Data mining and network analysis ITI8665

Home assignment 2

Distance function, clustering.

Starting date 15.03.2019 deadline 28.03.2019 10:00

General requirements:

- No plagiarism in any form. Please cite all the sources you used.
- Prepare your solution in such a way, that after extracting files from the archive into a single folder it may be executed on any computer with MATLAB. Data file for evaluating your solutions will follow the same structure as during the practice: single array where rows correspond to the elements and columns correspond to different dimensions.
- Prepare a short write-up with the analysis of achieved results. Maximum 2 pages 12pt.
- NB! NO E-MAIL SUBMISSIONS!!!
- Submit write-up as **PDF** file by means of ained.ttu.ee environment ained.ttu.ee
- Upload your code and all necessary files to https://gitlab.cs.ttu.ee grant developer rights to sven.nomm@ttu.ee (sven.nomm@taltech.ee)
- During the practice on 28.03 you will have to demonstrate your solution and will be asked few questions. Note it is mandatory to attend practice on 28.03 and demonstrate your solutions.
- If you are unsure about using some third party function contact your teacher.

Exercise 1. k - nn

Program in MATLAB your own implementation of the k - nearest neighbors method. Evaluate performance of your method with respect to different metric functions, different number of nearest neighbors and classes with different shapes. You may use standard MATLAB functions for mean, standard deviation, covariance matrix but not for distance functions!

Exercise 2. Decision trees

Program in MATLAB your own implementation of binary decision trees.

Exercise 3. Regression

Program in MATLAB your own implementation of regression model building. It is strongly advised to program mean squares method yourself.

Exercise 4. Gradient descend

Program in MATLAB gradient descent algorithm.

Good luck!