

Methods for dimension reduction and clustering in Fintech surveys

§ Name of instructor

Dimitris Karlis , Athens University of Economics and Business, Greece.

§ Short description

Basic Ideas/Principles on Dimension Reduction. Principal Components Analysis: the machinery, how many components, variants of the method, applications with real data.

§ § Schedule

Correspondence analysis. The machinery, type of stances, the idea of inertia.

Visualization. Applications with real data. Multiple Correspondence Analysis, how and why. Applications

Clustering: hierarchical and distance based methods. Distances and their properties.

Agglomerative clustering. K-means and variants. Assessing the number of clusters, cluster validation, quality of clustering

Model- based clustering: the finite mixture model. Advantages and disadvantages.

The EM algorithm. Selecting the number of clusters. Certain examples and applications

§ Introductory background

McNicholas, P.D. (2016), Mixture Model-Based Classification, Boca Raton FL:

Chapman & Hall/CRC Press.

Jolliffe, I. T. (2002). Principal component analysis for special types of data (pp. 338-372). Springer New York.

Giordani, P., Ferraro, M. B., & Martella, F. (2020). An introduction to clustering with R. Springer Singapore.

§ Facilities Required

- Software: R, open source, with the R package `clustrd` installed

- Course Material. All course materials, including the data and R scripts for the examples, will be made available for course participants.