Assignment 4

This assignment will be graded by 100 point scale and could compose 10% of your total grade. You must send the Python code to na.melkonyan@gmail.com.

Deadline: 31 May 2019, 11:59 pm (GMT+4)

Programming

Don't use native Python functionality in case the task states to *implement*.

Data

Please, download the data to your working folder. On the webpage you can find its description. https://archive.ics.uci.edu/ml/datasets/Iris

Report

Use methods you know so far to extract knowledge from Iris data (KDD process). The report is intended to the reader who is not familiar neither with the task nor with the problem. Reports should contain expressive language, details, plots and arguments. Every your conclusion or proposition should be backed up with an experiment.

Extra Points

Extra points will be granted for extra creativity, e.g. Sturges' formula has been used to find the number of bins in the histogram.

Rules

Take attention on the feedbacks of previous assignments.

Theoretical Question

Construct an example where K-means clustering algorithm ends-up with an empty cluster. E.g., algorithm was initialized let's say with 3 clusters, but after convergence 1 cluster had no points in it.