INFO284, Assignment 3 breakdown

27 March 2019   15:18

Goal:

Obtain practical knowledge of clustering algorithms

Process:

Scikit-learn classifier, use algorithms on dataset
1. k-Means clustering algorithm
   i. Introduction to Machine Learning with Python, p.172-193
2. Gaussian Mixture Models Clustering algorithm
   i. http://scikit-learn.org/stable/modules/mixture.html

Dataset, choose one
1. Engelsberger_full.xlsx
2. Engelsberger_short.xlsx

Compare results

Method
Fine-tune the parameters of the algorithms in Scikit-learn until we get clusters we are happy with.
Compare results.
Visualize the obtained clusters in a 2D scatter plot.
We may need to reduce the dimensionality of the clustered data points for the purposes of visualisation (not for the puposes of clustering)

Submit:

Code
1. Must work with dataset (Engelsberger_full.xlsx OR Engelsberger_short.xlsx)
   i. Do not submit the actual dataset
   ii. Must run on any operating system

Text
1. How to run the code
2. 1000-5000 character report
   i. describe how the clustering algorithms performed on the task.
   ii. how many clusters did we get?
   iii. state opinions at to why the results turned out the way they did.

Visualization
1. Visualisation of clustered data in a 2d scatter plot.