

# Machine Learning: Theory, Fairness, Privacy

Christos Dimitrakakis

September 19, 2024

# Outline

## The problems of Machine Learning (1 week)

- Introduction

- Course Contents

- Objective functions

- Pitfalls

## The problems of Machine Learning (1 week)

Introduction

Course Contents

Objective functions

Pitfalls

# Machine Learning And Data Mining

## The nuts and bolts

- ▶ Models
- ▶ Algorithms
- ▶ Theory
- ▶ Practice

## Problems

- ▶ Data collection
- ▶ Classification
- ▶ Regression
- ▶ Clustering
- ▶ Compression
- ▶ Reinforcement learning

# Machine learning

## Data Collection

- ▶ Downloading a clean dataset from a repository
- ▶ Performing a survey
- ▶ Scraping data from the web
- ▶ Deploying sensors, performing experiments, and obtaining measurements.

## Modelling (what we focus on this course)

- ▶ Simple: the bias of a coin
- ▶ Complex: a language model.
- ▶ The model depends on the data and the problem

## Algorithms and Decision Making

- ▶ We want to use models to make decisions.
- ▶ Decisions are made every step of the way.
- ▶ Decisions are automated algorithmically.

# The main problems in machine learning and statistics

## Prediction

- ▶ Will it rain tomorrow?
- ▶ How much will bitcoin be worth next year?

## Inference

- ▶ Does my poker opponent have two aces?
- ▶ What is the mass of the moon?
- ▶ What is the law of gravitation?

## Decision Making

- ▶ Should I go hiking tomorrow?
- ▶ Should I buy some bitcoins?
- ▶ Should I fold, call, or raise in my poker game?
- ▶ How can I get a spaceship to orbit the moon?

# The need to learn from data

## Problem definition

- ▶ What problem do we need to solve?
- ▶ How can we formalise it?
- ▶ What properties of the problem can we learn from data?

## Data collection

- ▶ Why do we need data?
- ▶ What data do we need?
- ▶ How much data do we want?
- ▶ How will we collect the data?

## Modelling and decision making

- ▶ How will we compute something useful?

# Learning from data

## Unsupervised learning

- ▶ Given data  $x_1, \dots, x_T$ .
- ▶ Learn about the data-generating process.

## Supervised learning

- ▶ Given data  $(x_1, y_1), \dots, (x_T, y_T)$
- ▶ Learn about the relationship between  $x_t$  and  $y_t$ .
- ▶ Example: Classification, Regression

## Online learning

- ▶ Sequence prediction: At each step  $t$ , predict  $x_{t+1}$  from  $x_1, \dots, x_t$ .
- ▶ Conditional prediction: At each step  $t$ , predict  $y_{t+1}$  from  $x_1, y_1, \dots, x_t, y_t, x_{t+1}$

## Reinforcement learning

Learn to act in an **unknown** world through interaction and rewards



# Course Contents

## Models

- ▶ k-Nearest Neighbours.
- ▶ Linear models and perceptrons.
- ▶ Multi-layer perceptrons (aka deep neural networks).
- ▶ Trees.
- ▶ Mixture models (bagging, boosting).
- ▶ Support vector machines.

## Algorithms

- ▶ (Stochastic) Gradient Descent.
- ▶ Linear programming.
- ▶ Quadratic programming
- ▶ Bayesian inference.
- ▶ Expectation maximisation.
- ▶ Monte Carlo Methods.

# Supervised learning

The general goal is learning a function  $f : X \rightarrow Y$ .

## Classification

- ▶ Input data  $x_t \in \mathbb{R}$ ,  $y_t \in [m] = \{1, 2, \dots, m\}$
- ▶ Learn a mapping  $f$  so that  $f(x_t) = y_t$  for unseen data

## Regression

- ▶ Input data  $x_t, y_t$
- ▶ Learn a mapping  $f$  so that  $f(x_t) = \mathbb{E}[y_t]$  for unseen data

# Unsupervised learning

The general goal is learning the data distribution.

## Compression

- ▶ Learn two mappings  $c, d$
- ▶  $c(x)$  compresses an image  $x$  to a small representation  $z$ .
- ▶  $d(z)$  decompresses to an approximate image  $\hat{x}$ .

## Density estimation

- ▶ Input data  $x_1, \dots, x_T$  from distribution with density  $p$
- ▶ Problem: Estimate  $p$ .

## Clustering

- ▶ Input data  $x_1, \dots, x_T$
- ▶ Assign each data  $x_t$  to cluster label  $c_t$ .

# Supervised learning objectives

- ▶ Data  $(x_t, y_t)$ ,  $x_t \in X$ ,  $y_t \in Y$ ,  $t \in [T]$ .
- ▶ i.i.d assumption:  $(x_t, y_t) \sim P$  for all  $t$ .
- ▶ Supervised decision rule  $\pi(a_t|x_t)$

## Classification

- ▶ Predict the labels correctly, i.e.  $a_t = y_t$ .
- ▶ Have an appropriate confidence level

## Regression

- ▶ Predict the mean correctly
- ▶ Have an appropriate variance around the mean

# Unsupervised learning objectives

- ▶ Reconstruct the data well
- ▶ Be able to generate data

# Reinforcement learning objectives

- ▶ Maximise total reward

# Pitfalls

## Reproducibility

- ▶ Modelling assumptions
- ▶ Distribution shift
- ▶ Interactions and feedback

## Fairness

- ▶ Implicit biases in training data
- ▶ Fair decision rules and meritocracy

## Privacy

- ▶ Accidental data disclosure
- ▶ Re-identification risk