The Moment Preservation Method of Cluster Analysis

Bernard Harris
Department of Statistics, University of Wisconsin
1210 W. Dayton Street
Madison, Wisconsin, 53706, U.S.A.
harris@stat.wisc.edu

1. Introduction and Summary.

The moment preservation method is a technique for cluster analysis and classification, which appears in the engineering literature. Its proponents claim that it is effective and computationally efficient. However, they do not appear to be cognizant of its mathematical and statistical properties.
The purpose of this report is to describe such properties. Essentially in using this procedure, one is determining solutions of the classical reduced moment problem. Such solutions have many useful attributes, including the maximization and minimization of certain types of linear functionals. This may be employed to optimize clustering and classification procedures under certain conditions. Some generalizations of the present technique are also readily implemented using the mathematical and statistical properties.

2. The Moment Preservation Method.

In this section, the present form of the moment preservation method is described for univariate data. Let $x_1, x_2, \ldots, x_n$ be n univariate data points from a distribution which is absolutely continuous. To use this procedure, the number of clusters must be chosen in advance. Assume that q ($q \geq n$) clusters are desired. Then, the first 2q-1 non-central sample moments are calculated. The unique discrete distribution with positive probability at q points, which possesses these moments, is determined. Thus for two clusters, the first three sample moments are calculated, determining the two point distribution, ($p_1, y_1; 1-p_1, y_2$, $0 < p_1 < 1$, $y_1 < y_2$). In general, the solution is given by (p_i, y_i, $i = 1,2, \ldots, q$: $\Sigma\, p_i = 1$, $y_1 < y_2 < \ldots < y_q$) ; y_i is the center of the ith cluster. The data points $\{x_i$, I $= 1,2,\ldots,n\}$ are assigned to the q clusters as follows. Arrange the n data points in increasing order. The observations, whose rank does not exceed np_1 are placed in the first cluster. From the remaining observations, those whose rank does not exceed $n(p_1 + p_2)$ are placed in the second cluster. This process is continued until all observations have been assigned to a cluster.

3. The Reduced Moment Problem and Its Solutions.

Let $\mu_1, \mu_2, \ldots, \mu_k$ be a realizable sequence of the first k non-central moments , that is, there exists a probability distribution for which these are the moments. The determination of the set of possible probability distributions on $a \leq x \leq b$ with these moments is called the (classical) reduced moment problem on [a,b] , ($a > -\infty$, $b < \infty$). There is an extensive literature dealing with solutions of this problem and with various extensions and modifications of it. Space limitations prevent even a cursory discussion of this at the present moment. However, a few basic results that are relevant will be noted.

Each of the following is a convex set

(1). The set of all probability distributions,
(2) The sets of probability distributions with finite support,
(3). The set of all discrete probability distributions.

For each of the above convex sets, the extreme points are the degenerate (one-point) distributions. In fact, the set of all distributions is the weak closure of the set of distributions with finite support. Essentially, all distributions are "limits" of finite convex combinations of the degenerate distributions.

(4). The set of realizable moments is a convex set.

(5). The set of distributions with a given reduced moment sequence is a convex set.

The extreme points of the set of realizable moments of probability distributions on [a,b] are the vectors $(d,d^2,..,d^k)$, $a \le d \le b$, that is, the moment sequences for the degenerate distributions.

Applying representation theorems based on the theory of finite dimensional convex sets, in particular, a well-known theorem of Caratheodory, it is possible to obtain parametric geometric representations of the solutions and hence candidates for clustering. The details are necessarily omitted here.

4. Final Remarks.

The theory outlined in this summary does not specifically require the use of the monomials, $x, x^2,..,x^k$. Any set of k linearly independent functions can be employed. However, if the functions are strictly convex, as is the case with monomials, with the exception of x, the computations are simplified. In addition, it is possible to use these representations to solve some variational or extremal problems, which can arise in cluster analytical problems. Let g(x) be an integrable function on a convex set of probability distributions. The objective is to maximize (minimize)

$\int g(x)\,dx$ over the solution set of the reduced moment problem and to characterize the distribution which attains this maximum (minimum). Several methods for solving this are available. The potential application to cluster analysis applies if g(x) is some measure of merit of the clustering, such as the probability of misclassification. This calculation is simplest when g(x) is a well-behaved function; in particular, if g(x) is completely monotonic or absolutely monotonic. In general, solutions to such extremal problems are attained by probability distributions with small carrier sets.

Extensions to multivariate data can be similarly developed and will not be presented here.

Sommaire.

Une methode pour construction les grappes est se presente′. Cet methode est par ingenieurs utilise′. In cet compte, traits caracte′ristiques mathematiques sont presente′.