

Note on Clustering Criterion

Shang Gao^{1, a *} and Hong Mei Li^{1, b}

¹School of Computer Science and Engineering, Jiangsu University of Science and Technology, Zhenjiang 212003, China

^agao_shang@just.edu.cn, ^bLihongmei@just.edu.cn

* please mark the corresponding author with an asterisk

Keywords: Clustering problem; Criterion; Classification; Pattern recognition

Abstract. The principle of the sum of square errors criterion is introduced. The most common criterion is the sum of squared errors criterion, the weighted sum of square of the distance criterion and the sum of distance between-class distance criterion. The shortage of the sum of square errors criterion is point out. The sum of absolute errors criterion, maximum error criterion and the sum of power of errors criterion are proposed. The results are illustrated by numerical examples.

Introduction

Cluster analysis or clustering is the assignment of a set of observations into subsets (called clusters) so that observations in the same cluster are similar in some sense. Clustering ([1]-[3]) is a method of unsupervised learning, and a common technique for statistical data analysis used in many fields, including machine learning, data mining, pattern recognition, image analysis, information retrieval, and bioinformatics. The joining or tree clustering method uses the dissimilarities (similarities) or distances between objects when forming the clusters. Similarities are a set of rules that serve as criteria for grouping or separating items. These distances (similarities) can be based on a single dimension or multiple dimensions, with each dimension representing a rule or condition for grouping objects ([4]-[11]). For example, if we were to cluster fast foods, we could take into account the number of calories they contain, their price, subjective ratings of taste, etc. The most straightforward way of computing distances between objects in a multi-dimensional space is to compute Euclidean distances. If we had a two- or three-dimensional space this measure is the actual geometric distance between objects in the space (i.e., as if measured with a ruler). However, the joining algorithm does not "care" whether the distances that are "fed" to it are actual real distances, or some other derived measure of distance that is more meaningful to the researcher; and it is up to the researcher to select the right method for his/her specific application. The most common criterion is the sum of squared errors criterion, the weighted sum of square of the distance criterion and the sum of distance between-class distance criterion. The shortage of the sum of square errors criterion is point out. The sum of absolute errors criterion, maximum error criterion and the sum of power of errors criterion are proposed.

The Sum of Square Errors Criterion

The clustering criteria should reflect the similarity between classes, or separation of functions. If the clustering criterion function is well chosen, clustering quality will be high.

Given a set of observations $X = \{x_1, x_2, \dots, x_n\}$, where each observation is a d-dimensional real vector, k-means clustering aims to partition the n observations into k sets ($k \leq n$) $S = \{S_1, S_2, \dots, S_k\}$ so as to minimize the within-cluster sum of squares (WCSS) ([1]-[3]):

$$\min J_2 = \sum_{j=1}^c \sum_{x \in S_j} \|x - m_j\|^2 \quad (1)$$

where m_j is the mean of points in S_j , and $m_j = \frac{1}{n_j} \sum_{x \in S_j} x$, $j = 1, 2, \dots, c$.

Sum of squared error criterion is to find J_2 to the minimum. Sum of squared error criterion applicable to all types of samples which are Intensive and the numbers of samples are almost same, and different kinds of samples are clearly separated from the case. It is shown in Figure 1. The distribution of the sample consists of 3 types, each type of sample numbers are not much difference (10 or so). The he sample are more intensive, sum of square error is smaller, and distances between the categories are farther.

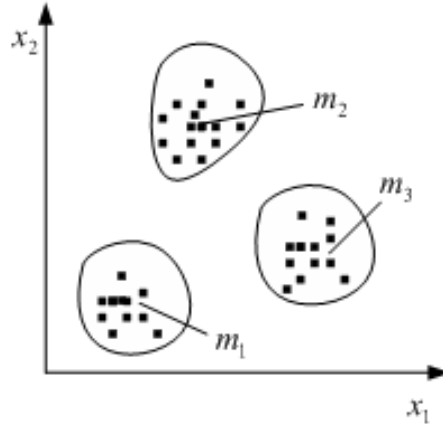
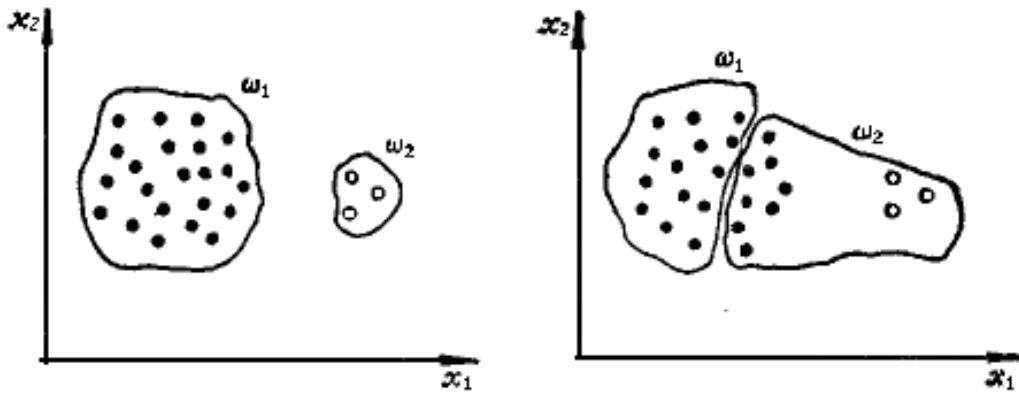


Figure. 1 Pattern classification diagram

If the difference of the numbers of different set are great, it is possible to separate sample number of types in error using the sum of squared error criterion in order to achieve the total minimum. As shown in Figure 2, the number of class ω_1 is more than the class ω_2 . In accordance with the sum of squared error criterion, it is possible to separate part of class ω_1 to class ω_2 . In this clustering, the he sum of squared error may be smaller.



(a) The correct classification

(b) The error classification

Figure. 2 Comparison of Classification

For example: a sample, there are 5 samples as shown in Figure 3. From the figure 3, we can get $x_1 \sim x_4 \in \omega_1$ and $x_5 \in \omega_2$. The correct classification is shown by the dotted line. The result is shown by the solid line using the least sum of square error criterion.

If the samples are divided by the dotted line, the sum of square error is calculated as following.

$$\omega_1 : X_1 = \{x_1, x_2, x_3, x_4\}, \quad m_1 = \frac{1}{4} \sum_{k=1}^4 x_k = (0, 0)^T$$

$$\omega_2 : X_2 = \{x_5\}, \quad m_2 = x_5 = (2, 0)^T$$

$$J_2^{(1)} = \sum_{x_k \in X_1} \|x_k - m_1\|^2 + \sum_{x_k \in X_2} \|x_k - m_2\|^2 = 4$$

If the samples are divided by the solid line, the sum of square error is calculated as following.

$$\omega_1 : X_1 = \{x_1, x_2, x_3\}, \quad m_1 = \frac{1}{3} \sum_{k=1}^3 x_k = \left(-\frac{1}{3}, 0\right)^T$$

$$\omega_2 : X_2 = \{x_4, x_5\}, \quad m_2 = \frac{1}{2} \sum_{k=4}^5 x_k = (1.5, 0)^T$$

$$\begin{aligned} J_2^{(2)} &= \sum_{x_k \in X_1} \|x_k - m_1\|^2 + \sum_{x_k \in X_2} \|x_k - m_2\|^2 \\ &= \frac{8}{3} + \frac{1}{2} = \frac{19}{6} = 3.17 \end{aligned}$$

So, $J_2^{(1)} > J_2^{(2)}$. If we use the sum of square error criterion, we will get the wrong result.

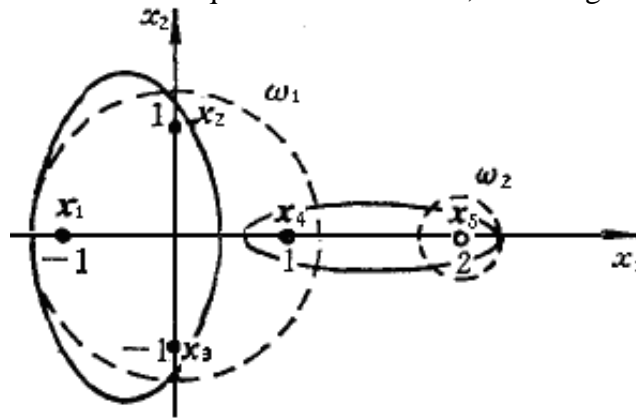


Figure 3. A example of clustering

Other Clustering Criterion and Discussion

Since the sum of squared error criterion are sometimes not perfect, we discuss the other clustering criterions. From the norm defined view, the sum of absolute errors criterion and the maximum error criterion are given.

The sum of absolute errors criterion is

$$J_1 = \sum_{j=1}^c \sum_{x \in S_j} \|x - m_j\| \quad (2)$$

The maximum error criterion is

$$J_\infty = \max_{x \in S_j} \|x - m_j\| \quad (3)$$

In the above example, if the samples are divided by the dotted line, $J_1^{(1)} = 4$ and $J_\infty^{(1)} = 1$.

If the samples are divided by the solid line, $J_1^{(2)} = \frac{5 + 2\sqrt{10}}{3} < J_1^{(1)}$ and $J_\infty^{(2)} = \frac{\sqrt{10}}{3} > J_\infty^{(1)}$.

According to the sum of absolute errors criterion, it will also get the wrong result. According to the maximum error criterion it will gain the right result. It shows that using the different criteria, classification results are different.

The general clustering criterion can be defined as

$$J_p = \sum_{j=1}^c \sum_{x \in S_j} \|x - m_j\|^p \quad (4)$$

This criterion can be called the sum of P power of errors criterion.

In the above example $J_p^{(1)} = 4$, $J_p^{(2)} = 2(\frac{\sqrt{10}}{3})^p + (\frac{1}{3})^p + 2(\frac{1}{2})^p$.
 If $p \geq 14$ then $J_p^{(2)} > J_p^{(1)}$. It shows that it will gain the right result if $p \geq 14$.

Acknowledgment

This work was sponsored by Artificial Intelligence of Key Laboratory of Sichuan Province (2012RYJ04) , Jiangsu 333 Project , Qing Lan Project and the National Natural Science Foundation of China under Grant and the National Natural Science Foundation of China under Grant 61305058.

References

- [1] Z. H. Huang and C. Y. Wu. Pattern recognition. Zhejiang University Press, 1991.pp.40-62 (in Chinese).
- [2] Y. L. Cai. Pattern recognition. Northwest Telecommunications Industry Press,1986. pp.17-32(in Chinese).
- [3] R.O. Duda, P.E. Hart, and D.G. Stork. Pattern Classification. John Wiley & Sons, 2001.
- [4] B. Stanisaw. A validity criterion for fuzzy clustering. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), v 8342, p 132-151, 2014.
- [5] C. Carlos, M. Henry; U. Richar, and M. Martha etc . Enrique. Clustering of web search results based on the cuckoo search algorithm and Balanced Bayesian Information Criterion; Information Sciences, v 281, p 248-264, October 10, 2014.
- [6] H. Zeng and Y.M. Cheung. Learning a mixture model for clustering with the completed likelihood minimum message length criterion. Pattern Recognition, v 47, n 5, p 2011-2030, May 2014.
- [7] W.G. Sheng, S. Y. Chen and F. Michael, etc. Multilocal search and adaptive Niching based Memetic algorithm with a consensus criterion for data clustering. IEEE Transactions on Evolutionary Computation, v 18, n 5, p 721-741, October 1, 2014.
- [8] H.R. Li, H. L. Deng, and Q. F. Wu, etc. A novel K-means clustering based on I-divergence criterion. Journal of Computational Information Systems, v 9, n 5, p 2017-2024, March 1, 2013.
- [9] K. Hoyt and C. Bertrand. A bayesian criterion for cluster stability.Statistical Analysis and Data Mining, v 6, n 4, p 346-374, August 2013.
- [10] H. L. Deng, H.R. Li, and Q. F. Wu, etc A novel K-means clustering based on max entropy criterion.ICIC Express Letters, v 7, n 8, p 2243-2248, 2013.
- [11] B. Soheila and E. Mehdi, S.Hamid. A new feature ranking criterion based on density function of subtractive clustering.13th Iranian Conference on Fuzzy Systems, IFSC 2013, 2013, 13th Iranian Conference on Fuzzy Systems, IFSC 2013.