Work I

**Pattern Recognition and Machine Learning Report**

冯宇婷 2111712336

Computer Technology Master 1701

By learning and practicing, I finished my homework for MachineLearning and Pattern Recognition.The following is the first homework report, and the report is divided into three parts: section 1, I show the relation of the code ;then in section2, I show the result of the code that running on the matlab;Section3,I illustrate the GMM,EM and K-means and I give the definition of these terms,and show that how can we get it in math.

# 1.Code Analysis

Firstly, get parameter of EM, call the interface EM\_init\_kmeans to get weights of Priors, the average values of Mu, and covariance of Sigma.



Fig. 1. Get Priors, Mu, Sigma

How to initialize the parameter, divide data4model into nbStates, and calculate the date of Priors, Mu and Sigma. The code of EM\_init\_kmeans in Fig. 2.

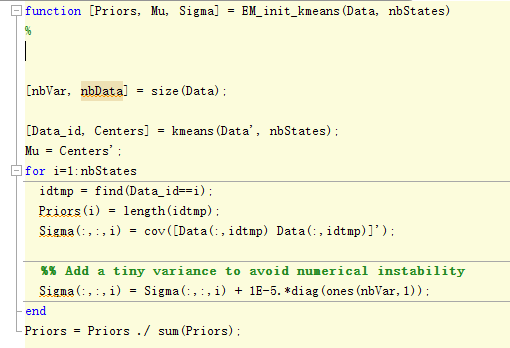


Fig. 2. The code of EM\_init\_kmeans

Next, train GMM model. Using EM interface to train the model. EM interface is shown in Fig. 3.



Fig. 3. Train the GMM model

Now, let’s see EM interface. Firstly, initialize the Priors, Mu and Sigma.

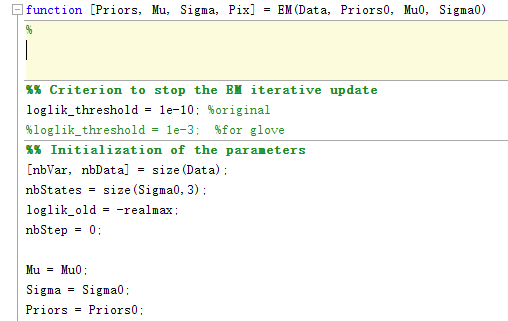


Fig.4. Initialize the Priors, Mu and Sigma

Next, we can see the entire EM process, step one is E-step,

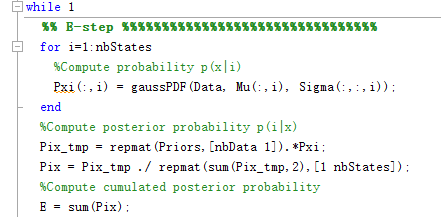


Fig. 5. E-step

E-step compute cumulated posterior probability.

Step two is M-step.

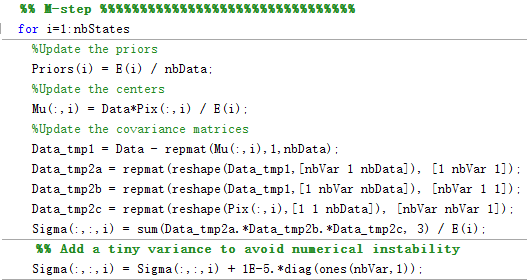


Fig. 6. M-step

M-step update the priors, centers, covariance matrices.

Step three, judging whether the operation is stopped.

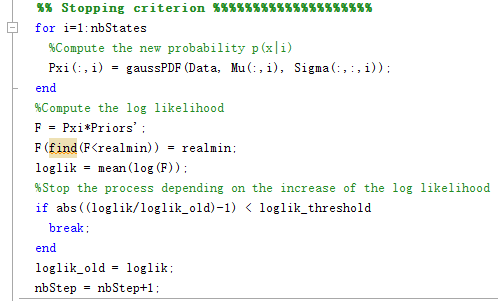


Fig. 7. Stopping criterion

When the increase is smaller than the log likelihood, the code stop. Otherwise, go step one to carry out E-step.

After EM, I can get Priors, Mu, Sigma, the data is shown in Fig. 8 and Fig. 9.

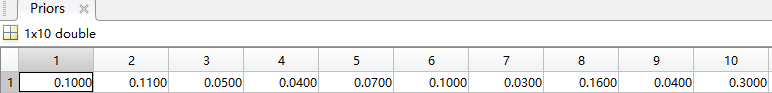


Fig. 8. The data of priors

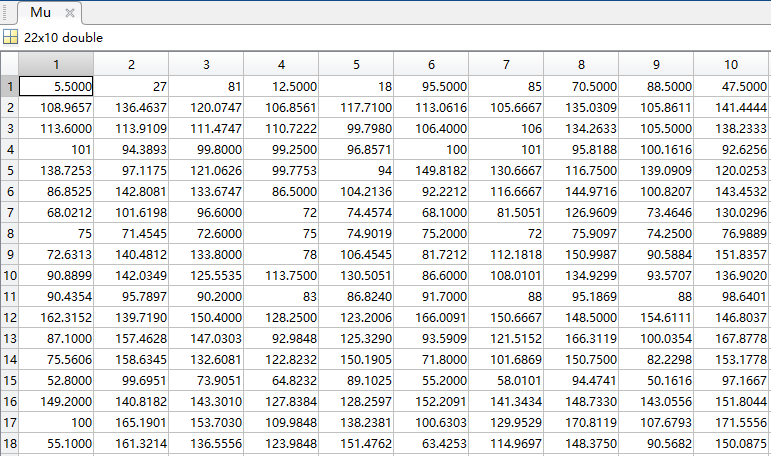


Fig. 9-a. The data of mu

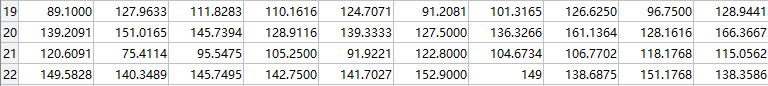


Fig. 9-b. The data of mu

Finally, recognize the test\_data, the recognition step is shown in Fig. 10.

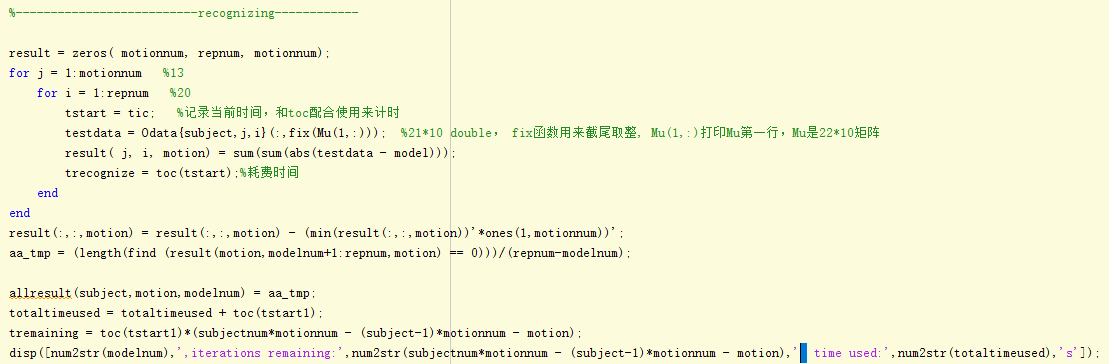


Fig. 10. The recognition step

# 2. Program Results

The recognition rate is shown in Fig. 11.

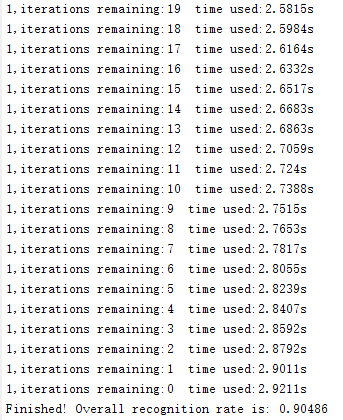


Fig.11. The recognition rate

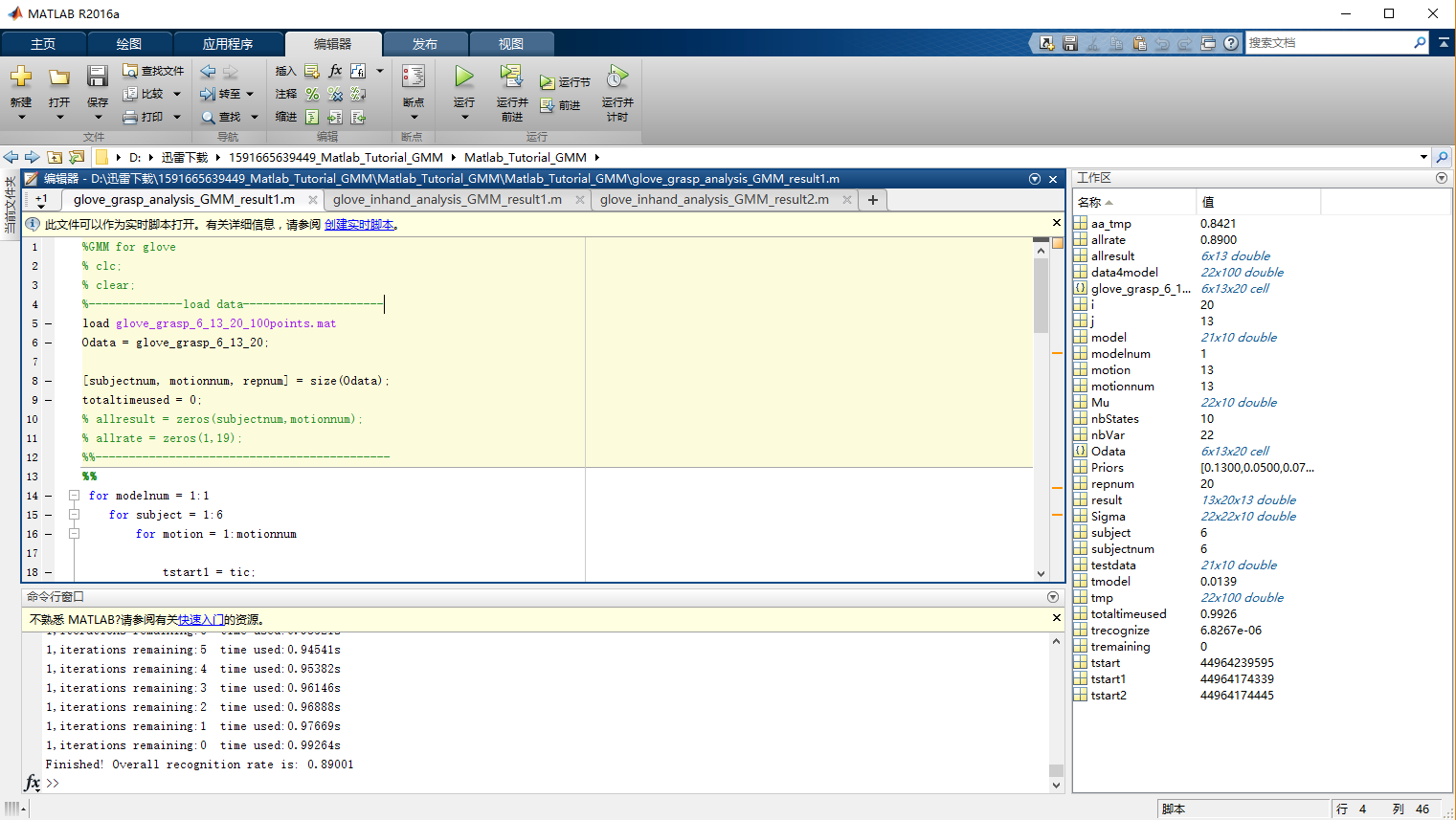


Fig.12. Glove\_grasp\_analysis\_GMM\_result1

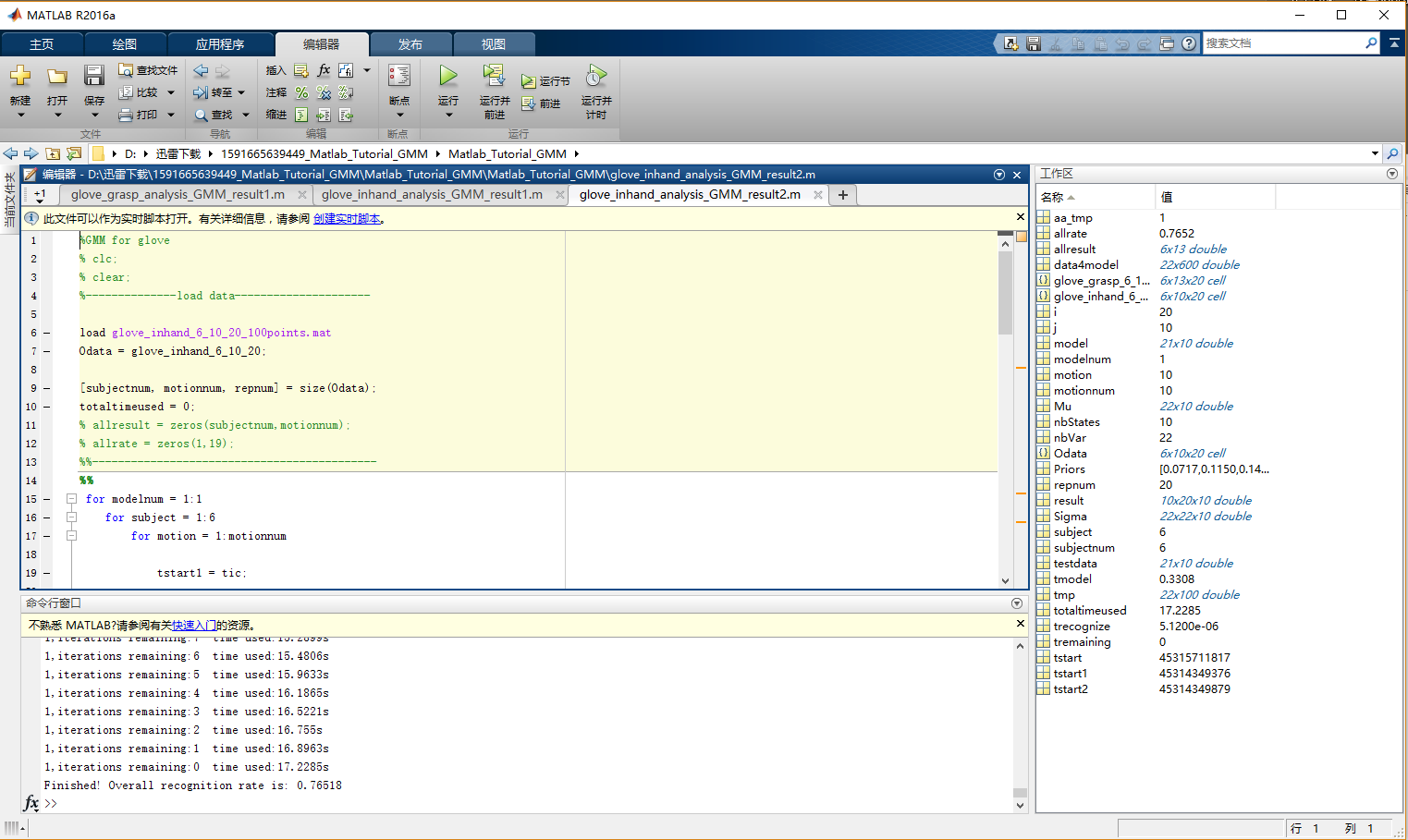


Fig.13. Glove\_grasp\_analysis\_GMM\_result2

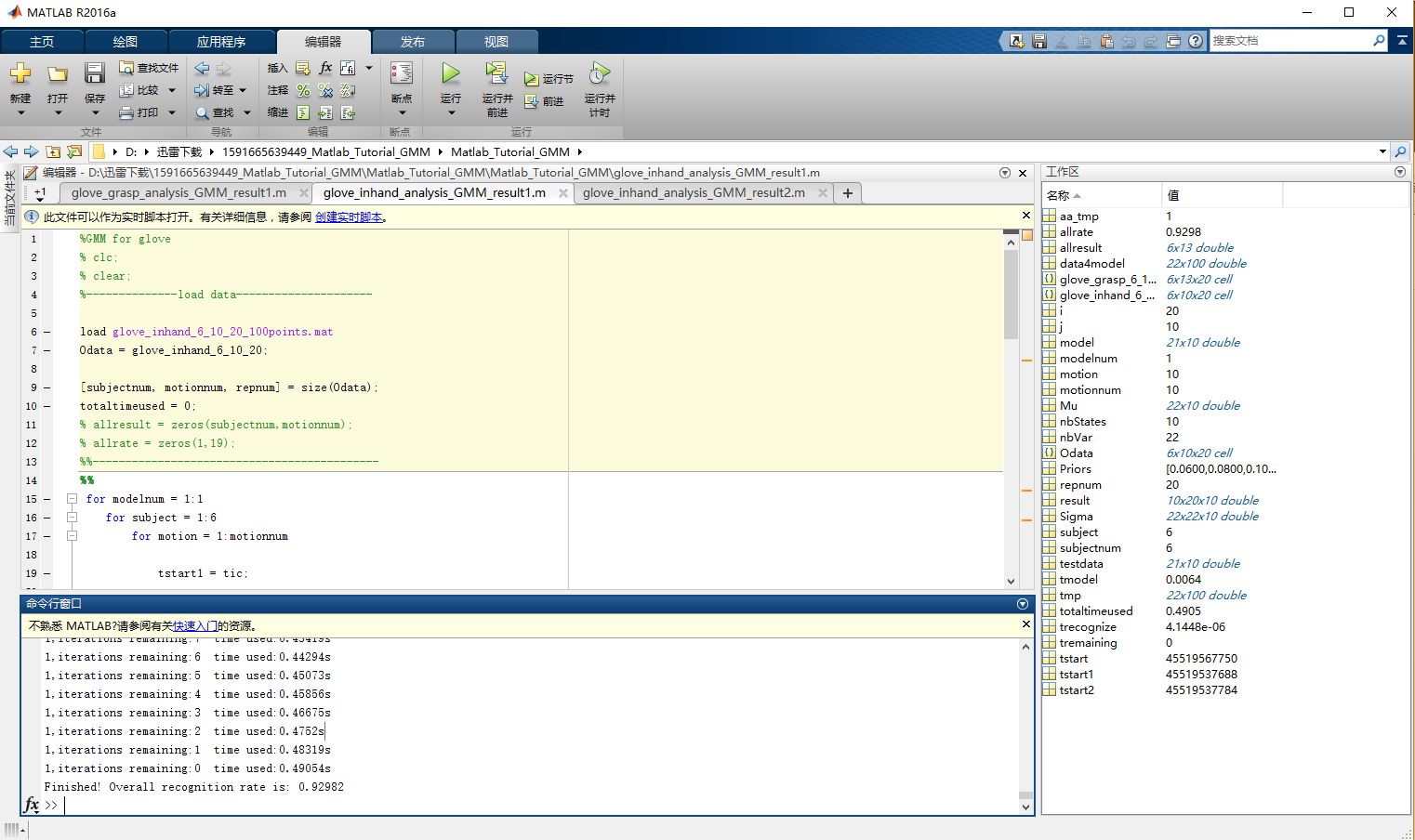


Fig.14. Glove\_inhand\_analysis\_GMM\_result1

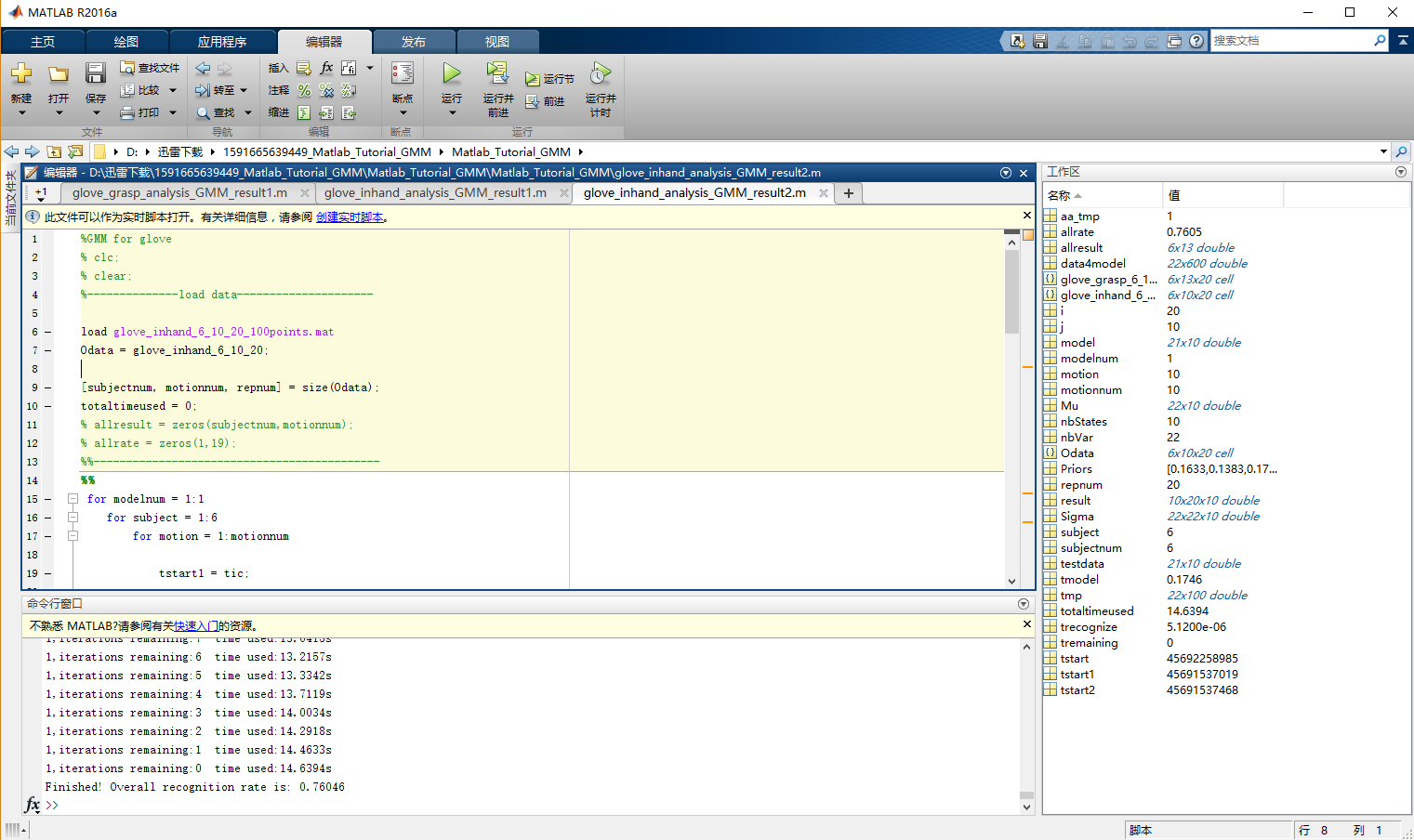


Fig.15. Glove\_inhand\_analysis\_GMM\_result2

3. Understanding of Kmeans,GMM,EM and KNN

3.1 Kmeans algorithm

k-means algorithm is a method of vector quantization, originally from signal processing, that is popular for cluster analysis in data mining. k-means algorithm aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. This results in a partitioning of the data space into Voronoi cells.

The problem is computationally difficult (NP-hard). However, there are efficient heuristic algorithms that are commonly employed and converge quickly to a local optimum. These are usually similar to the expectation-maximization algorithm for mixtures of Gaussian distributions via an iterative refinement approach employed by both algorithms. Additionally, they both use cluster centers to model the data; however, k-means clustering tends to find clusters of comparable spatial extent, while the expectation-maximization mechanism allows clusters to have different shapes.

Given a set of observations , where each observation is a d-dimensional real vector, k-means clustering aims to partition the n observations into k ,but we should promise , sets  so as to minimize the within-cluster sum of distance functions of each point in the cluster to the K center. In other words, its objective is to find:

(1)

Here, is the mean of points in .

Given an initial set of k means , the algorithm proceeds by alternating between two steps:

Assignment step: Assign each observation to the cluster whose mean yields the least within-cluster sum of squares. Since the sum of squares is the squared Euclidean distance, this is intuitively the "nearest" mean.

(2)

where each is assigned to exactly one, even if it could be assigned to two or more of them.

Update step: Calculate the new means to be the centroids of the observations in the new clusters.

(3)

The algorithm has converged when the assignments no longer change,otherwise, assignment step and update step are repeatedly implemented.

3.2 GMM algorithm

The abbreviation for Gaussian mixture model is GMM. In statistics, a mixture model is a probabilistic model for representing the presence of subpopulations within an overall population, without requiring that an observed data set should identify the sub-population to which an individual observation belongs. Formally a mixture model corresponds to the mixture distribution that represents the probability distribution of observations in the overall population.

The Gaussian mixture distribution can be written as a linear superposition of Gaussians in the form:

(4)

There are introduce a K-dimensional binary random variable having a 1-of-K representation in which a particular element is equal to 1 and all other elements are equal to 0. The values of , therefore satisfy and and we see that there are K possible states for the vector z according to which element is nonzero. We shall deﬁne the joint distribution in terms of a marginal distribution and a conditional distribution The marginal distribution over is speciﬁed in terms of the mixing coefﬁcients , such that

(5)

where the parametersmust satisfy

(6)

and

(7)

In order to be valid probabilities. Because z uses a 1-of-K representation, we can also write this distribution in the form,

(8)

Similarly, the conditional distribution of x given a particular value for z is a Gaussian,

(9)

The joint distribution is given by ), and the marginal distribution of x is then obtained by summing the joint distribution over all possible states of z to give,

(10)

The formula is made use of (8) and (9). Thus the marginal distribution of x is a Gaussian mixture of the form (4).

Another quantity that will play an important role is the conditional probability of z given x. We shall use to denote , whose value can be foundusing Bayes’ theorem,

(11)

We shall view as the prior probability of , and the quantity as the corresponding posterior probability once we have observed x.

**Maximum likelihood:** Suppose we have a dataset of observations ,from (4) the log of the likelihood function is given by,

(12)

**EM for GMM:** Given a Gaussian mixtur model, the goal is to maximize the likelihood function with respect to the parameters. The steps as follows,

1. Initialize the means , covariances and mixing coefﬁcients , and evaluate the initial value of the log likelihood.
2. **E step**. Evaluate the responsibilities using the current parameter values,

(13)

1. **M step**. Re-estimate the parameters using the current responsibilities,

(14)

(15)

(16)

where

(17)

1. Evaluate the log likelihood,

(18)

And check for convergence of either the parameters or the log likelihood. If the convergence criterion is not satisﬁed return to step 2 to process E step.

3.3 EM algorithm

The expectation maximization algorithm(EM algorithm) is a general technique for ﬁnding maximum likelihood solutions for probabilistic models having latent variables.

Consider a probabilistic model in which we collectively denote all of the observed variables by X and all of the hidden variables by Z. The joint distribution is governed by a set of parameters denoted . Our goal is to maximize the likelihood function that is given by

(19)

Here we are assuming Z is discrete, although the discussion is identical if Z comprises continuous variables or a combination of discrete and continuous variables, with summation replaced by integration as appropriate.

Deﬁne over the latent variables, and we observe that, for any choice of , the following decomposition holds

(20)

Where deﬁne

(21)

(22)

To verify the decomposition (20), we ﬁrst make use of the product rule of probability to give

(23)

Which we then substitute into the expression for ). This gives rise to two terms, one of which cancels while the other gives the required log likelihood after noting that is a normalized distribution that sums to1.

The following describes the EM algorithm operation. Given a joint distribution , over observed variables X and latent variables Z, governed by parameters , the goal is to maximize the likelihood function with respect to .

1. Choose an initial setting for the parameters .
2. **E step**: Evaluate
3. **M step**: Evaluate given by

(24)

Where

(25)

1. Check for convergence of either the log likelihood or the parameter values. If the convergence criterion is not satisﬁed, then let

(26)

and return to step 2.

3.4 Relationship between Kmeans, GMM, EM

**EM algorithm:** The GMM algorithm is a specific example of the EM algorithm. EM algorithm to solve the problem is: to cluster the data, assuming the data obey the probability distribution of heterozygosity, the distribution of specific parameters unknown, involving random variables in two groups, one group can observe the other group of unobservable . Now use the maximum likelihood estimation to get the distribution parameters. If the two groups of random variables involved are observable, the problem can be solved immediately, and the solution of the distribution parameter can be obtained by maximizing the likelihood function. The EM algorithm first assigns initial values ​​to the desired distribution parameters so that the expectation of the hidden variables can be calculated. Then, the maximum likelihood is obtained by using the expected variables of the hidden variables and the observable variables. Thereby updating the distribution parameters. And then use the updated distribution parameters to calculate the expectations of implied variables, and then with the expected data can be observed with the distribution of the updated parameters. That is to say, the EM algorithm consists of two steps: concealment variable expectation computation (E-step) and likelihood function maximum value calculation (M-step), assigning an initial value to the distribution parameter and then iteratively executing E-step and M- Step until the algorithm converges.

**GMM algorithm:** The GMM algorithm, as an example of the family of EM algorithms, specifies that the distribution of each participating heterozygosity is a Gaussian distribution, that is, the distribution parameters appear as mean Mu and variance Sigma. The parameters of each Gaussian distribution are calculated iteratively using the EM algorithm as the framework for the computation.

**Comparison of GMM and Kmeans:** Similarity: They are the iterative implementation of the algorithm, and iterative strategy is the same: the algorithm begins when the first need to calculate the parameters assigned to the initial value and then the two steps are then alternately performed, one step being the estimation of the data(Kmeans is to estimate the cluster of each point; GMM is to calculate the expectations of implicit variables.); The second step is to recalculate the parameter values with the estimated values calculated in the previous step and update the target parameters(Kmeans is to calculate the cluster center position; GMM is to calculate the center of each Gaussian distribution and covariance matrix).

Difference**:** (1).The parameters to be calculated are different: k-means is the cluster center position; GMM is the parameter of each Gaussian distribution.(2). The method of calculating the target parameters is different: k-means is the mean value of the position of all the elements in the current cluster; GMM is the probability-based algorithm. The distribution parameter is solved by computing the maximum of the likelihood function.