**PRML一**

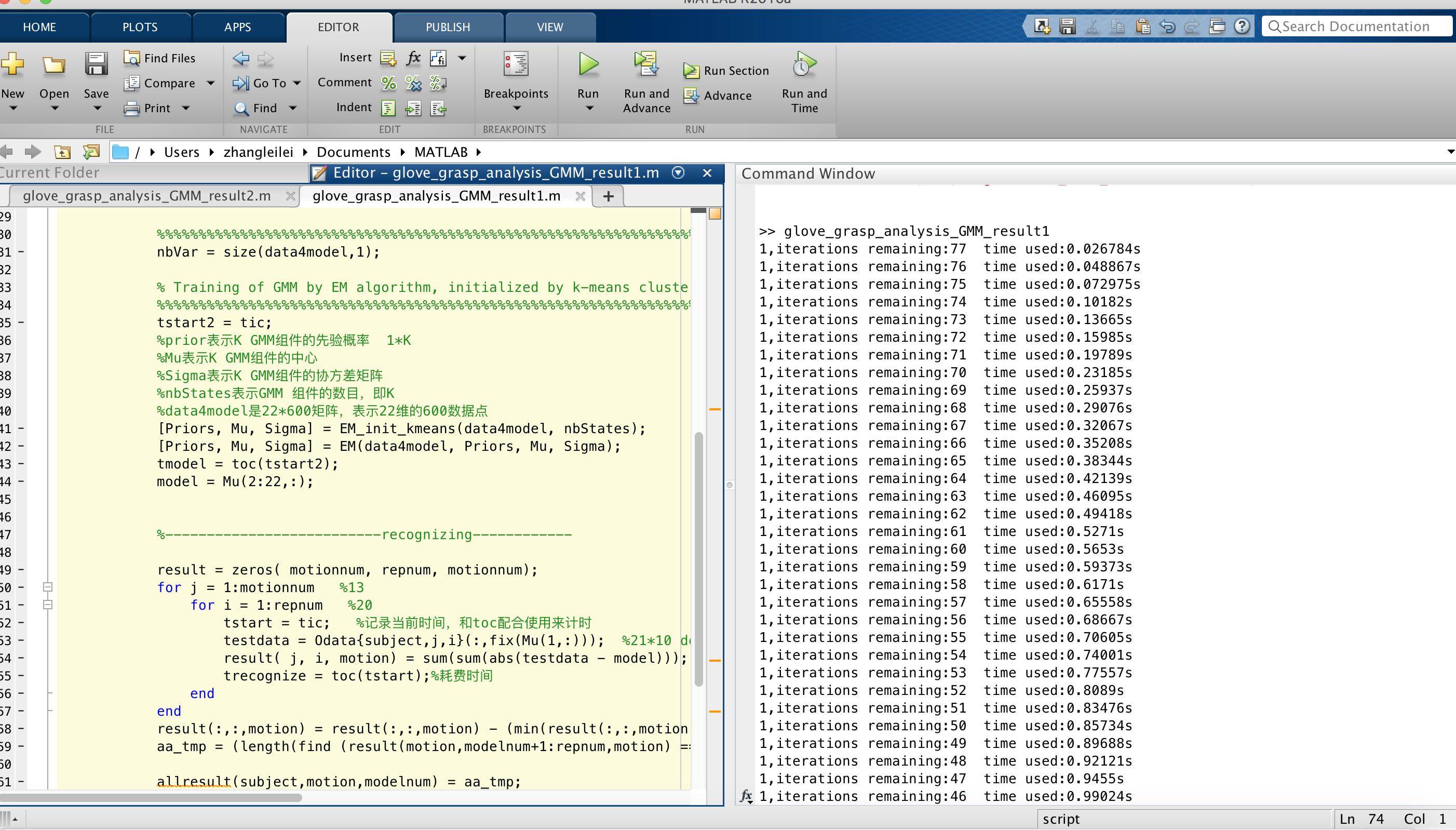
**Pattern Recognition and Machine Learning Repor**

张雷雷 2111712343

Computer Technology Master 1702

Program running

Run screenshots：



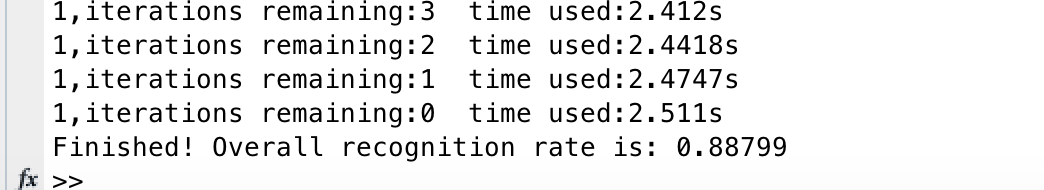
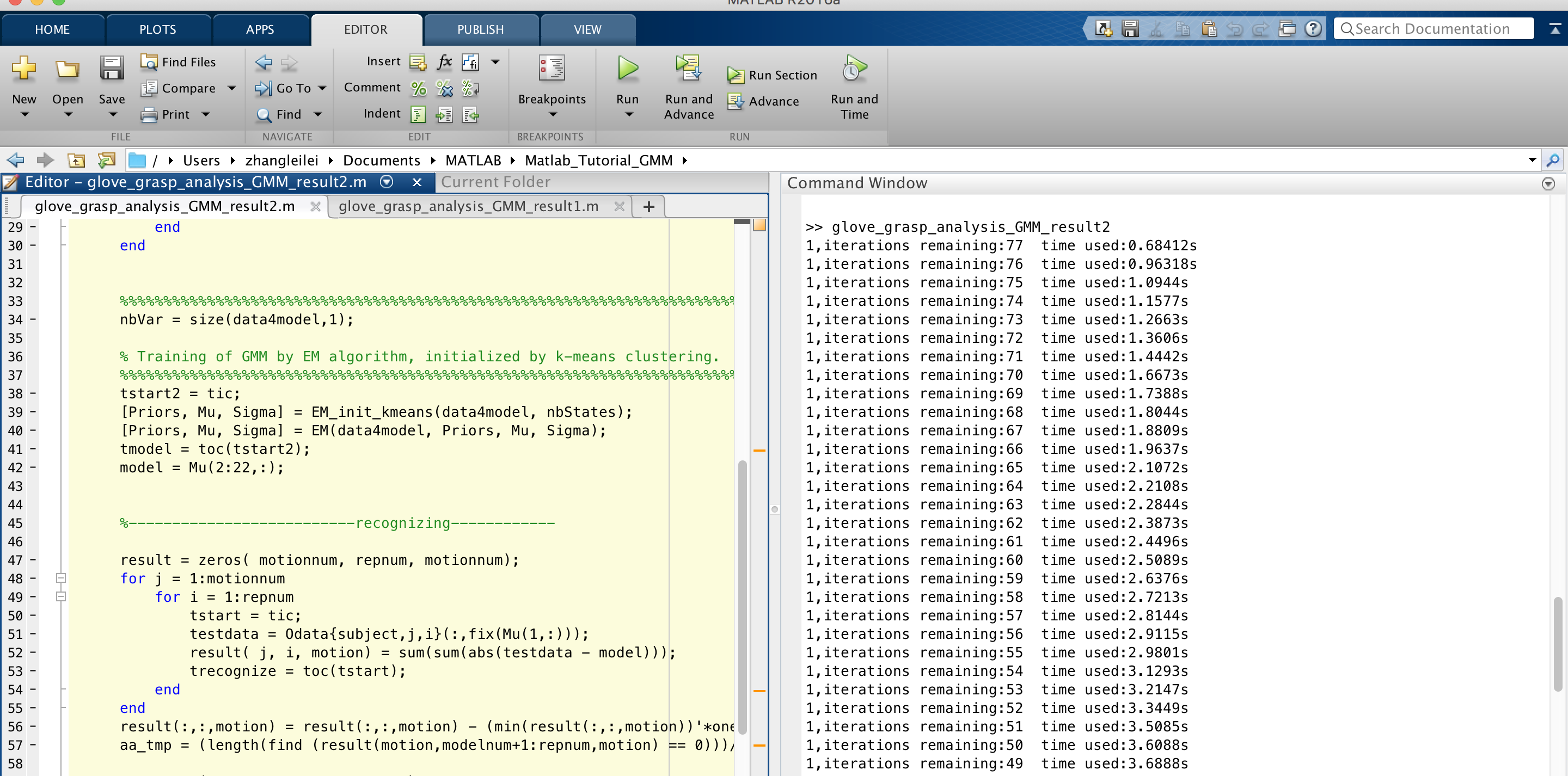


Fig0-Glove\_grasp\_analysis\_GMM\_result1



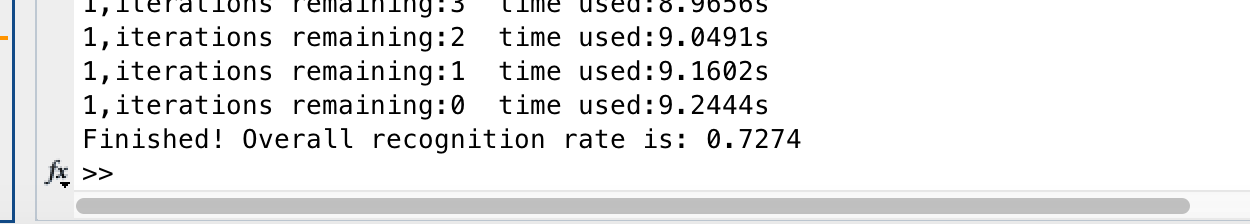
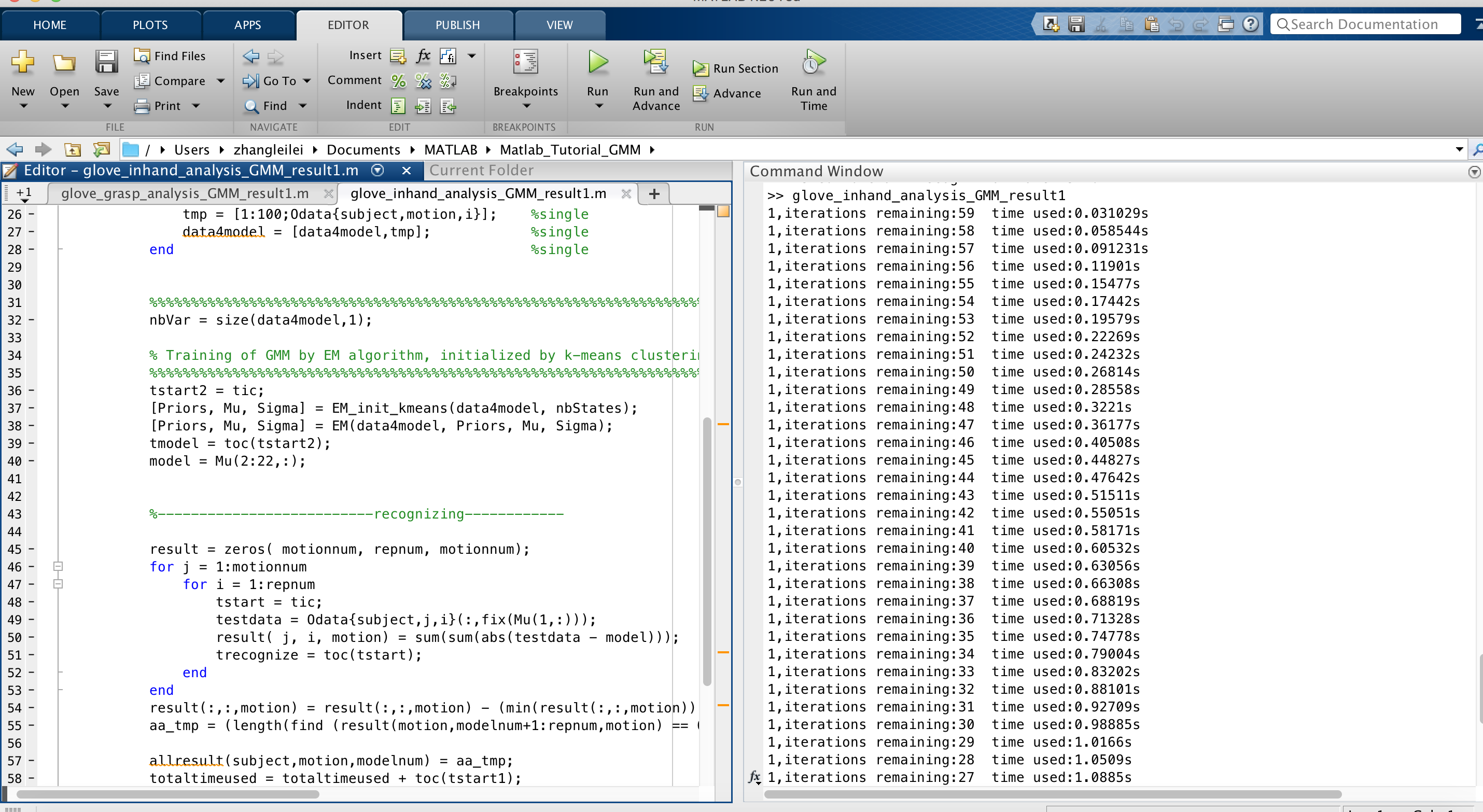


Fig1-Glove\_grasp\_analysis\_GMM\_result2



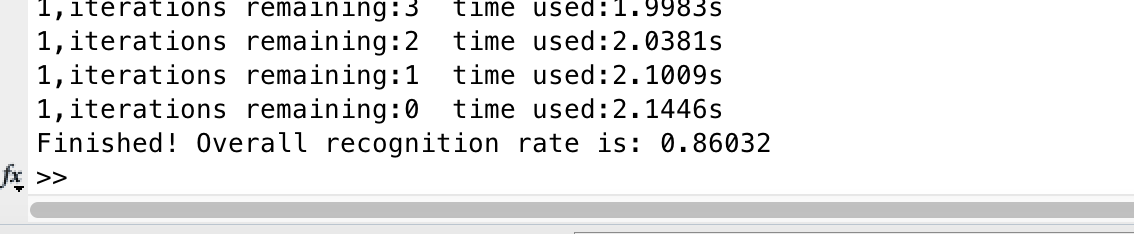
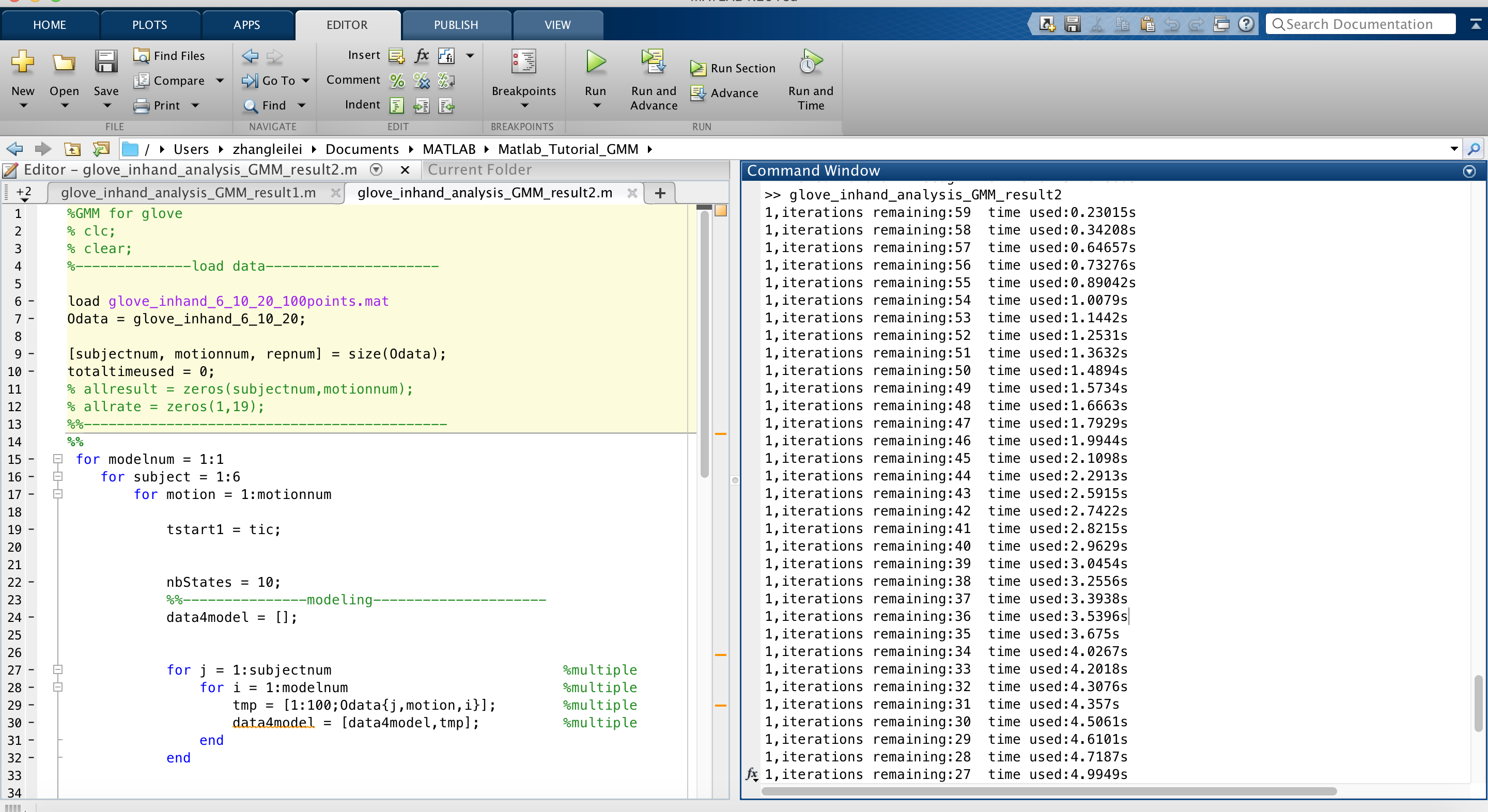


Fig2-Glove\_inhand\_analysis\_GMM\_result1



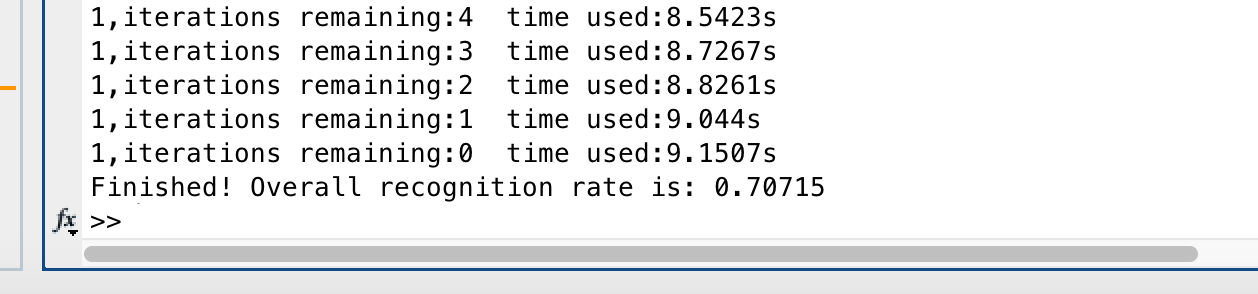


Fig3-Glove\_inhand\_analysis\_GMM\_result2

Word description：

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.

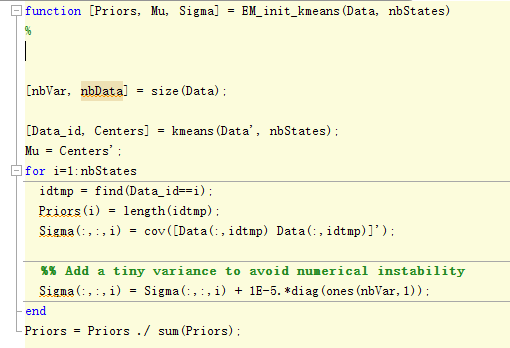


Fig. 2. The code of EM\_init\_kmeans

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.

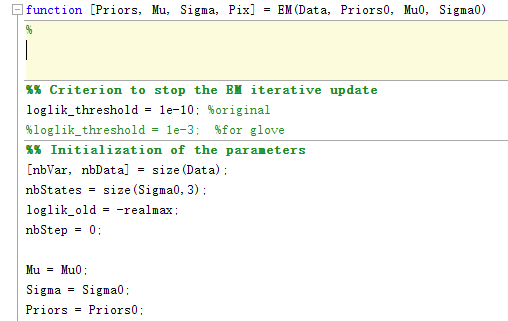


Fig.4. Initialize the Priors, Mu and Sigma

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

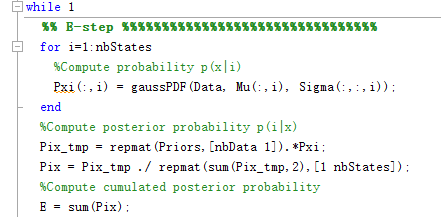


Fig. 5. E-step

E-step compute cumulated posterior probability.

Step two is M-step.

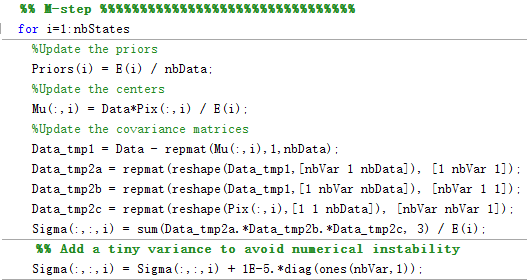


Fig. 6. M-step

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

Step three, judging whether the operation is stopped.

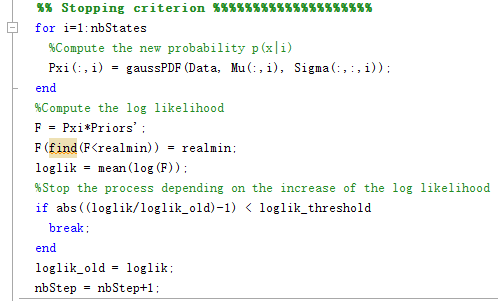


Fig. 7. Stopping criterion

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.

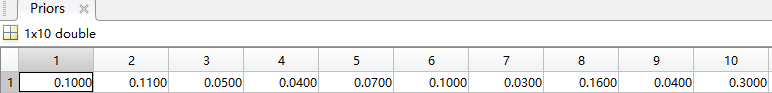


Fig. 8. The data of priors

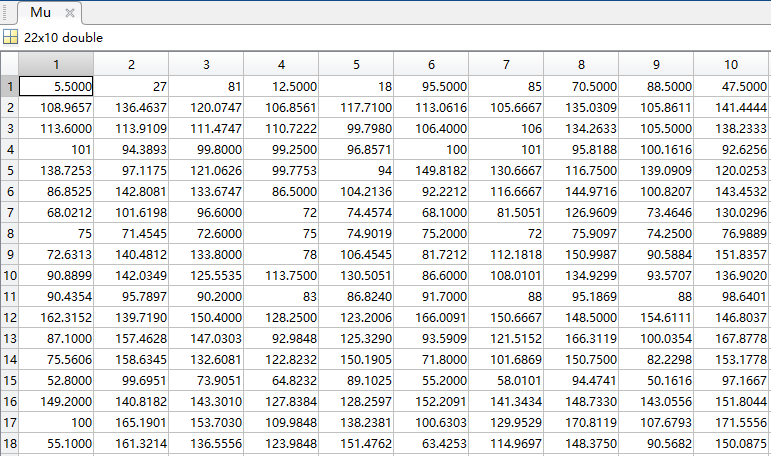


Fig. 9-a. The data of mu

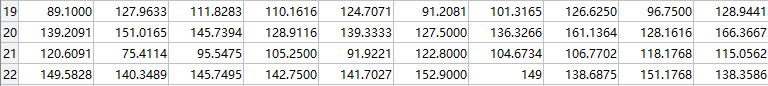


Fig. 9-b. The data of mu

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

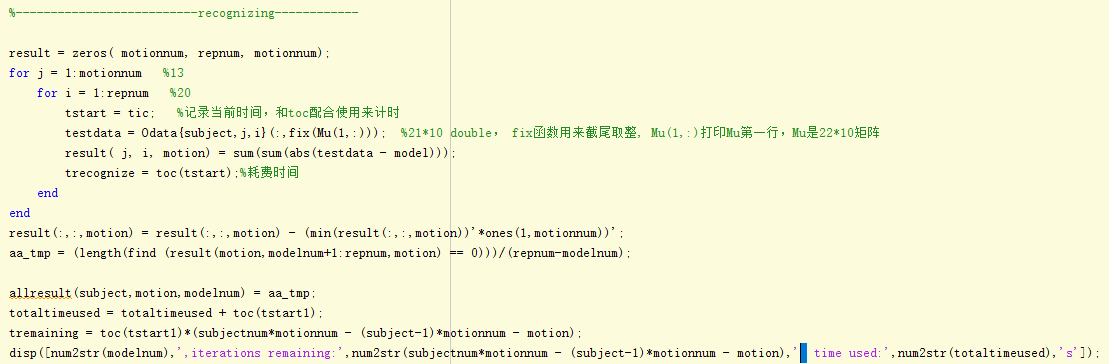


Fig. 10. The recognition step

# Understanding and Relationships of K-means, GMM and EM Algorithms

**K-means algorithm:**

The most classic partition-based clustering method is one of the ten classic data mining algorithms. The basic idea of ​​K-means algorithm is: to cluster k points in space and classify the objects closest to them. Through the iterative method, the values ​​of each cluster center are updated one by one until the best clustering result is obtained. The k-means algorithm accepts the parameter k; and then divides the n data objects input in advance into k clusters in order to make the obtained cluster satisfy: the objects in the same cluster have higher similarity; and the objects in different clusters Similarity is smaller. Clustering similarity is an algorithm that uses a center object (gravity center) obtained by averaging the objects in each cluster. First, k objects are randomly selected as the initial clustering centers from n data objects. The rest of the other objects are assigned to clusters that are most similar to them (represented by the cluster centers) based on their similarity (distance) to the cluster centers, and then each new cluster Of the cluster centers (the mean of all objects in the cluster); and repeat this process until the standard measure function begins to converge. Generally, the mean square error is used as a standard measure function, k clusters have the following characteristics: each cluster itself is as compact as possible, and the clusters are separated as much as possible.

**GMM algorithm：**

Gaussian model is to use Gaussian probability density function (normal distribution curve) to accurately quantify things, a thing is divided into several based on the Gaussian probability density function (normal distribution curve) to form the model. The principle and process of establishing Gaussian model for image background: The gray histogram of image reflects the occurrence frequency of a certain gray value in the image, and can also be regarded as the estimation of the gray probability density of the image. If the image contains a larger target area and a background area, and the background area and the target area have a certain difference in grayscale, the grayscale histogram of the image has a bimodal-valley shape, in which one peak corresponds to The target, the other peak corresponds to the center of the background gray. For complex images, especially medical images, are generally multi-peaked. By considering the multimodal characteristics of the histogram as a superposition of multiple Gaussian distributions, the image segmentation problem can be solved. In the intelligent monitoring system, the detection of the moving target is the central content, but in the moving target detection and extraction, the background target is very important for the target identification and tracking. Modeling is an important part of the background object extraction

GMM commonly used in clustering, if you want to take a random point from the distribution of GMM, can actually be divided into two steps: first randomly selected among these K Component, the probability of each Component is selected is actually Its coefficient πk, select the Component, and then consider the distribution of this Component alone to select a point on it - here has returned to the normal Gaussian distribution, into a known problem.。

**EM algorithm：**

Refers to the maximum expectation algorithm, an iterative algorithm used in statistics to find the maximum likelihood estimate of a parameter in a probabilistic model that depends on unobservable implicit variables. MLE estimation of parameters from non-complete data sets is a very simple and practical learning algorithm. This method can be widely used to deal with defect data, censored data, so-called incomplete data with noise. Suppose we estimate that both A and B parameters are known, both of them are unknown at the beginning, and we know B's information by knowing A's information. We can consider giving some initial value of A to get the estimated value of B, and then from the current value of B, re-estimate the value of A, the process continues until convergence.。

EM algorithm to solve the problem is: to cluster the data, assuming that the data obeys the heterozygosity of several probability distributions, the specific parameters of the distribution is unknown, the random variables involved are two groups, one of which can observe another group of unobservable . We now use 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 to the distribution parameter can be obtained by taking the maximum of the likelihood function。

First, the EM algorithm initializes the distribution parameters required to obtain the expectation of the hidden variables, and then maximizes the likelihood function of the distribution parameters by using the data of the hidden variables and the observable variables to obtain a set of solutions Thus updating the distribution parameters. Then, we use the updated distribution parameters to calculate the expectation of the implicit variables, and then use this expectation to update the distribution parameters with the observable data, that is, the EM algorithm includes the hidden variables E-step and the maximum likelihood function calculation M-step) two processes, initially assigning an initial value to the distribution parameters and then iteratively performing E-step and M-step until the algorithm converges

**The relationship between the three algorithms：**

The GMM algorithm is similar to the K-means algorithm and is used for clustering. The GMM algorithm is a concrete example of the family of EM algorithms. As an example of the family of EM algorithms, the GMM algorithm specifies that each of the participating hybrids has a Gaussian distribution, that is, the distribution parameters are represented by mean Mu and variance Sigma. By using the EM algorithm as a framework for calculation, the parameters of each Gaussian distribution are iteratively calculated. . In GMM, we define the number of components as K, which is the same as K in K-means, and we need another way to determine the value of K. In K-means, the initialization of each cluster center will affect the clustering effect. Similarly, the GMM algorithm is also sensitive to the initialization of the centroid of each component.

Comparing the K-Means algorithm with the GMM EM solution, we find that the two have strong similarities. The K-Means algorithm performs a "hard assignment" to the clustering of data points, that is, each data point belongs to a unique cluster. The EM solution of GMM is based on the posterior probability distribution and "softly allocates" data points. That is, each individual Gaussian model contributes to the data clustering, but the contribution is very small. In fact, we can classify the K-Means algorithm as a special case of EMM for GMM.

**The same point：**clustering effect similar to the initial sensitivity of the shortcomings are similar, the solution is similar, dichotomy. Is an iterative algorithm, and the iterative strategy is also the same: the algorithm starts the implementation of the need to calculate the initial parameters, and then alternately perform two steps, one step is to estimate the data (k-means is to estimate each Point belongs to the cluster; GMM is the expectation of calculating hidden variables;); The second step is to recalculate the parameter values ​​with the estimated values ​​calculated in the previous step and update the target parameters (k-means is to calculate the position of the cluster centers; GMM is to calculate each Gaussian distribution Center position and covariance matrix)

**Differences:** The GMM outputs the probability that a data point belongs to each class, using the maximum likelihood method to determine the class. In terms of rigor, using probability to describe the classification of data points, the GMM is clearly much better than the K-mean; the parameters to be calculated are different: k-means is the centroid position; GMM is the parameter for each Gaussian distribution; Is different: k-means is to calculate the mean value of the positions of all the elements in the current cluster; GMM is a probability-based algorithm that calculates the distribution parameters by calculating the maximum value of the likelihood function.