Data Mining Exercise 1: 27.1.2022

Zoom 8.15

Remember to enroll yourself on the examination of the current course no later than 8 days before an examination date!

1. Installation instructions of Matlab are given in course web page. Install Matlab and select Home open. Open dm1.m file and play with it. It contains some repeatedly needed example tasks. In the first task in the file we generate successive integers from 1 to 100. How could you generate values from 500 to 20? How about values from 0 to 100 with the interval of 0.1? If you have not used Matlab before, visit https://se.mathworks.com/help/matlab/getting-started-with-matlab.html. Commands can be written to the command window or they can be run from script file by selecting the command (or set of commands) right clicking the mouse on the selected area and selecting "Evaluate selection ".
2. In file dm1.m there are two signals $y_1[n]=\sin(2\pi nf/F)$ and $y_2[n]=\cos(2\pi nf/F)$ where n gets the values from 1 to 100. Find the intersection point of the signals visually. You may need such functions as abs(y) and min(y). Why you cannot find the exact point of intersection?
3. Difference (diff(y)) is a counterpart of derivative in discrete world, while cumulative sum (cumsum(y)) resembles integral respectively. Evaluate $\frac{d}{dn}y_1[n]$ considering n as a continuous real variable. On the basis of your result, generate its values in discrete points [1 100] and compare it to the difference of signal $y_1[n]$.
4. Data analysis is usually done using observation matrix. In an observation matrix rows are the observations and columns are the variables that are considered to describe the observations as well as possible. Let us consider that we have a matrix **A**. In Matlab you can access to the value v that is in the position of (i,j) using notation v=**A**(i,j). In this notation i means row index and j means column index. An entire row **r** can be selected using notation **r**=**A**(i,:). Load in file Matrix.txt and plot its 5th row. You can also listen to the content of 5th row using function soundsc. (load ('Matrix.txt')). Calculate also the means (mean) of rows, columns and whole matrix.
5. Let us consider that our task would be to find the lowest sounding signal from the file Matrix.txt. Outline one method that could be used in the task.
6. Import file inco13par.txt (Home import data) (select Numeric matrix, default is column vectors). How many cases and variables the resulting matrix contains? What are the numbers of missing values for each variable? You may need such functions as find() and isnan().