



Exercise Sheet 8

Intelligent Systems

Classification

This exercise sheet will be discussed on February 3, 2020

Exercise 1 - DBSCAN and Outlier Detection

- A. Calculate the *Local Outlier Factor (LOF)* of the points A_1 and N in Figure 1.
- B. Draw the distribution of ascending *kdist*s of every point with $k = 1, 2, 3$.
- C. How can you estimate the parameter ϵ by given a percentage of noise?
- D. Find parameters $\epsilon > 0$, $\text{min_pts} \in \mathbb{N}$ s.t.
 - $A_i, i = 1, 2, 3$ is clustered as a cluster
 - $B_j, j = 1, 2$ is clustered as a cluster
 - N is marked as noise.
- E. Find parameters $\epsilon > 0$, $\text{min_pts} \in \mathbb{N}$, and points C_k s.t.
 - $A_i, i = 1, 2, 3, B_j, j = 1, 2$ is clustered as a cluster.
 - N is marked as noise.

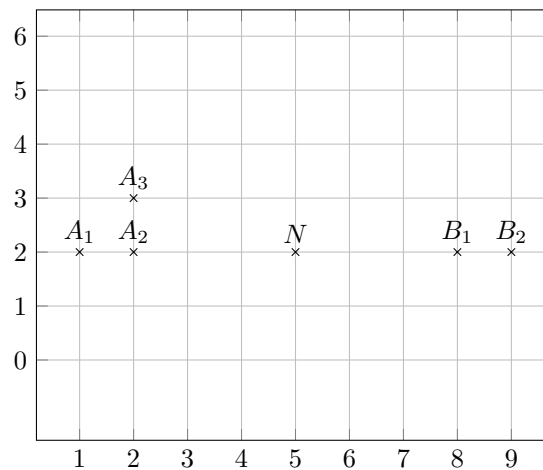


Abbildung 1: Points.

Exercise 2 - Classification algorithms

- A. Observe the data set in Table 1. First, create a 1-R Classifier that is able to predict whether a person is going to visit the party this evening by using the information of his/her amount of money, whether he/she writes an exam tomorrow, or if his/her heartthrob will come to the party.
- B. Extend your 1-R Classifier to a Decision Tree. Which features should be placed on higher levels of the tree?
- C. Apply the Naïve Bayes Classifier on the same data set. Calculate also the probabilities $P(Yes|E1)$ and $P(No|E6)$.

Sample	Money	Exam	Heartthrob	Party
E1	10	Yes	Yes	Yes
E2	13	No	Yes	Yes
E3	11	Yes	No	No
E4	12	No	No	Yes
E5	7	Yes	Yes	Yes
E6	5	Yes	No	No
E7	6	No	Yes	Yes
E8	8	No	No	No

Tabelle 1: Party Datensatz

Exercise 3 - Classification with Python

Download the Jupyter Notebook *7_Classification.ipynb* and the dataset *usStatesData2.csv* from *OpenOlat*. The task is to predict the vote label only by given the features in the dataset (e.g. *percentage of weapon owners*). Which classification metrics would you apply to test the performance of your classifier?