UNSUPERVISED MACHINE LEARNING

Postgraduation in Applied Artificial Intelligence & Machine Learning

COURSE SYLLABUS

Jorge Caiado
ISEG School of Economics and Management/University of Lisbon,
CEMAPRE/REM - Research in Economics and Mathematics and GlobalSolver
Email: jcaiado@iseg.ulisboa.pt
Web: http://jcaiado100.wix.com/jorgecaiado

Lisboa, 2024

OBJECTIVES

This course introduces key topics in unsupervised learning with Python, such as clustering, principal component analysis, multidimensional scaling, anomaly and outlier detection, specifically for data-intensive applications. It focuses on utilizing unsupervised learning tasks and algorithms in various applications, including customer segmentation, dimensionality reduction, anomaly and fraud detection, and pattern discovery in large datasets.

PROGRAMME

- Hierarchical Cluster Analysis (Class #1)
- K-Means Clustering (Class #2)
- Principal Component Analysis (PCA) (Class #3)
- Multidimensional Scaling (MDS) (Class #4)
- Anomaly and Outlier Detection: DBSCAN (Class #5)
- Anomaly and Outlier Detection: Isolation Forests, Local Outlier Factor, Elliptic Envelope, and One-Class SVM (Class #6)

SELECTED BIBLIOGRAPHY

- Chopra, R. and England, A. (2019). *Data Science with Python*, Packt Publishing Ltd.
- E.A. Maharaj, P. D'Urso and Caiado, J. (2019). *Time Series Clustering and Classification,* CRC Press, Taylor & Francis Group, Boca Raton (United States).
- M. Kubat (2021). *An Introduction to Machine Learning*, Springer.
- Pate, A. A. (2019). *Hands-On Unsupervised Learning Using Python: How to Build Applied Machine Learning Solutions from Unlabeled Data,* O'Reilly.
- Tripathy, B.K., Sundareswaran, A. and Ghela, S. (2022). *Unsupervised Learning Approaches for Dimensionality Reduction and Data Visualization*, CRC Press, Taylor & Francis Group.

ASSESSMENT

The **Unsupervised Machine Learning (UML)** course assessment is a single element: **Regular Assessment (100%)**. This involves a group project (3-5 students) on "UML with Python." Choose a research topic, gather literature sources and cross-sectional data. Apply cluster analysis, PCA , MDS and/or anomaly detection to analyze finance, economic, social, or business phenomena. Justify each step, explain choices, and present relevant tables and graphs. **Project Submission:** 12-July-2023. **Presentation and discussion:** 15-July-2023 (19:00-20:30, on-line). Students have exactly 15 minutes for the presentation, and we will reserve approximately 5 minutes for the discussion.

Some useful datasets:

https://medium.datadriveninvestor.com/top-8-sources-for-machine-learning-and-analytics-datasets-5d2d94ada8ab

https://vincentarelbundock.github.io/Rdatasets/datasets.html (1300 datasets)