Machine Learning & Deep Learning
Artificial Intelligence

Jacopo Parretti

I Semester 2025-2026

Indice

Parte I

Introduction to Machine Learning

1   What is Machine Learning?

Machine Learning is a branch of Artificial Intelligence that enables software to use data to find solutions to specific tasks without being explicitly programmed to do so.

1.1   Machine Learning vs Statistical Modelling

In traditional statistical modelling, we follow a structured process:

1. Collect data

2. Verify and clean the data (correct or discard if not clean)

3. Use the clean data to test hypotheses, make predictions and forecasts

In contrast, Machine Learning takes a different approach:

- It is the **data** that determines which analytic techniques to use

- The computer uses data to **train** algorithms to find patterns and make predictions

- Algorithms are no longer **static** but become **dynamic**

1.2   Conventional Programming vs Machine Learning

The fundamental difference between conventional programming and machine learning can be understood through their inputs and outputs:

1.2.1   Conventional Programming

- **Input**: Program + Data

- **Output**: Result

- You write the program, give it data, and get results

1.2.2   Machine Learning

- **Input**: Data + Result

- **Output**: Program

- You give the computer data and desired results, and it learns to generate a program/algorithm

1.3   The Machine Learning Recipe

The ML recipe consists of 4 fundamental steps:

1. Collect data

2. Define a family of possible models

3. Define an objective (error) function to quantify how well a model fits the data

4. Find the model that minimizes the error function (training/learning a model)

1.3.1   The Key Ingredients

Every machine learning problem requires five essential ingredients:

- **Task**: The problem we want to solve

- **Data**: The information we use to learn

- **Model hypothesis**: The family of functions we consider

- **Objective function**: How we measure success

- **Learning algorithm**: How we find the best model

2   Machine Learning Tasks

A task represents the type of prediction being made to solve a problem on some data. We can identify a task with the set of functions that can potentially solve the problem. In general, it consists of functions assigning each **input** $x \in \mathcal{X}$ an **output** $y \in \mathcal{Y}$.

2.1   Classification Task

Definizione 2.1 (Classification)**.** Find a function $f \in \mathcal{Y}^{\mathcal{X}}$ assigning each input $x \in \mathcal{X}$ a **discrete** label, $f(x) \in \mathcal{Y} = \{c_1, \ldots, c_k\}$.

In classification, the output space is a finite set of discrete categories or classes. The goal is to learn a decision boundary that separates different classes.

2.2   Regression Task

Definizione 2.2 (Regression)**.** Find a function $f \in \mathcal{Y}^{\mathcal{X}}$ assigning each input $x \in \mathcal{X}$ a **continuous** label, $f(x) \in \mathcal{Y} = \mathbb{R}$.

In regression, the output is a continuous value, allowing for predictions of quantities such as prices, temperatures, or probabilities.

2.3   Density Estimation Task

Definizione 2.3 (Density Estimation)**.** Find a probability distribution $f \in \Delta(\mathcal{X})$ that best fits the data $x \in \mathcal{X}$.

There is no reference to an output space (as in classification and regression tasks). Instead, we make a reasoning on the **input** x itself, trying to model the underlying probability distribution that generated the data.

2.4   Clustering Task

Definizione 2.4 (Clustering)**.** Find a function $f \in \mathbb{N}^{\mathcal{X}}$ that assigns each input $x \in \mathcal{X}$ a **cluster index** $f(x) \in \mathbb{N}$.

All points mapped to the same index form a cluster. The goal is to group similar data points together without prior knowledge of the groups.

2.5   Dimensionality Reduction Task

Definizione 2.5 (Dimensionality Reduction)**.** Find a function $f \in \mathcal{Y}^{\mathcal{X}}$ that **maps** each (**high-dimensional**) input $x \in \mathcal{X}$ to a **lower-dimensional** embedding $f(x) \in \mathcal{Y}$, where $\dim(\mathcal{Y}) < \dim(\mathcal{X})$.

This task aims to reduce the number of features while preserving the essential information in the data.

3 Data in Machine Learning

3.1 Data Representation

We represent inputs of our algorithm as $x \in \mathcal{X}$ and outputs as $y \in \mathcal{Y}$. The dataset is the set of both, where each input is associated with an output.

Key considerations:

- The quality of the data is crucial for the performance of the algorithm

- It is equally important to have a large amount of data to learn effective models

- Poor quality or insufficient data can lead to poor model performance

3.2 Features

Definizione 3.1 (Features)**.** Features are the measurable properties or characteristics of the phenomenon being observed.

Examples:

- To identify a flower: color, petal length, petal width, sepal length, sepal width

- To identify a fruit: color, shape, size, texture, weight

- To classify images: pixel values, edges, textures, shapes

We can say that features are how our algorithm "sees" the data and are in general represented with (fixed-size) vectors.

4 Data Distributions and Learning Types

In machine learning, the information about the problem we want to solve is represented in the form of a data distribution, usually denoted as p_{data}.

4.1 Supervised Learning

4.1.1 Classification and Regression

The data distribution is over pairs of inputs and outputs:

$$p_{\text{data}} \in \Delta(\mathcal{X} \times \mathcal{Y})$$

Here:

- $\mathcal{X}$ is the input space

- $\mathcal{Y}$ is the output space (labels or targets)

- The goal is to learn a mapping from inputs to outputs using labeled data

4.2   Unsupervised Learning

4.2.1   Density Estimation, Clustering, and Dimensionality Reduction

The data distribution is only over the input space:

$$p_{\text{data}} \in \Delta(\mathcal{X})$$

Here:

- We do not have explicit labels or targets

- The goal is to discover structure, patterns, or representations within the data itself

4.3   Summary of Learning Types

Task Type	Data Distribution	Learning Type
Classification, Regression	$p_{\text{data}} \in \Delta(\mathcal{X} \times \mathcal{Y})$	Supervised Learning
Density Estimation, Clustering, Dimensionality Reduction	$p_{\text{data}} \in \Delta(\mathcal{X})$	Unsupervised Learning

Tabella 1: Comparison of learning types based on data distribution

Nota. In **supervised learning**, each data point consists of an input and a corresponding output (label). In **unsupervised learning**, each data point consists only of the input, and the algorithm tries to find patterns or structure without explicit labels.

This distinction is fundamental in machine learning, as it determines the type of algorithms and approaches used to solve different problems.