

# 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 movement
- 2 Statistical Learning, Machine Learning frameworks

# Contents

## 1 Introduction

- Organization, tooling
- Motivations and Use Cases
- Big Data movement

## 2 Statistical Learning, Machine Learning frameworks

# Contents

## 1 Introduction

- Organization, tooling
- Motivations and Use Cases
- Big Data movement

## 2 Statistical Learning, Machine Learning frameworks

# Contents

## 1 Introduction

- Organization, tooling
- Motivations and Use Cases
- Big Data movement

## 2 Statistical Learning, Machine Learning frameworks

# Contents

## 1 Introduction

- Organization, tooling
- Motivations and Use Cases
- Big Data movement

## 2 Statistical Learning, Machine Learning frameworks

# Contents

## 1 Introduction

- Organization, tooling
- Motivations and Use Cases
- Big Data movement

## 2 Statistical Learning, Machine Learning frameworks



# Outline

## 1 Introduction

- Organization, tooling
- Motivations and Use Cases
- Big Data movement

## 2 Statistical Learning, Machine Learning frameworks

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

- 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

# Outline

## 1 Introduction

- Organization, tooling
- Motivations and Use Cases
- Big Data movement

## 2 Statistical Learning, Machine Learning frameworks

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

- 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

# Outline

## 1 Introduction

- Organization, tooling
- Motivations and Use Cases
- **Big Data movement**

## 2 Statistical Learning, Machine Learning frameworks

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

- 3V definition :
  - Volume, 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

# Outline

## 1 Introduction

- Organization, tooling
- Motivations and Use Cases
- Big Data movement

## 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.



# Outline

## 1 Introduction

- Organization, tooling
- Motivations and Use Cases
- Big Data movement

## 2 Statistical Learning, Machine Learning frameworks

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

- 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  $Y$ .

# Outline

## 1 Introduction

- Organization, tooling
- Motivations and Use Cases
- Big Data movement

## 2 Statistical Learning, Machine Learning frameworks

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

- 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.

# Outline

## 1 Introduction

- Organization, tooling
- Motivations and Use Cases
- Big Data movement

## 2 Statistical Learning, Machine Learning frameworks

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

- 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...

# Outline

## 1 Introduction

- Organization, tooling
- Motivations and Use Cases
- Big Data movement

## 2 Statistical Learning, Machine Learning frameworks

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

# The 3 types of Machine Learning

From the most difficult to the less

Unsupervised Learn.	Supervised Learn.	Reinforcement Learn.
Observations Understand Clustering	+ targets (labels) Predict Classification or Regression	++ rewards Decide Action Policy Strategy



# ML Cake

From the Yann Le Cun's lesson at College de France

## Reinforcement Learning (cherry)

- The machine predicts a scalar continue reward given once in a while.
- **A few bits for some samples**

## Supervised Learning (icing)

- The machine predicts a category or a few numbers for each input
- **10→10,000 bits per sample**

## Unsupervised Learning (cake)

- The machine predicts any part of its input for any observed part.
- Predicts future frames in videos
- **Millions of bits per sample**

