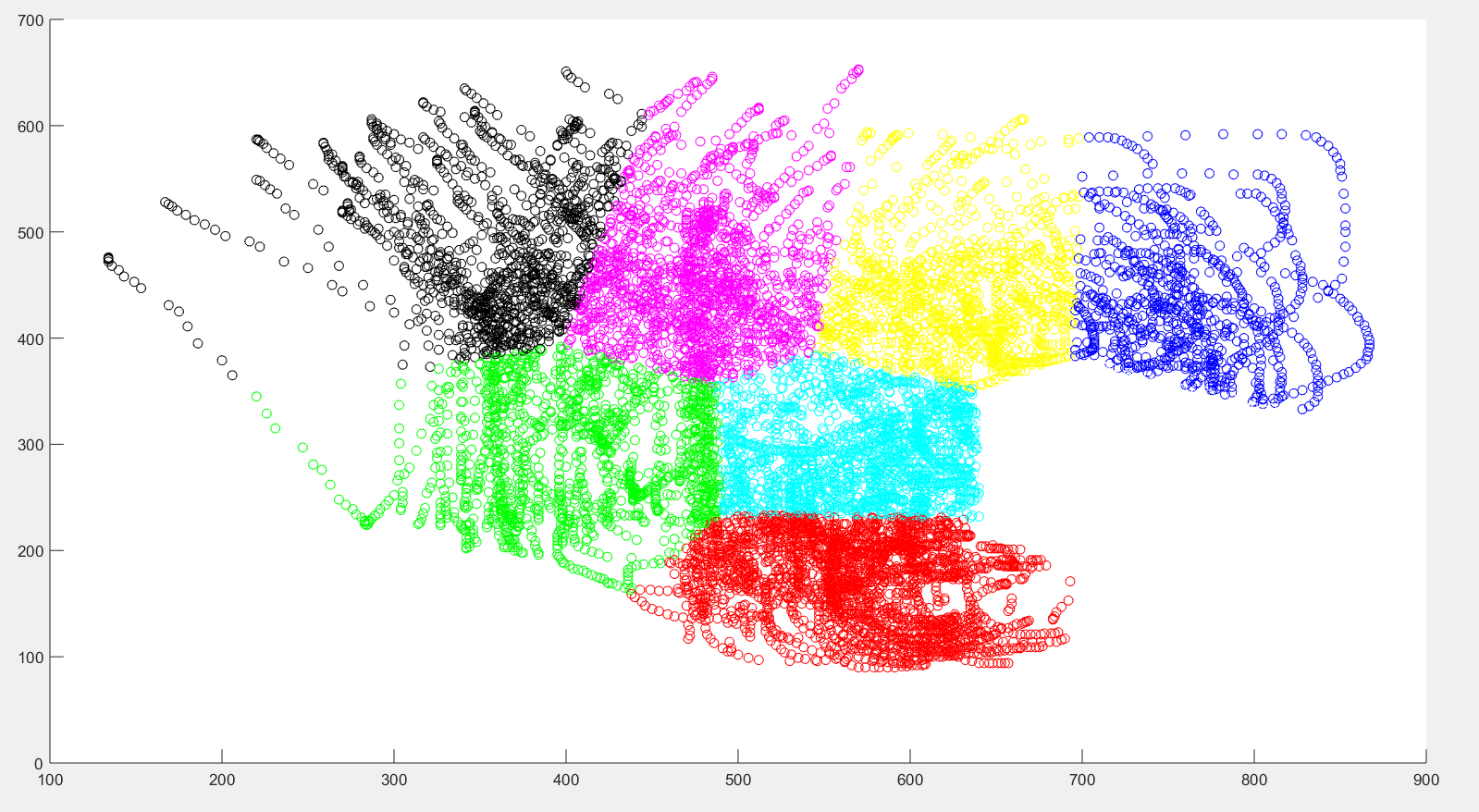
## （2）kmeans



## Implement the standard Hidden Markov Model.

HMM input: M = [Pi, A, B], sequences of observations.

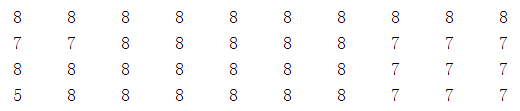
**For sequences of observations:**

I tried three ways to get a sequence of observations.

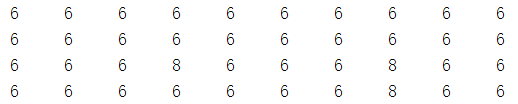
1. **get a sequence of observations by k\_means**

Interpolate in the supplied sequence, calculate the class of interpolation points, and obtain the following data：

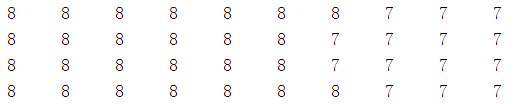
A



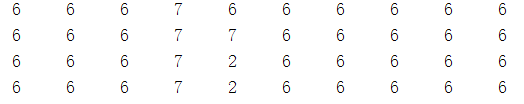
E



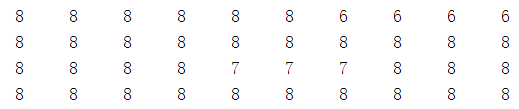
I



O

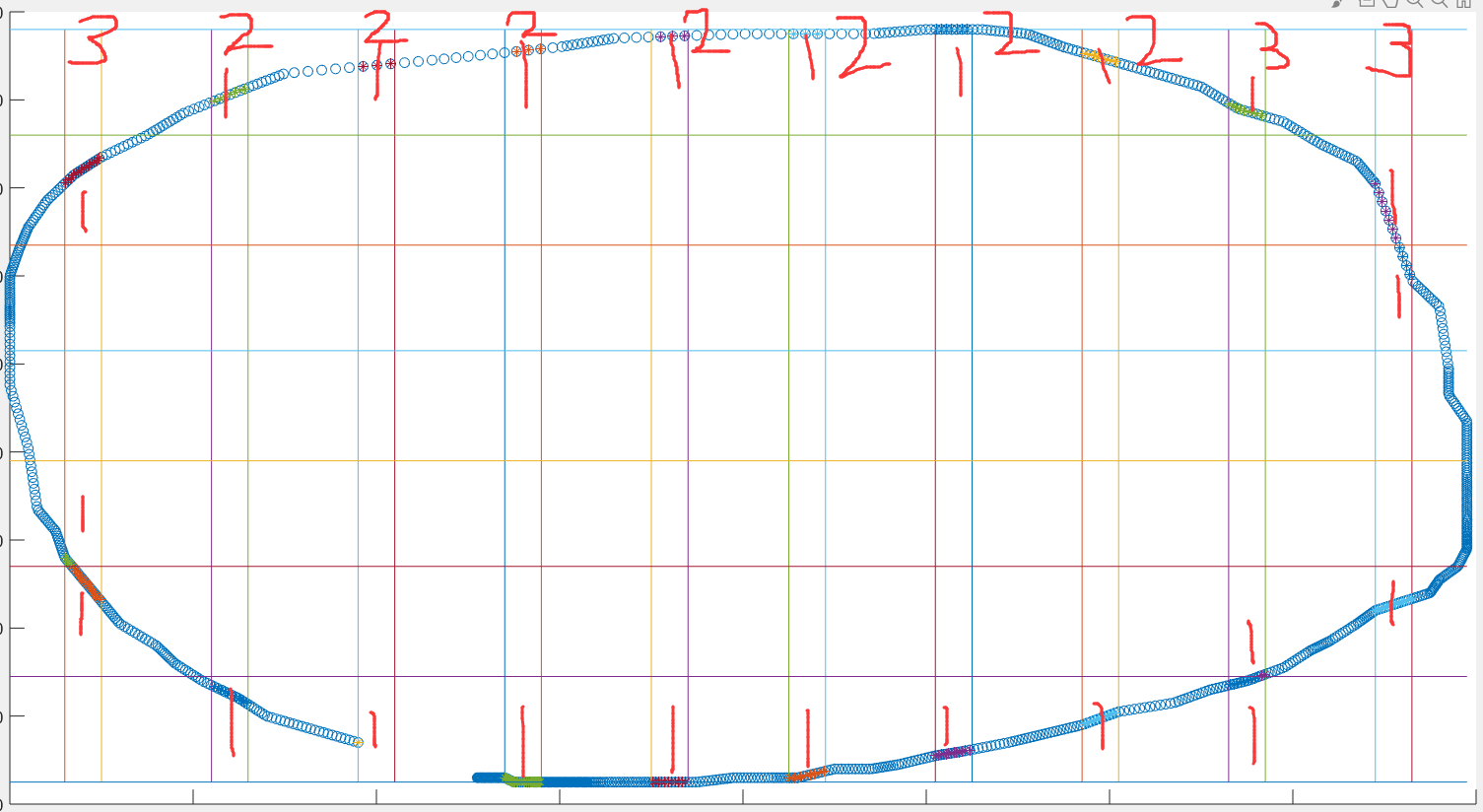


U

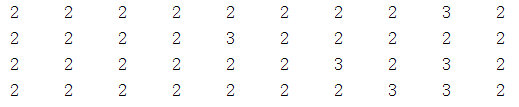


1. **get a sequence of observations by drawing**

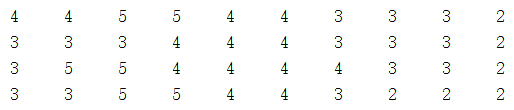
Interpolate in the supplied sequence to make the data denser and compute as shown below：



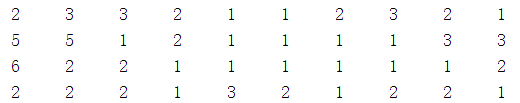
A



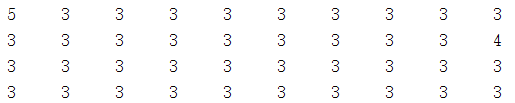
E



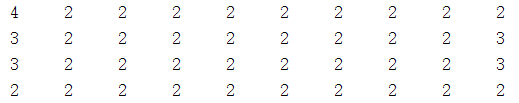
I



O

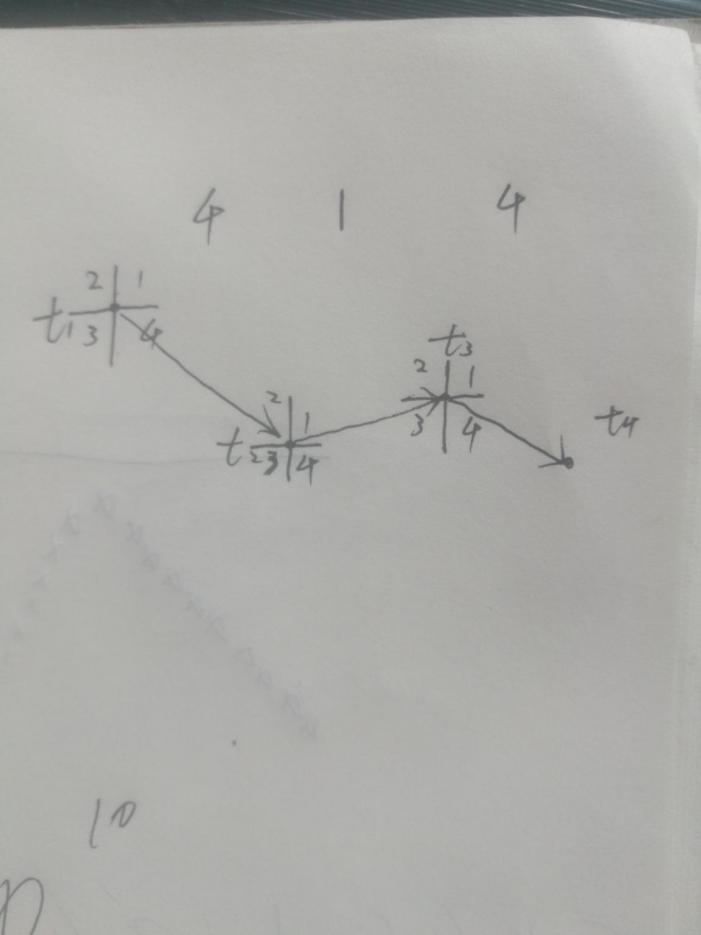


U

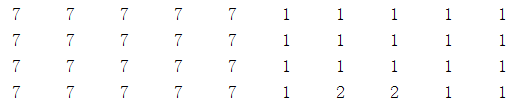


1. **get a sequence of observations by point toward**

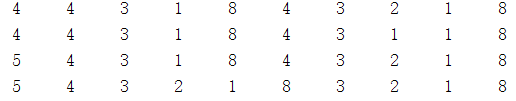
Interpolate in the supplied sequence and calculate as shown below：



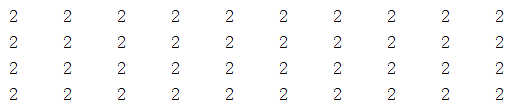
A



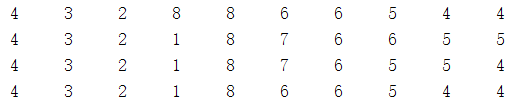
E



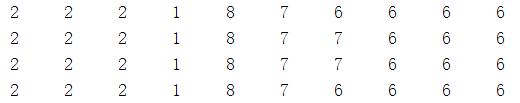
I



O



U



If I just look at the observation sequence, I can see that in the first method, A and I are easy to be misjudged, and E and O are easy to be misjudged. In the second method, A and U are easy to misjudge. In the third method, O and U are easy to misjudge.

On the other hand, the first and second methods use only coordinate information, and the third method uses time information and coordinate information, so I use the third method.

**For M = [Pi, A, B]:**

M can be obtained by estimating the hidden state.

The given sequence is interpolated and classified by function ‘discretize’ to get the estimated hidden state.

## (4)

Length of observation sequence = 10

Number of observation States = 8

Number of hidden states = 4

confusion matrix:

20 0 0 0 0

0 19 0 1 0

0 0 19 0 1

0 0 0 20 0

0 0 0 0 20

## (5)

1.

Length of observation sequence = 10

Number of observation States = 8

Number of hidden states = 1

confusion matrix:

19 0 0 0 1

0 19 0 1 0

0 0 20 0 0

0 0 0 20 0

0 0 0 0 20

The number of hidden states does not need to be too large.

2.

Length of observation sequence = 5

Number of observation States = 8

Number of hidden states = 4

confusion matrix:

20 0 0 0 0

1 17 0 2 0

0 0 20 0 0

0 0 0 20 0

0 0 0 0 20

Length of observation sequence = 10

Number of observation States = 16

Number of hidden states = 4

confusion matrix:

20 0 0 0 0

0 16 0 4 0

0 0 13 2 5

0 0 0 20 0

0 0 0 0 20

When number of observation States is larger than length of observation sequence,the accuracy of hidden Markov model is greatly reduced