

A Complete Gradient Clustering Algorithm Formed with Kernel Estimators

Piotr Kulczycki

Polish Academy of Sciences, Systems Research Institute, Center for Stochastic Data Analysis Methods,
ul. Newelska 6, PL-01-447 Warsaw, Poland
Cracow University of Technology, Department of Automatic Control, ul. Warszawska 24, PL-31-155 Cracow, Poland
kulczycki@ibspan.waw.pl ; kulczycki@pk.edu.pl

Malgorzata Charytanowicz

Polish Academy of Sciences, Systems Research Institute, Center for Stochastic Data Analysis Methods,
ul. Newelska 6, PL-01-447 Warsaw, Poland
Catholic University of Lublin, Institute of Mathematics, al. Raclawickie 14, PL-20-950 Lublin, Poland
malgorzata.charytanowicz@ibspan.waw.pl ; mchmat@kul.lublin.pl

Abstract — The aim of this paper is to provide a gradient clustering algorithm in its complete form, suitable for direct use without requiring a deeper statistical knowledge. The values of all parameters are effectively calculated using optimizing procedures. Moreover, an illustrative analysis of the meaning of particular parameters is shown, followed by the effects resulting from possible modifications with respect to their primarily assigned optimal values. The proposed algorithm does not demand strict assumptions regarding the desired number of clusters, which allows the number obtained to be better suited to a real data structure. Moreover, a feature specific to it is the possibility to influence the proportion between the number of clusters in areas where data elements are dense as opposed to their sparse regions. And finally, the algorithm – by the detection of one-element clusters – allows the identification of atypical elements, which enables their elimination or designation to more numerous clusters, thus increasing the homogeneity of the data set.

Keywords: data analysis and mining, clustering, gradient procedures, nonparametric statistical methods, kernel estimators, numerical calculations.