The Research on How to Determine the Sorts of Cluster

Du zi-fang
People's University of China
Peking, China

Xu wen
Municipal Social Economic Investigation of
Zhengzhou of Statistical Bureau of China
No. 74 Huzhu Road,
Zhengzhou China

Cluster analysis is a simple and widely used method of multivariate statistical analysis with the popularization of multivariate statistical analysis, high function computers and commonly used statistical analysis software. Cluster analysis is being widely used by more and more people in their research work. In cluster analysis, how to determine the sorts of cluster K can be regarded as a problem of this kind.
This essay discusses about this problem and gives us a specific method to solve it.

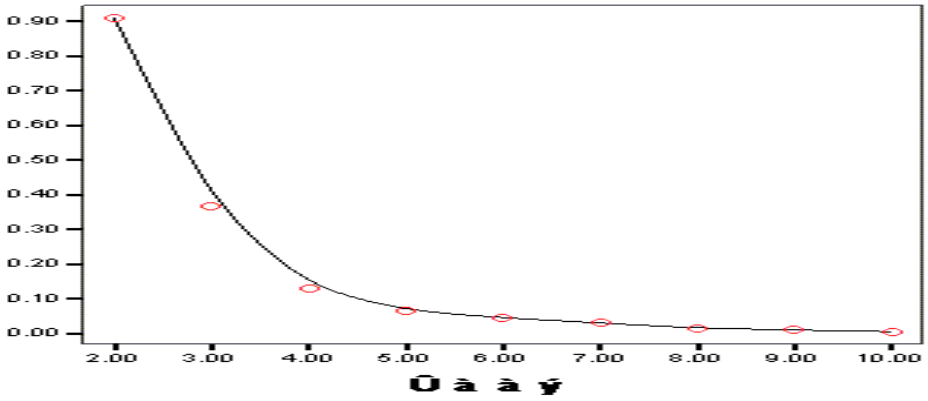
1. The proposal of the problem

1T》the form of [Case 1], the cluster problem in continuous data can give all the possible sorts' optimum cut (point) and the corresponding goal function value. We can draw a plane function figure similar to the main composition analysis in Scree Plot according to what was mentioned above.。
[Case 1] In order to understand the regularity of children's growth, we obtain the mean weight of a boy per year from his birth to eleven years old and it is shown in the following chart.：
Table 1:

Age	1	2	3	4	5	6	7	8	9	10	11
Increasedweight (kilos)	9.3	1.8	1.9	1.7	1.5	1.3	1.4	2	1.9	2.3	2.1

How can we classify age period by cluster analysis?

Using Fisher algorithm, first calculate all the possible sorts' diameter of the sum of internal variation in measurement sort.
Then calculate the minimum goal function related to the optimum cut (point) under the various conditions.
Finally take the goal function in last line above the chart as ordinate and the sorts of cluster analysis as abscissa to draw the diagram:
But how many sorts of keys should we choose at last?



2. The methods proposed in the paper

Method1try to get the key of K $D_u = \frac{\Delta e_{k-1} - \Delta e_k}{e_k}\ \delta\ \max$

Firstmakean(statisticalamount)indexfunctiontowellshowoutanddescribethemathematic character at the curve turning point and the smooth tendency curve. In this connection, considering the monotone decreasing of a e_k so, there is no turning point in opposite direction: the differences at the curve turning point of b, will be relatively greater than others between its neighbors. By using the elastic concept of function to form $\frac{\Delta e_{k-1}}{e_{k-1}} \Diamond \frac{\Delta e_k}{e_k}$ and $\Delta\Delta e_k = \Delta e_{k-1} - \Delta e_k$,then simplify it into $D_u = \frac{\Delta e_{k-1} - \Delta e_k}{e_k}$

When we get the greatest value, the index function can explain that the differences are relatively greater than others between its neighbors.

In order to prove this property of incex function, to use some practical data to compare with:

[Case1]to get the key of K $D_u = \frac{\Delta e_{k-1} - \Delta e_k}{e_k}\ \delta\ \max_{=3}$

Table 2

k	2	**3**	4	5	6	7	8	9	10
e	0.909	**0.368**	0.128	0.065	0.045	0.03	0.015	0.01	0.005

[Case2]to get the key of K $D_u = \frac{\Delta e_{k-1} - \Delta e_k}{e_k}\ \delta\ \max_{=3}$

[Case3]to get the key of K $D_u = \frac{\Delta e_{k-1} - \Delta e_k}{e_k}\ \delta\ \max_{=3(\text{or } k=6)}$

[Case4]to get the key of K $D_u = \frac{\Delta e_{k-1} - \Delta e_k}{e_k}\ \delta\ \max_{=5}$