

Contents

- 1 Introduction
 - Organization, tooling

Contents

- 1 Introduction
 - Organization, tooling
 - Motivations and Use Cases

Contents

- 1 Introduction
 - Organization, tooling
 - Motivations and Use Cases
 - Big Data mouvement
- 2 Statistical Learning, Machine Learning frameworks

Contents

- 1 Introduction
 - Organization, tooling
 - Motivations and Use Cases
 - Big Data mouvement
- 2 Statistical Learning, Machine Learning frameworks

Contents

- 1 Introduction
 - Organization, tooling
 - Motivations and Use Cases
 - Big Data mouvement
- 2 Statistical Learning, Machine Learning frameworks

Contents

- 1 Introduction
 - Organization, tooling
 - Motivations and Use Cases
 - Big Data mouvement
- 2 Statistical Learning, Machine Learning frameworks

Contents

- 1 Introduction
 - Organization, tooling
 - Motivations and Use Cases
 - Big Data mouvement
- 2 Statistical Learning, Machine Learning frameworks

Contents

- 1 Introduction
 - Organization, tooling
 - Motivations and Use Cases
 - Big Data mouvement
- 2 Statistical Learning, Machine Learning frameworks

- 4 courses of 3 hours each, with practical examples
- 3 Practical works of 4 hours each, with R/RStudio and python/anaconda programs / libraries
- 3 short quizzes
- Slides from several MOOCs, advanced materials

- How to understand datas, correlation but not causation
- How to handle variety of datas (see 3V description later)
- see materials for supervised (*) and unsupervised (*) use-cases

(*) Definition to come later

- 3V definition :
 - Volumen, Velocity, Variety
 - + Veracity, ++
- Position of data mining vs ML vs Statistical Learning vs AI

Contents

1 Introduction

Contents

1 Introduction

Contents

1 Introduction

2 Statistical Learning, Machine Learning frameworks

Contents

1 Introduction

2 Statistical Learning, Machine Learning frameworks

- Supervised learning

Contents

1 Introduction

2 Statistical Learning, Machine Learning frameworks

- Supervised learning
- Unsupervised learning

Contents

1 Introduction

2 Statistical Learning, Machine Learning frameworks

- Supervised learning
- Unsupervised learning
- The Goals of Unsupervised Learning

Contents

1 Introduction

2 Statistical Learning, Machine Learning frameworks

- Supervised learning
- Unsupervised learning
- The Goals of Unsupervised Learning
- Reinforcement learning

Contents

1 Introduction

2 Statistical Learning, Machine Learning frameworks

- Supervised learning
- Unsupervised learning
- The Goals of Unsupervised Learning
- Reinforcement learning
- Summary

Starting point :

- Outcome measurement Y (also called dependent variable, response, target) ;
- Vector of p predictor measurements X (also called inputs, regressors, covariates, features, independent variables). X is a matrix of dimension (N,p) , where n is the number of measurements ;
- In the **regression problem**, Y is quantitative (e.g price, blood pressure) ;
- In the **classification problem**, Y takes values in a finite, unordered set (survived/died, digit 0-9, cancer class of tissue sample) ;
- We have training data $(x_1, y_1), \dots, (x_N, y_N)$. These are observations (examples, instances) of these measurements.

- In unsupervised learning, we observe only the features X_1, X_2, \dots, X_p .
- We are not interested in prediction, because we do not have an associated response variable Y .

- The goal of Unsupervised Learning is to discover interesting things about the measurements : is there an informative way to visualize the data ? Can we discover subgroups among the variables or among the observations ?
- We discuss two methods :
 - **principal components analysis (PCA)**, a tool used for data visualization or data pre-processing before supervised techniques are applied, and
 - **clustering**, a broad class of methods for discovering unknown subgroups in data.

- In reinforcement learning, measurements are done one by one.
(e.g. temporal series)
- After each prediction a reward is given
- Use Case : Chess or Go player
- toward common sense...

3 columns table
image Yann Le Cun ML cake