

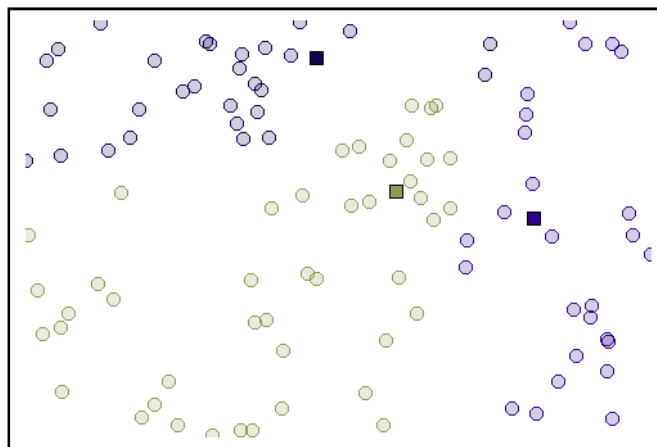
Foundations of Information Retrieval

Exercise 9

Issue date: 12.12.2013
Exercise session on: 09.01.2014
Questions to: Elena Demidova: demidova@L3S.de

Clustering

1. The figure below describes a state of the k -means algorithms with $k=3$. In the figure, the squares represent centroids and circles represent the data points. The color encoding corresponds to the current cluster assignment. Sketch the changes that will be performed by the k -means algorithm in the next step.



2. The documents that use different terms (e.g. automobile, VW Golf, etc.) for one concept (i.e. car) will be most probably assigned to the same cluster if using term vectors for clustering. Explain why.
3. Explain similarities and differences between classification and clustering. Pay specific attention to the following aspects:
 - a) Training of the system
 - b) Number of cluster / classes
 - c) Presentation of the results