UNIVERSIDAD
DE GRANADA

SSC IMPROVE: Data Analysis and Pattern Recognition

The goal of this course is to introduce to the students basic concepts of Data Analysis and Pattern Recognition. Data Analysis techniques provide tools to visualize and understand patterns in the data. These techniques are useful for the massive data collected from the volcano monitoring system, and help to identify sensor problems/misfunctioning and sensor diversity. Pattern Recognition and Machine Learning techniques allow to infer a model from known data and clustering unknow data according to the model. Data analysis is completely complementary to Machine Learning fitting model.

This course will be imparted by instructors of the ETSIIT of the Granada University:

- Professor Dr. José Camacho
- Professor Dra. Carmen Benítez
- Dr. Manuel M. Titos
- Dr. Guillermo Cortés
- Dr. Michael Sorochan Armstrong

Sofware: Matlab and Phyton

Dates: 24-25-26 of January 2024
Place: Escuela Técnica Superior de Ingeniería Informática y de Telecomunicación (ETSIIT) Granada

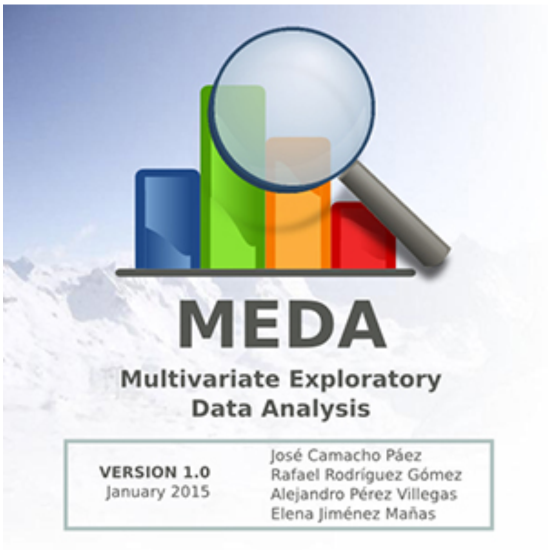
Wednesday 24/01/2024
Topic: Data Analysis

9:00-11:00

Introduction to MEDA and the MEDA Toolbox (MEDA, Matlab)

The Multivariate Exploratory Data Analysis (MEDA) Toolbox in Matlab is a set of multivariate analysis tools for the exploration of data sets. There are several alternative tools in the market for that purpose, both commercial and free. The PLS_Toolbox from Eigenvector Inc. is a very nice example. The MEDA Toolbox is not intended to replace or being a competitor of any of these toolkits. Rather, the MEDA Toolbox is a complementary tool that includes several of our recent contributions to the field. Thus, traditional exploratory plots based on Principal Component Analysis (PCA) or Partial Least Squares (PLS), such as score, loading and residual plots, are combined with new methods: MEDA, oMEDA, SVI plots, ADICOV, EKF & CKF cross-validation, CSP, GPCA, ...

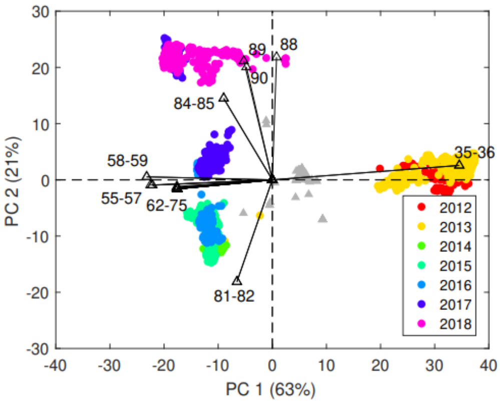
The MEDA Toolbox can be used to analyze normal size data sets (several hundreds of observations times several hundreds of variables) There is also an extension of the toolbox for large data sets, with millions of items, under folder Big Data.



11:30-13:30

Principal Component Analysis (MEDA, Matlab)

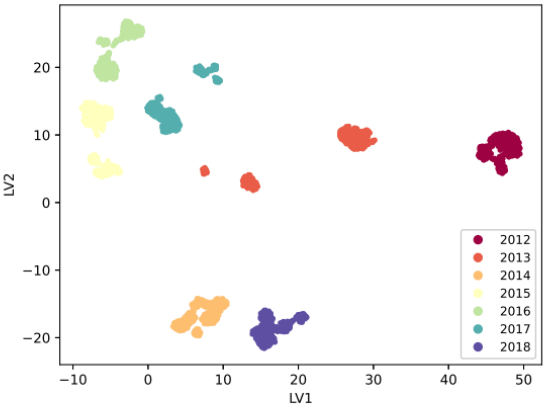
PCA is one of the most widely used data-analysis methods across disciplines. It has been widely employed for the derivation of useful multivariate visualizations. The PCA factorization is useful for the visualization of multivariate data, since it allows us to explore the distribution of the observations (rows) and of the features (columns) of the data in separate plots of much lower dimension, and hence easier to visualize, while they retain most of the information in the data.



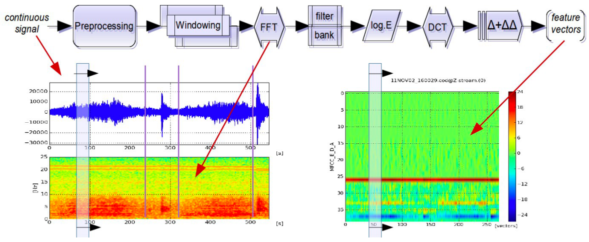
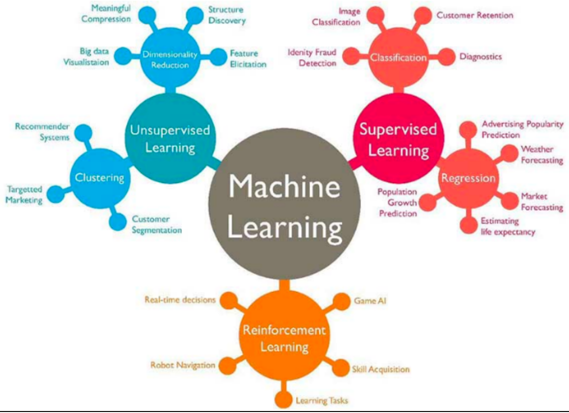
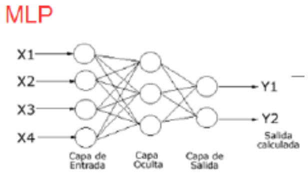
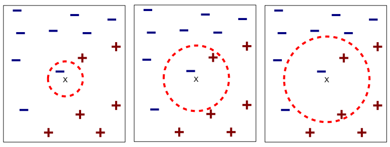
15:00-17:00

UMAC (Python)

UMAP is a popular non-linear dimensionality reduction method to efficiently visualize high-dimensional data with no obvious linear relationships. UMAP can be seen as a non-linear extension of PCA where we can explore complex projections to find interesting patterns in our data. However, the scores do not correspond to linear distances within the data itself, and consequently the contributions of each variable are difficult to interpret.



Thursday 25/01/2024
Topic: Pattern Recognition

9:00-11:00. **Introduction to Feature Extraction** In order to classify and label events, it is not suitable to work directly with the samples captured by the sensor. The registered signal is contaminated with noise from sources external to the volcano, and hence it is necessary to use algorithms that allow for the extraction of relevant information from the signal less affected by the noise. The goal is to characterize the signal through a set of values which is most representative and useful for a subsequent classification task.	 continuous signal Preprocessing Windowing FFT filter bank log.E DCT Δ+ΔΔ feature vectors
11:30-13.30 **Introduction to Machine Learning** Machine learning is an important component of the growing field of data science. Through the use of statistical methods, algorithms are trained to make classifications or predictions, and to uncover key insights in data mining projects. These insights subsequently drive decision making within applications and businesses, ideally impacting key growth metrics. In this course we will introduce: • Learning algorithms • Performance measure • Unsupervised/ Supervised Learning	 Machine Learning Unsupervised Learning Dimensionality Reduction Meaningful Compression Structure Discovery Big data Visualistaion Feature Elicitation Clustering Recommender Systems Targetted Marketing Customer Segmentation Supervised Learning Classification Image Classification Customer Retention Idenity Fraud Detection Diagnostics Regression Advertising Popularity Prediction Weather Forecasting Population Growth Prediction Market Forecasting Estimating life expectancy Reinforcement Learning Real-time decisions Game AI Robot Navigation Skill Acquisition Learning Tasks
15:00 17:00 Classic algorithms • Supervised learning algorithms: RF, ANN • Unsupervised learning examples: KNN	 MLP X1 X2 X3 X4 Y1 Y2 Capa de Entrada Capa Oculta Capa de Salida Salida calculada  (a) 1-nearest neighbor (b) 2-nearest neighbor (c) 3-nearest neighbor

Friday 26/01/2024
Topic: Exercises with volcano data

9:00_11:00	Data Analysis Exercises
11:30-13: 30	Machine Learning Exercises