# Introduction

Nathanaël Carraz Rakotonirina

Mathématiques Informatique et Statistique Appliquées (MISA)
Université d'Antananarivo

# Notation/Terminology

## Supervised learning

We want a mapping $f$ from inputs $\boldsymbol{x} \in \mathcal{X}$ ($\mathcal{X} = \mathbb{R}^D$) to outputs $y \in \mathcal{Y}$.

- $\boldsymbol{x}$ : features, predictors, covariates
- $y$ : label, target, response
- $\mathcal{D} = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^N$ : training set ($(\boldsymbol{x}_i, y_i)$ is the ith observation or example)
- $D$ : number of features
- $N$ : number of training examples
- $\boldsymbol{X} \in \mathcal{R}^{N \times D}$ : Design matrix

**Classification** : $\mathcal{Y} = \{1, 2, ..., C\}$
where $C$ is the number of classes

**Regression** : $\mathcal{Y} = \mathbb{R}$

## Unsupervised learning

Only inputs $\mathcal{D} = \{\boldsymbol{x}_i\}_{i=1}^N$ are observed without any corresponding outputs. We want to find structure, "meaning" in data.

## Reinforcement learning

An agent learns to interact with its environment. A policy defines which action to take in response to each possible environment state. A reward signal is received in response to the actions the agent takes.

## Other types of ML

- Semi-supervised learning
- Self-supervised learning

- **"Pure" Reinforcement Learning** (cherry)
  - The machine predicts a scalar 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
  - Predicting human-supplied data
  - **10→10,000 bits per sample**
- **Unsupervised/Predictive Learning** (cake)
  - The machine predicts any part of its input for any observed part.
  - Predicts future frames in videos
  - **Millions of bits per sample**

Figure: Machine learning types as layers of a cake by Yann Le Cun at NIPS'2016

## Models

- A (statistical) **model** $\mathfrak{F}$ is a set of distributions or densities or functions.
- A **parametric model** $\mathfrak{F}$ is a set of distributions or densities or functions that can be parametrized by a finite number of parameters.

$$\mathfrak{F} = \left\{ f(x; \boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta \right\}$$

  sometimes just $f(x; \boldsymbol{\theta})$ where $\boldsymbol{\theta}$ is the parameter
- A **non-parametric model** is a set $\mathfrak{F}$ that cannot be parameterized by a finite number of parameters.

# Machine Learning Systems

A Machine learning system is generally composed of:

- Dataset
- Model (with parameters and hyperparameters)
- Loss function
- Optimization (learning) algorithm
- Regularization strategy

## Other related fields

- Predictive Analytics
- Statistics
- Data Mining
- Data Science
- Business Intelligence
- Artificial Intelligence

See ML glossary
https://developers.google.com/machine-learning/glossary/

# Prerequisites

## Must-have

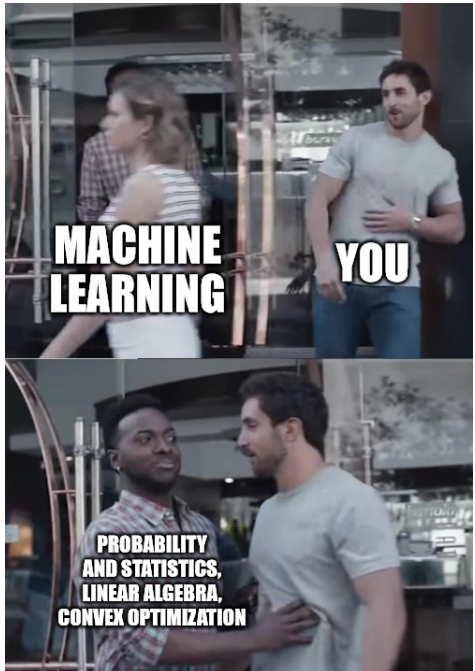
- Programming
- Probability and statistics
- Linear algebra
- Convex optimization
- Multivariate calculus

## Recommended

- Analysis of algorithms
- Information Theory
- Numerical computation

## Practical skills

- Python
- Bash
- SQL
- Git
- Scrum
- DevOps
- Writing and presentation skills


MACHINE LEARNING
YOU
PROBABILITY AND STATISTICS, LINEAR ALGEBRA, CONVEX OPTIMIZATION