Mateo W. Racca

UNIVERSITAT ROVIRA i VIRGILI
Escola Tècnica
Superior d'Enginyeria

NEURAL AND EVOLUTIONARY COMPUTING

(MESIIA): Assignment #3: Unsupervised learning with PCA, t-SNE, k-means, AHC and SOM

Part 1

The chosen dataset is commonly known as the HCV dataset (Hepatitis C Virus) from the UCI Machine Learning Repository. Below is the detailed description of the dataset and a link to the source webpage:

HCV Dataset Description:

Domain: Medical / Hepatology.

Objective: The dataset is typically used for research and analysis in the medical domain, often for the purpose of understanding the factors associated with Hepatitis C Virus infection and its stages.

Features: The dataset includes various medical measurements such as liver enzymes, bilirubin, albumin, and other blood chemistry measurements. It also contains demographic information like age and gender.

Target: The dataset categorizes patients into different categories based on the stage of Hepatitis or other clinical diagnoses related to HCV. The possible values are `'0=Blood Donor'`, `'0s=suspect Blood Donor'`, `'1=Hepatitis'`, `'2=Fibrosis'`, `'3=Cirrhosis'`

Data Points: It initially has 615 instances, each representing different patients or clinical cases.

The dataset can be retrieved from: https://archive.ics.uci.edu/dataset/571/hcv+data