Machine Learning

Duration 3 day(s) (ML)

Understand and develop machine learning algorithms

Description

The Big Data revolution has generated a dramatic increase in data volumes, storage space and computing power to process them. Taking full advantage of this favorable environment, Machine Learning has significantly increased the efficiency of tools we use every day, such as search engines and referral systems. These fields of application, which are constantly expanding, today range from medicine to industry and the financial sector. This training offers an extended presentation of Machine Learning, as it is used today in the professional world. It was designed to be accessible to an audience from a variety of backgrounds, the prerequisites being few. The Machine Learning methodologies and its main algorithms will be presented, in their concepts as in their typical cases of use. Each time, implementations based on diversified domains will be proposed. They will take the form of labs implemented in Python language and using the most common libraries. Constructed in a didactic way, these labs will allow a tangible approach of the reality of Machine Learning: the predictive power of the models, like their limitations, will be studied notably through the quantitative analysis of the obtained results. The modern and very attractive subject of Deep Learning, based on neural networks, will be the subject of a first introduction.

Goals

- Understanding Data Science
- Understanding Machine Learning
- To know how to model a Machine Learning problem
- Types of Machine Learning
- Problems of Machine Learning
- Most used algorithms through application examples
- Introduction to Deep Learning and Neural Networks

Public

Developers, future Data Scientists, Architects, Functionals, Project Managers

Prerequisites

- Basic knowledge in algebra (matrices) and statistics
- Knowledge in programming, ideally in Python

Structure

50% Theory, 50% Practice

Program

Introduction: Big Data & Data Science

Reminders on algebra (vectors and matrices) & statistics

Machine Learning: Automatic Learning

- Definition and history
- Examples of machine learning applications
- Modeling a problem in Machine Learning
- Types of learning (Supervised / Unsupervised)

Steps in learning

- Choice of model
- Learning: calculation of model parameters
- Over-learning
- Validation, cross validation, test
- Model comparison criteria

Getting started with Python

- The Jupyter Notebook
- Introduction to Python Programming
- Basic structures and operations in Python
- **LAB 1**: Getting started with Python
- Data recovery - Exploration and preprocessing of data (use of Pandas and Numpy libraries)
- Visualization of data (use of the Matplotlib library)
- **LAB 2**: Data exploration and pre-processing

Learning algorithms

Regression

- Use case: Prediction of selling prices of houses
- Metrics of the regression
- Linear regression
- Principle and functioning
- Cost / loss function
- Optimization function (Gradient Descent algorithm)
- **LAB 3**: Linear regression
- Regressions: multiple, Ridge, Lasso
- **LAB 4**: Multiple regression, Ridge & Lasso regression

Classification

- Use case: Detection of spam emails
- Metrics of the classification
- Logistic Regression
- SVM (Vector Support Machine)
- **LAB 5**: Logistic Regression & SVM
- Decision trees
- Random forests
- **LAB 6**: Decision Trees and Random Forests
- K-NN (the nearest K neighbors)
- **LAB 7**: K-NN

Segmentation & clustering

- Use cases: Segmentation of articles
- Distances K-means (K-Means)
- **LAB 8**: K-averages
- Spectral clustering
- Hierarchical clustering
- **LAB 9**: Spectral & Hierarchical clustering

Recommendation Systems

- Use case: Recommendation system for an e-commerce site
- Content-based filtering
- **LAB 10**: Content-based filtering
- Collaborative filtering
- **LAB 11**: Collaborative filtering

Dimensional reduction

- Use case: example of the Iris data
- PCA (Principle Component Analysis)

Deep learning: Deep learning

- Use case: Image classification
- Multilayer perceptron neuron networks
- Convulsion neural networks
- Recurrent neural networks
- Autoencoders networks

Challenges and perspectives