POISSON DISTRIBUTION BASED INITIALIZATION FOR FUZZY CLUSTERING

Tomas Vintr[*], Vanda Vintrova[†], Hana Rezankova[‡]

Abstract: A quality of centroid-based clustering is highly dependent on initialization. In the article we propose initialization based on the probability of finding objects, which could represent individual clusters. We present results of experiments which compare the quality of clustering obtained by k-means algorithm and by selected methods for fuzzy clustering: FCM (fuzzy c-means), PCA (possibilistic clustering algorithm) and UPFC (unsupervised possibilistic fuzzy clustering) with different initializations. These experiments demonstrate an improvement in the quality of clustering when initialized by the proposed method. The concept how to estimate a ratio of added noise is also presented.

1. Introduction

Cluster analysis is represented by a set of multi-dimensional data analysis techniques used for object classification when objects are characterized by values of selected attributes. Methods of cluster analysis belong to a class of methods called *unsupervised learning*. Their aim is to divide a set of objects into subsets so that the similarity of objects within a subset and the difference of objects from different subsets are maximized. These subsets are called *clusters*.

---

[*]Tomas Vintr
University of Economics, Prague, Faculty of Informatics and Statistics, Department of Statistics and Probability, E-mail: `tomas.vintr@vse.cz`
[†]Vanda Vintrova
University of Economics, Prague, Faculty of Informatics and Statistics, Department of Statistics and Probability, E-mail: `vandav@email.cz`
[‡]Hana Rezankova
University of Economics, Prague, Faculty of Informatics and Statistics, Department of Statistics and Probability, E-mail: `rezanka@vse.cz`