

# **Advantages and disadvantages of non and semi-parametric methods for data classification and categorization.**

## **Preamble.**

One major role of statistical agencies the world over is the analysis and classification of data. Generally, classification methods are based upon arbitrarily chosen (occasionally criticised) criteria. Non-parametric and semi-parametric classification methods afford agencies and researchers the ability to assess the extent, magnitude and importance of latent groupings within their data products thus facilitating examination of the relevance of existing categorization criteria.

## **AIM**

The aim of this class is to introduce the basic concepts of nonparametric and semi-parametric classification methods with particular emphasis on different types of income data (truncated, grouped data..). Real datasets will be used for the empirical analyses and codes in R will be provided.

## **OUTLINE OF THE TRAINING COURSE**

1. Data issues
2. Kernel density estimation
  - 2.1. From histogram to kernel density
  - 2.2. Density estimation with sampling weights
  - 2.3. Bandwidth selection
  - 2.4. Adaptive kernel estimation
  - 2.5. Multivariate and conditional density
  - 2.6. Kernel density estimation in R
3. Finite mixture models
  - 3.1. Mixtures of distributions
  - 3.2. Estimation procedure
  - 3.3. Number of components and number of groups
  - 3.4. Group profiles: covariates entering the mixing weights
  - 3.5. Mixture of regressions: covariates entering the component distributions
  - 3.6. Estimating mixtures in R