C | A | U

Kiel University
Christian-Albrechts-Universität zu Kiel

Exercise Sheet 9
Intelligent Systems
January 06, 2020

Anomalies and Classification

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 minimum *kdists*

$$min_y(kdist(x, y)),$$

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 $\epsilon > 0$, 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 $\epsilon > 0$, 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.

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 $P(No|E6)$.
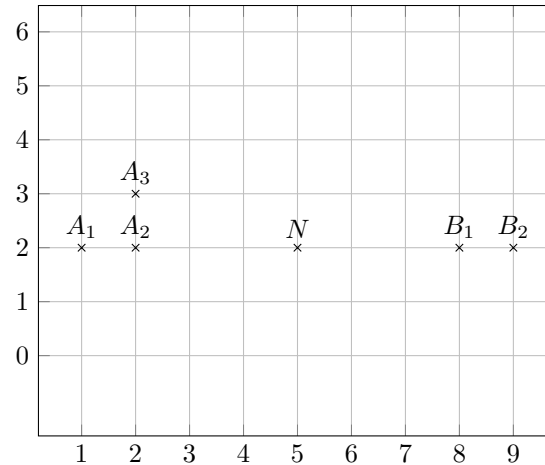
Figure 1: Points.

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

Table 1: Party Datensatz

Exercise 3 - Classification with Python

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

Signature Task - Classification Challenge

In classification tasks, the labels (here: the author of the signature) in the training data is given to the model for training. The task is to build a classifier that is able to predict the author of a signature. To present your results make a short presentation with 4 slides describing the following steps:

A. Preprocessing (Outlier detection/Normalisation/Segmentation)

B. Extraction of your lower dimensional features

C. Dimensionality reduction via PCA

D. Classification (Bayes Classifier, Decision Trees, SVM, NN, etc.)