Introduction to Data Science, WMCS16002, semester 1a 2020

1 Assignment 3 - Part B

Due October 8, 2020 12:00h (midday)
Submit by creating a pull request on GitHub

1.1 Genetic Algorithm (15 points)

Download the Matlab script **myGeneticAlgorithm.m**, which contains an incomplete implementation of a **genetic algorithm** that can be used for feature selection. Fill in the following:

a) the function called **getnewpopulation()**, (6 points)
b) the function called **getOffSpring()**, and (6 points)
c) the calculation of a fitness score in the function (3 points) **getScore()**

1.2 Application: Feature Selection (5 points)

Download the Matlab script **myFeatureSelectionwithGA.m** and the file **wine.data** that contains the WINE data set[1]. The Matlab script contains an implementation that uses the script **myGeneticAlgorithm.m** (mentioned above) for feature selection and it is applied to the WINE data set.

Run multiple experiments with the script **myFeatureSelectionwithGA.m** and compare the outcome of the feature selection approach based on the genetic algorithm with no feature selection (i.e. using the full feature set). Use graphical representation (of your choice) to illustrate the accuracy rates of both approaches for multiple experiments.

[1] http://archive.ics.uci.edu/ml/datasets/Wine/