Practical Machine Learning in R

Clustering

Lars Kotthoff\textsuperscript{1,2}
larisko@uwyo.edu

\textsuperscript{1}with slides from Bernd Bischl and Michel Lang
\textsuperscript{2}slides available at http://www.cs.uwyo.edu/~larisko/ml-fac
Unsupervised Clustering

Goal: Group data by similarity, or estimate membership probabilities
$k$-means clustering

- pick $k$ cluster centers randomly
- assign each data point to a cluster by shortest mean distance
- centroid (point with smallest mean distance to all points) of each cluster becomes new center
- repeat until convergence
\( k \)-means clustering

- easy to understand, runs quickly
- need to specify number of clusters
- clusters are spherical
$k$-means clustering
$k$-means clustering
$k$-means clustering

Iteration #2
$k$-means clustering

Iteration #3
$k$-means clustering
$k$-means clustering
$k$-means clustering
$k$-means clustering

Iteration #7
$k$-means clustering
$k$-means clustering

Iteration #9
$k$-means clustering
$k$-means clustering

Iteration #11
$k$-means clustering

Iteration #12
$k$-means clustering

Iteration #13
$k$-means clustering

Iteration #14

By Chire - Own work, GFDL, https://commons.wikimedia.org/w/index.php?curid=59409335
EM – Expectation Maximization

- maximize likelihood of clusters, given data
- estimate distribution of data as mixture of distributions
- compute *expectation* of clusters for fixed model
- determine model parameters that *maximize* fixed clusters
  - repeat until convergence
- can determine number of clusters automatically
EM – Expectation Maximization

Different cluster analysis results on "mouse" data set:

Original Data  k-Means Clustering  EM Clustering
DBScan

- density-based clustering
- find core points (with a large number of neighbors)
- find connected core points, and which core points other points are assigned to
- number of clusters and shape determined automatically
- need to specify minimum number of points in a cluster and density threshold
DBScan
Exercises

http://www.cs.uwyo.edu/~larsko/ml-fac/03-clustering-exercises.Rmd