Data mining and network analysis ITI8665

Home assignment 3

Distance function, clustering.
Starting date 18.04.2019 deadline 02.05.2019 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 02.05 you will have to demonstrate your solution and will be asked few questions. Note it is mandatory to attend practice on 02.05 and demonstrate your solutions.
- If you are unsure about using some third party function contact your teacher.

Exercise 1.
Choice 1: Program in MATLAB or Python your own your own implementation of neural network with 2 hidden layers and 2 neurons on each layer. Perform the training process. NB for this choice it is not allowed to used MATLAB or third party neuron functions.

Choice 2: Program in MATLAB or Python your own SOM implementation. Provide an example to demonstrate its performance.

Exercise 2.
Implement Ada Boost algorithm in MATLAB or Python. Implementation of Ada Boost is expected to be your own. But individual weak classifiers may be standard MATLAB or Python functions. In this assignment special attention will be payed to the intermediate results. Please observe special conditions:

a. Design your own datasets to train and validate "Boosted classifier"

b. During the implementation is it necessary to demonstrate evolution of the weights and errors. It is preferable to demonstrate performance of each classifier.

c. Implementation should allow to extract parameters of the weak classifiers.

d. Strictly no symbolic computations for this exercise!

e. Graphical illustrations are mandatory!

Good luck!