1 Overview

The goal of this session is to get familiar with core data mining tasks and the data mining process *without* diving into the details of how algorithms work. We provide JUPYTER notebooks, but feel free to use any tool you know/like: e.g. WEKA, MATLAB, EXCEL, GNUPLOT, your own programs and scripts – as long as they help gain insights into the data. The exercises can be done in groups of 4-5 people.

2 Exploratory Data Analysis

Download `eda.csv` and `eda.ipynb` from Toledo. Use the guidelines from the lecture to get familiar with these data and identify their interesting properties. Look at attribute distributions, relations between attributes, missing data, and so on.

Each group gets to present one most interesting pattern that they discover. Make sure that the pattern is easy to grasp, e.g. by using suitable visualizations.

3 Data Mining Process: Pre-processing the data

Download `bikes_train.csv` and `bikes.ipynb` from Toledo. This dataset is related to the two-year historical log corresponding to years 2011 and 2012 from Capital Bikeshare system, Washington D.C., USA. In addition, the corresponding weather and seasonal information have been added to the dataset. [1]

1. In your group, discuss possible use cases of these data.[2]
2. In the dataset provided, bike rental usage is aggregated on hourly basis. Make use of the functionality `pandas` provides for convenient access to time series data and compute the daily system usage.
3. Bike-sharing rental process is highly correlated to the environmental and seasonal settings. Try to find interesting patterns in the data by applying visualization techniques similar to that of the Exercise 1.

Each group will present their use cases and interesting findings.

[1] Raw data is is publicly available on `http://capitalbikeshare.com/system-data`
Weather information are extracted from `http://www.freemeteo.com`

[2] See, for example, the following articles making use of similar data:
`http://toddwschneider.com/posts/a-tale-of-twenty-two-million-citi-bikes-analyzing-the-nyc-bike-share-syste`