Statistical and Machine Learning methods for classification and clustering
Proposed project for MISCADA

Advisor: Georgios P. Karagiannis

Durham University, 2022

Description

The project includes the study, assessment, and modification of recent developments statistical and machine learning methods for clustering and classification. Clustering analysis involves statistical methods for finding groups in data set automatically, while classification involves statistical methods for assigning future outcomes to groups given previously classified observations automatically. They are suitable for the statistical analysis and modeling of heterogeneous populations, however the particular focus of the project is on data sets which present modern challenges such as with high-dimensional data, network data, functional data, and others. Clustering and classification have applications in engineering, environment, health, biology, social sciences, and others.

Potential project directions

Potential directions for the project/dissertation can be: Bayesian methods for classification, or clustering ; variable selection methods for clustering ; supervised and semi-supervised methods clustering ; clustering and classification methods for high-dimensional data, network data, non-Gaussian data sets, or functional data.

Project specific intended learning outcomes

By the end of this project, students will be able to design and implement suitable Statistical and Machine learning methods for modeling, inference, and predictions in heterogeneous populations. You will be exposed to the use of suitable software required to the practical implementation of the methodology.

Requirements

Knowledge of Bayesian Statistics, and Regression. ; Knowledge of R or Python

Contact details

For further information, feel free to contact
  Georgios Karagiannis (Office CM126b)
  Email: georgios.karagiannis@durham.ac.uk
  Web: https://www.maths.dur.ac.uk/users/georgios.karagiannis/