**Clustering analysis is a scientific and effective method to study the "clustering".**

The data obtained from the experimental tests are raw data，which is a non-classified, irregular and intricate variable . If we want to make it possible to reflect the regularity or special classification, clustering analysis is an ideal method , which can make the data or variables show a certain classification characteristics.

**What is the clustering method?**

Clustering is a process of classifying data into different classes or clusters, so objects in the same cluster have a large similarity, and objects between different clusters have a large degree of dissimilarity.

Cluster analysis is an exploratory analysis. In the process of classification, people do not need to give a classification standard in advance. Cluster analysis can be classified automatically from the sample data. Different methods used in cluster analysis often lead to different conclusions. Different researchers have clustered the same set of data, and the number of clusters is not necessarily consistent.

K-Means clustering algorithm is one of the best-known and most well-known clustering algorithms utilized in a variety of areas. The most important characteristic of the method is that it is unsupervised.

K-means has been widely applied to each Engineering and science, such as psychology, biology, medicine and so on.

K-means has a lot of advantages such as simple principle , easy way to implement and effective clustering results.

However , K-means also has many disadvantages such as sensitive outlier, more iterations and local optimal solution.

1.Read training data (T1, T2, .., Tn), where Tn is data training.

2. Choose a number of desired clusters, k .

3. Choose k starting points to be used as initial estimates

of the cluster centroids. These are the initial starting

values.

4. Calculate Distance using some distance for each

training data.

5. Examine each point in the given dataset and assign it

to the cluster whose centroid is nearest to it.

6. When each point is assigned to a cluster, recalculate the

new k centroids.

7. Repeat steps 3 and 4 until no point changes its cluster

assignment, or until a maximum number of passes

through the data set is performed.

**What is the difference between clustering and classification？**

Ok,the question is so wonderful.I remember that it is a problem my girlfriend turned to me,and I think she is the most intelligent and beautiful girl in the world so that I fall in love with her.

Classification is a way to label the object based on a certain standard, and then distinguish the classification according to the label.

On the contrary, clustering is a process to find out the reason why there is aggregation between things.

Firstly, category is defined in advance in Classification and the number of class is unchanged. There is no predefined category, and the number of classes is uncertain.

Next, the classifier of classification needs to be trained by manually labeled training data and the way is supervised. On the contrary, Clustering does not require manual labeling and pre-training of classifiers. Categories are automatically generated during clustering.