Ludwig-Maximilians-Universität München
Institut für Informatik
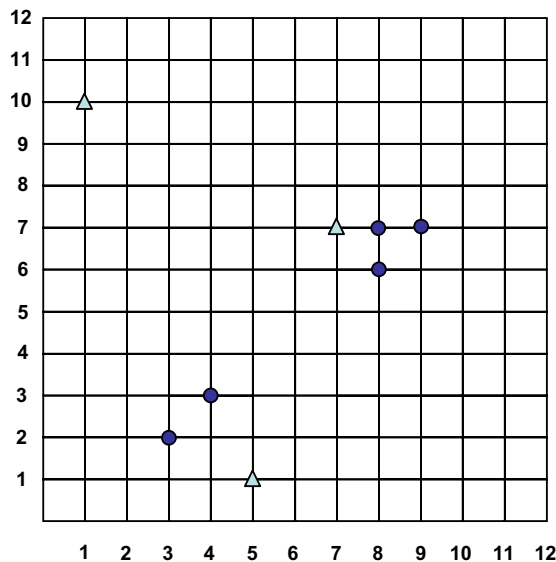Prof. Dr. Matthias Schubert
Felix Borutta

Managing Massive Multiplayer Online Games
SS 2019

Exercise Sheet 6: Knowledge Discovery and Data Mining
The assignments are due June 12, 2019

Assignment 6-1   *Instance-based learning: kNN-Classification*

Consider the following data set consisting of 8 points. The triangles are one class and the circles are another class.

Determine the classes of the given data points by using the k-nearest neighbors algorithm. If not stated differently, use the Manhattan distance (l_1 norm) as distance measure:

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

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

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

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

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

(e) Determine the class of point (6,1) for $k = 3$ using the majority class among the k-nearest neighbors. This time, employ a weighted version for the class decision, i.e., weight the class occurrences with the inverse Manhattan distance.

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

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

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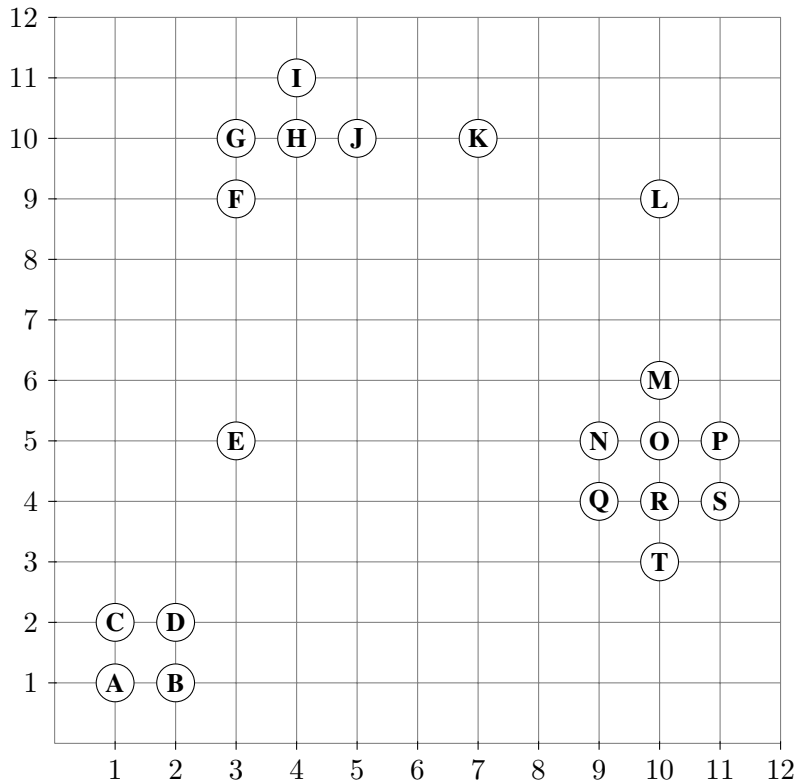
Cluster this dataset using DBSCAN. Use the Manhattan distance as distance function and the parameters $\epsilon = 1.1$ and $minPts = 3$.

Assignment 6-3   *Supervised Learning: Naive Bayes Classifier*

Given the following table of observations describing under which weather conditions person *A* was playing computer games.

Outlook	Temperature	Humidity	Play Computer Games
Sunny	Moderate	High	No
Sunny	High	Low	Yes
Rainy	Moderate	High	Yes
Rainy	High	High	No
Sunny	Moderate	Low	No
Sunny	Low	Low	No
Rainy	Low	Low	Yes

Outlook, *Temperature* and *Humidity* denote the observed features and *Play Computer Games* is the target variable.

Given the observation *o = (Outlook = Sunny, Temperature = High, Humidity = High)*, decide whether *A* is going to play computer games or not. Calculate the class probabilities by using the naive Bayes classifier.