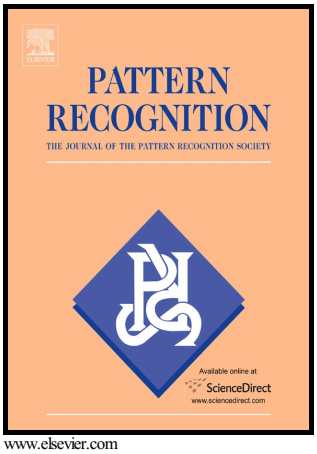
Kernel-based hard clustering methods with kernelization of the metric and automatic weighting of the variables

Marcelo R.P. Ferreira, Francisco de A.T. de Carvalho, Eduardo C. Simões



www.elsevier.com

PII: S0031-3203(15)00356-8
DOI: http://dx.doi.org/10.1016/j.patcog.2015.09.025
Reference: PR5529

To appear in: *Pattern Recognition*

Received date: 4 September 2014
Revised date: 18 September 2015
Accepted date: 21 September 2015

Cite this article as: Marcelo R.P. Ferreira, Francisco de A.T. de Carvalho and Eduardo C. Simões, Kernel-based hard clustering methods with kernelization of the metric and automatic weighting of the variables, *Pattern Recognition* http://dx.doi.org/10.1016/j.patcog.2015.09.025

Kernel-based hard clustering methods with kernelization of the metric and automatic weighting of the variables

Marcelo R.P. Ferreira[a,*], Francisco de A.T. de Carvalho[b], Eduardo C. Simões[b]

[a] *Departamento de Estatística, Centro de Ciências Exatas e da Natureza, Universidade Federal da Paraíba, CEP 58051-900 - João Pessoa (PB) - Brazil*
[b] *Centro de Informática, Universidade Federal de Pernambuco, Av. Jornalista Anibal Fernandes, s/n - Cidade Universitária, CEP 50740-560 - Recife (PE) - Brazil*

Abstract

This paper presents kernel-based hard clustering methods with kernelization of the metric and automatic weighting of the variables. The proposed methodology is supported by the fact that a kernel function can be written as a sum of kernels evaluated separately on each variable. Thus, in the proposed algorithms dissimilarity measures are obtained as sums of Euclidean distances between patterns and centroids computed individually for each variable by means of kernels. The main advantage of this approach over the conventional approach is that it allows the use of kernelized adaptive distances, which are suitable to learn the weights of the variables dynamically, improving the performance of the algorithms. Moreover, various partition and cluster interpretation tools are introduced. Experiments with a number of benchmark datasets corroborate the usefulness of the proposed algorithms and the merit of the partition and cluster interpretation tools.

Keywords: Kernel clustering, Kernelization of the Metric, Weighting of the variables, Adaptive distances.

*Corresponding Author. tel.:+55-83-32167785; fax:+55-83-32167785
Email addresses: marcelo@de.ufpb.br (Marcelo R.P. Ferreira), fatc@cin.ufpe.br (Francisco de A.T. de Carvalho), ecs4@cin.ufpe.br (Eduardo C. Simões)