

Title: k-means method

Author: Jana Hricová

Department: Department of Probability and Mathematical Statistics

Supervisor: prof. RNDr. Jaromír Antoch, CSc., Department of Probability and Mathematical Statistics

Abstract:

This thesis deals with the statistical method k-means, which is a part of an extensive set of methods and algorithms designed for cluster analysis of data. Results of the cluster analysis are widely used in other scientific activities, but also in marketing, management or in insurance etc. Statistical methods for cluster analysis are creating clusters from analyzed datasets, which consist of similar objects. Similarity of two objects is expressed by dis-/similarity measure.

The aim of this thesis was to introduce the k-means algorithm. This is a non-hierarchical method with given number of output clusters as input. We have applied this algorithm in the environment of mathematical software Matlab on simulated and real data and have interpreted the results using graphical and numerical outputs.

Keywords: k-means, cluster analysis, dissimilarity measure, silhouette