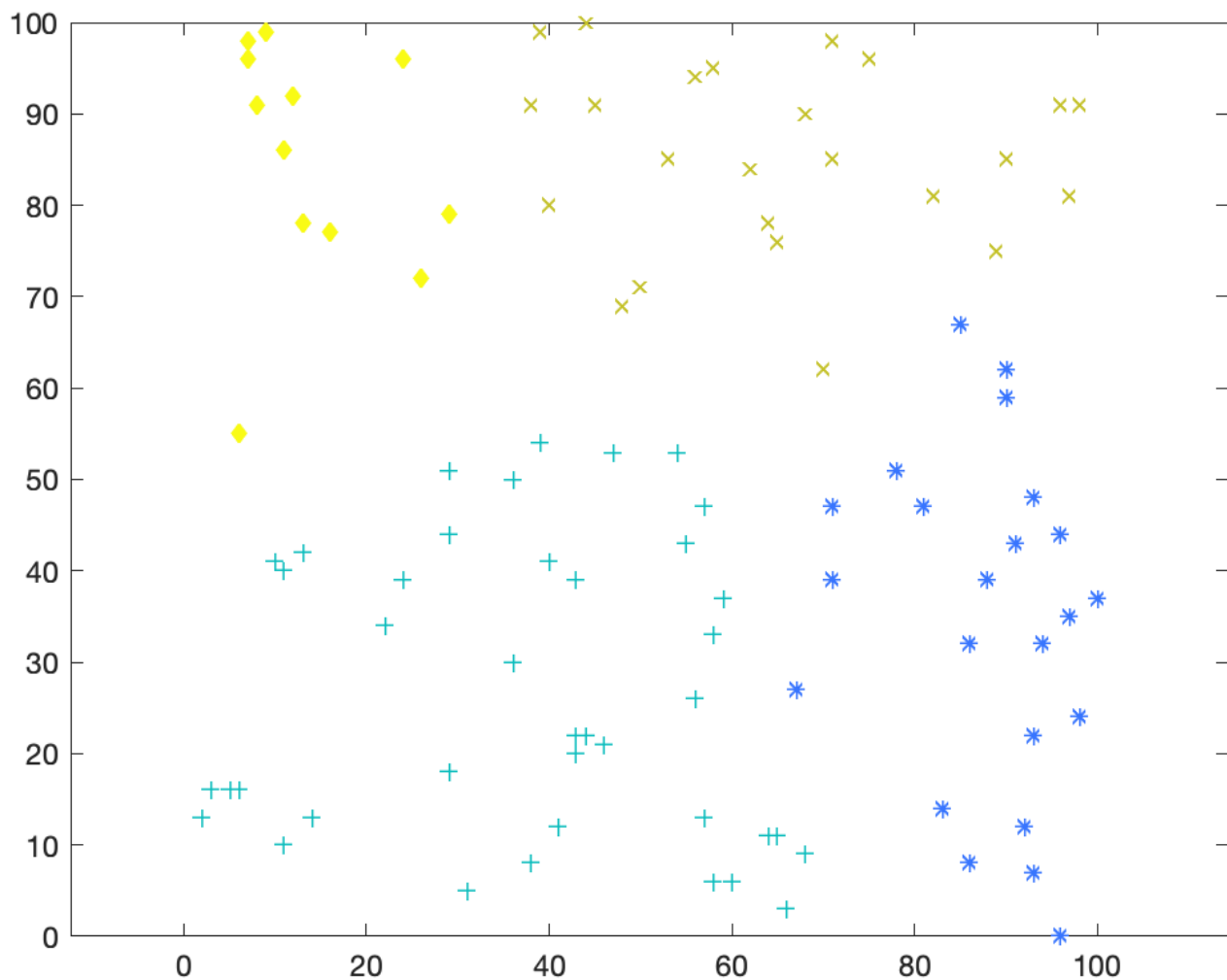

设置k=4 迭代次数=5时：

```
finished!
read a line:1,25,88.32,39.48
read a line:2,41,38.51219512195122,26.70731707317073
read a line:3,21,62.523809523809526,87.14285714285714
read a line:4,13,13.76923076923077,85.0
finished!

Process finished with exit code 0
```
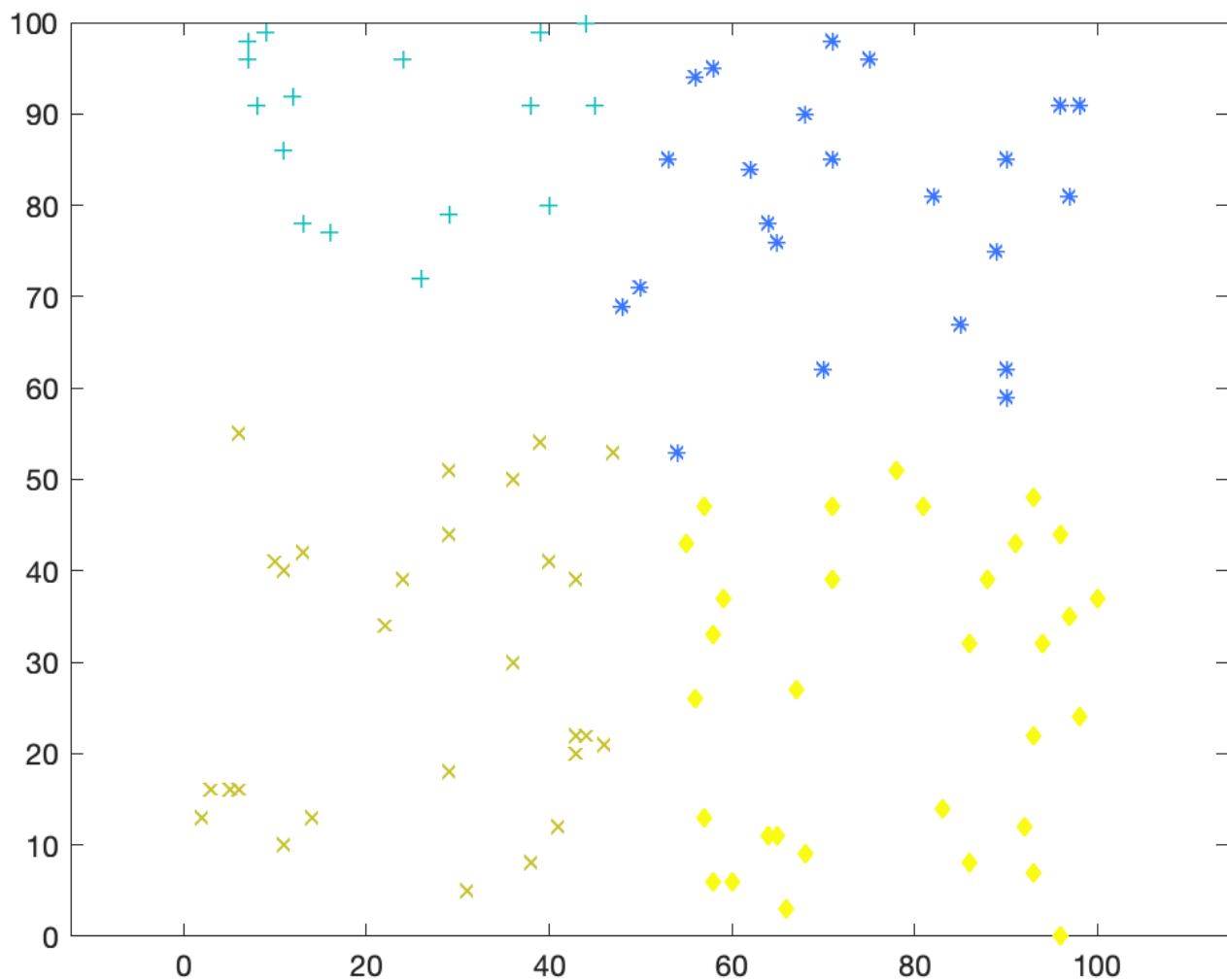
设置k=4 迭代次数=10时：

```
read a line:1,22,74.0,80.68181818181819
read a line:2,17,22.294117647058822,88.88235294117646
read a line:3,28,26.464285714285715,29.464285714285715
read a line:4,33,76.6969696969697,27.454545454545453
finished!

Process finished with exit code 0
```
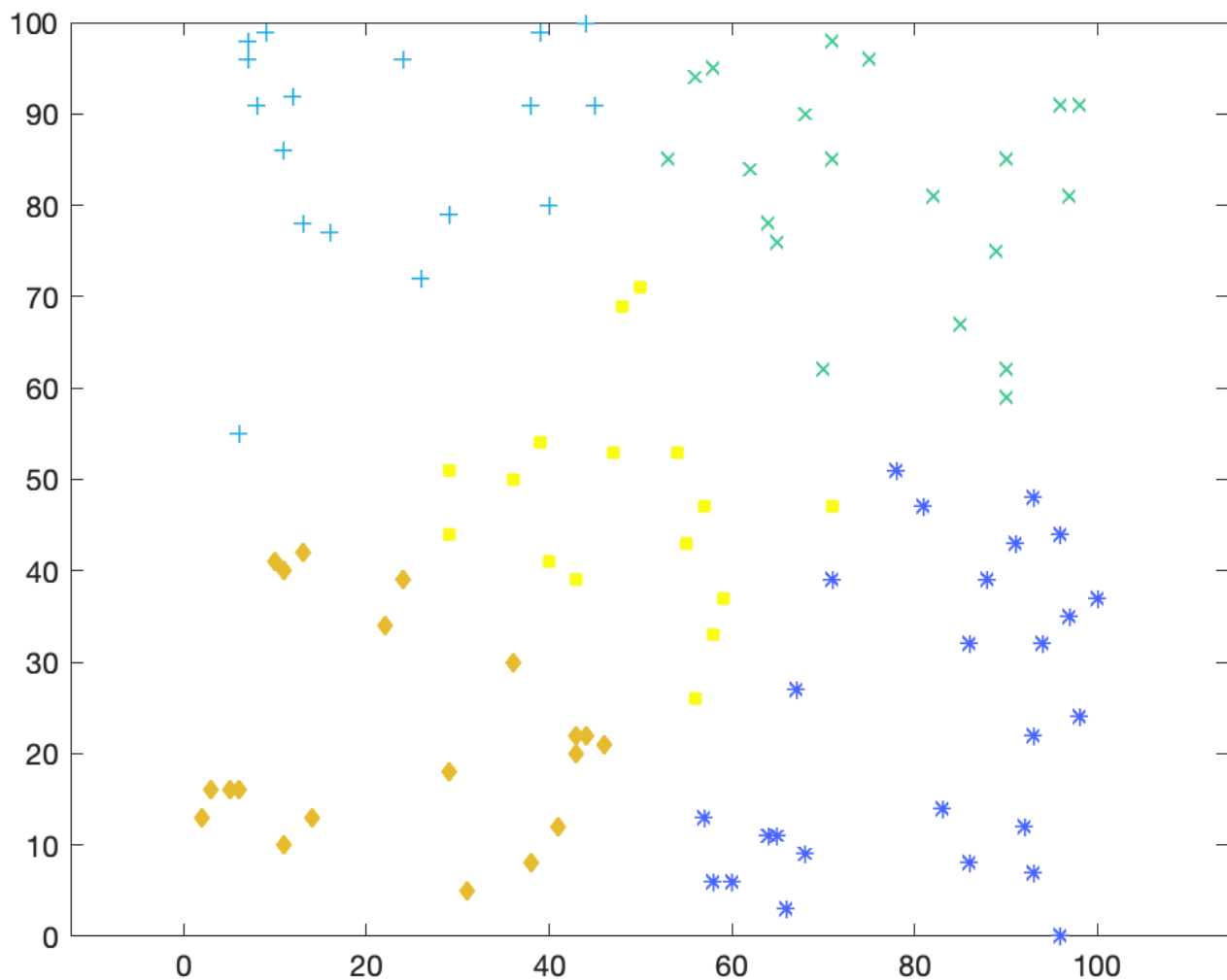
设置k=5 迭代次数=5时：

```
finished!
read a line:1,26,81.57692307692308,23.846153846153847
read a line:2,18,21.38888888888889,87.0
read a line:3,20,76.5,81.75
read a line:4,20,23.6,21.9
read a line:5,16,48.1875,47.375
finished!

Process finished with exit code 0
```
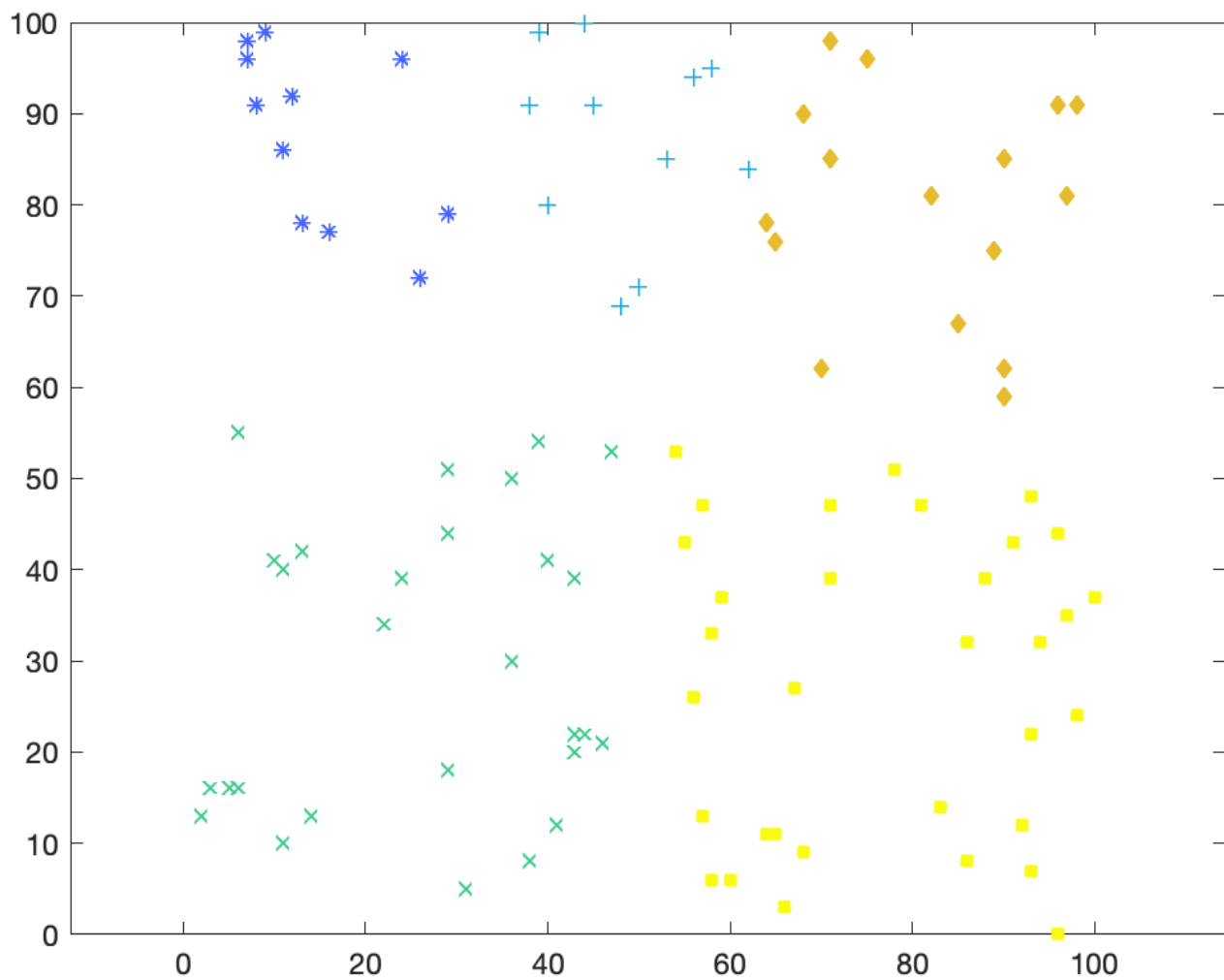
设置k=5 迭代次数=10时：

```
finished!
read a line:1,12,14.416666666666666,87.5
read a line:2,11,48.45454545454545,87.18181818181819
read a line:3,28,26.464285714285715,29.464285714285715
read a line:4,16,81.3125,79.8125
read a line:5,33,76.6969696969697,27.454545454545453
finished!

Process finished with exit code 0
```

经过尝试不同的k和迭代次数，可以显然看出结果不同；

个人认为取簇数为5，迭代次数为5时较好；

根据实际情况，比如数据代表的意义和作图损失函数，可以更好地调整簇数和迭代次数，进而达到更好的分类标准。