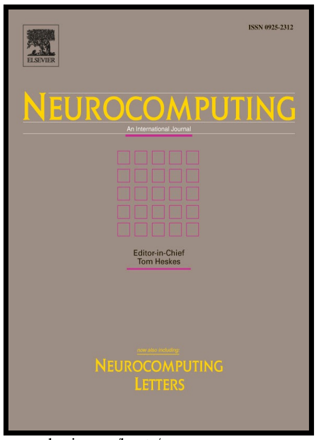
Author’s Accepted Manuscript



www.elsevier.com/locate/neucom

R Package CEC

P. Spurek, K. Kamieniecki, J. Tabor, K. Misztal, M. Śmieja

PII: S0925-2312(16)31443-6
DOI: http://dx.doi.org/10.1016/j.neucom.2016.08.118
Reference: NEUCOM17798

To appear in: *Neurocomputing*

Received date: 18 December 2015
Revised date: 29 July 2016
Accepted date: 3 August 2016

Cite this article as: P. Spurek, K. Kamieniecki, J. Tabor, K. Misztal and M. Śmieja, R Package CEC, *Neurocomputing*, http://dx.doi.org/10.1016/j.neucom.2016.08.118

R Package CEC

P. Spurek, K. Kamieniecki, J. Tabor, K. Misztal, M. Śmieja

Faculty of Mathematics and Computer Science, Jagiellonian University, Łojasiewicza 6, 30-348 Kraków, Poland.

Abstract

Cross-Entropy Clustering (CEC) is a model-based clustering method which divides data into Gaussian-like clusters. The main advantage of CEC is that it combines the speed and simplicity of k-means with the ability of using various Gaussian models similarly to EM. Moreover, the method is capable of the automatic reduction of unnecessary clusters. In this paper we present the **R** Package **CEC** implementing CEC method.

Keywords: clustering, Gaussian models, density estimation, R package

1. Introduction

Gaussian Mixture Model (GMM) is one of the most popular parametric clustering models implemented in various R packages, such as **mclust** [10], **Rmixmod** [11], **pdfCluster** [2], **mixtools** [3], etc. The model focuses on finding the mixture of Gaussians $f = p_1 f_1 + \ldots + p_k f_k$, where $p_1, \ldots, p_k > 0$ and $\sum_i p_i = 1$, which provides an optimal estimation of data set $X \subset \mathbb{R}^N$, measured by the negative log-likelihood cost function:

$$\mathrm{EM}(f, X) = -\frac{1}{|X|} \sum_{x \in X} \log \left(p_1 f_1(x) + \ldots + p_k f_k(x) \right), \tag{1}$$

where $|X|$ denotes the cardinality of X. Its minimization is iteratively performed with use of EM (Expectation Maximization) algorithm [5]. While the expectation step is relatively simple, the maximization step usually needs

[1]The work was supported by the National Centre of Science (Poland) [grants no. 2013/09/N/ST6/01178, 2014/13/B/ST6/01792, 2012/07/N/ST6/02192, 2014/13/N/ST6/01832].