
The p-k-median location problem: clustering data with respect to several patterns within each cluster.

Carlos Martín*^{†1} and Justo Puerto¹

¹Departamento de Estadística e Investigación Operativa (DESTIO) – Spain

Abstract

This paper introduces a new location problem with application in clustering data using more than one representative pattern. We assume to be given a set of demand points (set of data) and we wish to locate p configurations of k points each, in order to minimize the overall sum of distances of the demand points to the k points in their closest configuration. We consider the continuous and discrete versions of this problem. In the continuous case, the p configurations can be placed anywhere in a continuous space endowed with an l_p -norm. The discrete version assumes that the feasible locations for the configurations must belong to a given finite sets of points, from where the model has to choose the p configurations. The first problem can be formulated as a mixed integer non-linear programming problem whereas the second one is a mixed integer linear program. These problems generalize several well-known facility location problems as the p -median and round trip location problem. We study some properties of these models and give solution approaches to obtain their solutions. We present preliminary computational results, implemented in MOSEK, on several databases having in mind their application to cluster data with respect to k representative patterns rather than with respect to only one, as in the p -mean or p -median models.

Keywords: k , median, prototype based clustering, facility location, conic programming

*Speaker

[†]Corresponding author: carlosvalverdemartin@gmail.com