Ludwig-Maximilians-Universität München
Institut für Informatik
Prof. Dr. Matthias Schubert
Anna Beer

Munich, May 31st, 2017

Managing Massive Multiplayer Online Games

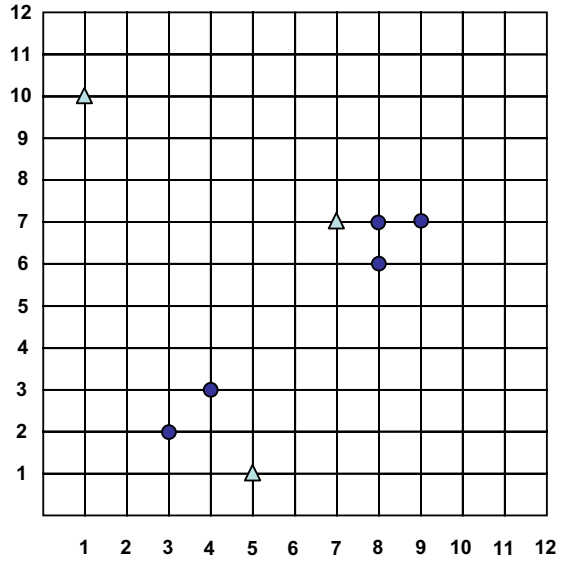
SS 2017

Exercise Sheet 6: Knowledge Discovery and Data Mining

Discussion: June 7th, 2017

Exercise 6-1 *Supervised Learning: Instance-based learning: classification with kNN*

The following data set with 8 points (e.g. two-dimensionally feature vectors) is given. The triangles build one class and the circles build the other.



In the following the classes of data points should be determined with the k-nearest neighbors algorithm. As distance function between two points the Manhattan distance (l_1 norm) shall be used:

$$L_1(x, y) = \sum_{i=1}^{d} |x_i - y_i|$$

(a) Determine the class of point (2,7) for $k = 2$ using the class of majority of its k-nearest neighbors, i.e. the point is assigned to the class which occurs most often among its k-nearest neighbors.

(b) Determine the class of point (2,7) for $k = 3$ using the class of majority of its k-nearest neighbors.

(c) Determine the class of point (2,7) for $k = 5$ using the class of majority of its k-nearest neighbors.

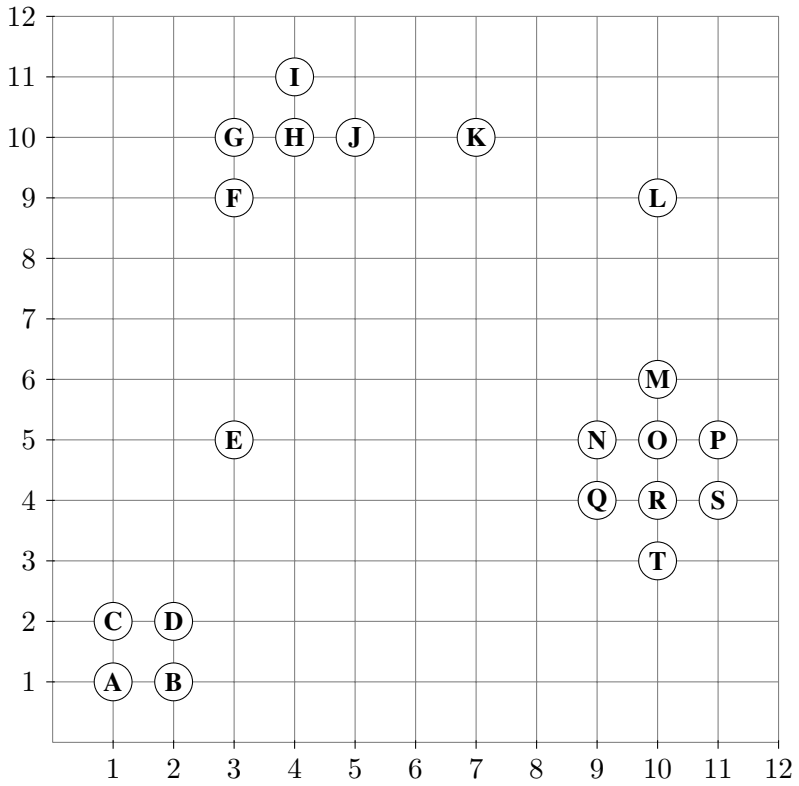
(d) Determine the class of point (6,1) for $k = 3$ using the class of majority of its k-nearest neighbors.

(e) Determine the class of point (6,1) for $k = 3$ using the class of majority of its k-nearest neighbors weighting the classes with inverse Manhattan distance.

$$L_1(x, y)^{-1} = \frac{1}{\sum_{i=1}^{d} |x_i - y_i|}$$

Exercise 6-2 *Unsupervised Learning: Clustering with DBSCAN*

The following dataset is given:



Cluster this dataset using DBSCAN. Use the Manhattan distance as distance function and the parameters $\epsilon = 1.1$ and $minPts = 3$.