Exploring clinical heterogeneous data using unsupervised machine learning

16 February 2023

Assignment IV – Part 1

Reading

- Z. Huang. *Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values*. Data Mining and Knowledge Discovery, 1998. See Paper.

- To read about the Silhouette score sklearn and Wikipedia.
- Good practices for writing source code in Python: https://peps.python.org/pep-0008/

Experimentation with k-prototypes

1. To implement the **k-prototypes** algorithm (see Section 5 of Huang's manuscript).
2. To download the dataset soybean disease and credit approval. Please see details in Huang's paper Section 6.1.1.
3. Share the Code and dataset in GitHub by the **1st March 2024**.

- NOTE_1. In case of any question regarding the assignment III, email me to adan.josegarcia@univ-lille.fr. In case you need a personal session for explanation of a topic please email me in advance to the deadline.

Dr. JOSÉ-GARCÍA
Building ESPRIT S3.11
https://adanjoga.github.io