MACHINE LEARNING TRAINING FOR DATA SCIENCE

THIS TRAINING COURSE ENABLES THE PARTICIPANTS TO REINFORCE THEIR THEORETICAL AND PRACTICAL KNOWLEDGE IN ORDER TO IMPLEMENT MACHINE LEARNING TECHNIQUES FOR THE AUTOMATIC ANALYSIS OF DATA. THE MAIN STATISTICAL METHODS FOR DATA ANALYSIS ARE PRESENTED, BOTH FOR DATA EXPLORATION (NON-SUPERVISED LEARNING) AND FOR PREDICTION (SUPERVISED LEARNING). EACH METHOD IS FIRST PRESENTED AND COMMENTED ON A THEORETICAL LEVEL, AND THEN ILLUSTRATED ON NUMERICAL EXPERIMENTS RUN WITH PUBLIC DATASETS USING R AND/OR PYTHON/SCIKIT-LEARN SOFTWARE.

OBJECTIVE

To know the main algorithms of automatic data analysis, and to know how to use them with R and/or python/scikit-learn.

EDUCATIONAL OBJECTIVES

The participants should be able to:

- Recognize the type of problem that they are facing: supervised or non-supervised learning, sequential learning, reinforcement learning ;
- Choose the right algorithm to use ;
- Use an R or python implementation of this algorithm.

TARGET PARTICIPANTS

This training session is for students, engineers, and computer scientists who wish to reinforce or extend their theoretical background and practical knowledge on automatic data analysis by statistical learning algorithms.

INFOS

DATE : upon request

DURATION :
Séquence 1 : 2 jours | 14 heures
4 Days | 28 hours

PRICE : 2 520€
Lunches and educational material included

PRACTICAL INFORMATION AND REGISTRATION:
+33 (0)5 61 55 92 53 | fcq@insa-toulouse.fr

RESPONSABLE DU STAGE :

- **Béatrice LAURENT-BONNEAU**
  Professeur des Universités
  Professeur au département Génie
  Mathématiques et Modélisation -
  INSAT - Institut de Mathématiques
  de Toulouse

A certificate of attendance will be delivered at the end of the training

PROGRAM

Day 1 ML Introduction and Unsupervised learning

- General presentation of statistical machine learning
- Principal component analysis
- Agglomerative Hierarchical Clusteringk-means, k-medoids and variants, DBSCAN
- Tutorials Execute successively the first tutorial and then the episodes one of each notebook: Ozone, Mars, HAR, MNIST. The segmentation of an image of Mars by clustering algorithms is specific to the first day.

N.B. It would be too long to execute all notebooks. So you can choose them accordingly to your level and/or your field of interest. There is a complexity progression from Ozone to MNIST.

Day 2 Supervised learning

- K nearest neighbors
- Gaussian linear model, logistic regression, model selection
- Lasso et variants
- Support Vector Machines
- Tutorials Episodes two of: Ozone, HAR, MNIST

Day 3 Supervised learning

- Decision Trees
- Bagging, Random Forests, Boosting
- Neural networks, deep learning
- Tutorials Episodes three of: Ozone, HAR, MNIST

Day 4 Other ML algorithms

- Sequential learning, multi-armed bandit problems
- Super-learning and expert aggregation
- Reinforcement learning (introduction)
- Tutorials Inventory Control

Joint training venture :

INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES TOULOUSE

Σ CERFACS

BIG DATA