

Machine Learning with PyTorch and Scikit-Learn From Coursera

Cristo Daniel Alvarado

23 de octubre de 2025

Índice general

1. Module 1	3
1.1. Overview	3
1.2. Giving computers the ability to learn from data	3
1.3. Introduction	3
1.3.1. The Three Different Types of Machine Learning	4
1.3.2. Supervised Learning	4
1.3.3. Classification for Predicting Class Labels	5
1.3.4. Regression for Predicting Continuous Outcomes	6

Lista de Códigos

CAPÍTULO 1

MODULE 1

The goal of this chapter is to explore the foundational concepts of machine learning, focusing on how algorithms can transform data into knowledge. We delve into the practical applications of supervised and unsupervised learning, equipping you with the skills to implement these techniques using Python tools for effective data analysis and prediction.

Observación 1.0.1 (Learning Objectives)

Learning Objectives:

- Analyze data patterns to make future predictions.
- Design systems for supervised and unsupervised learning.
- Implement machine learning algorithms using Python tools.

1.1. OVERVIEW

The goal basically is to learn machine learning using Python libraries such as Scikit-Learn and PyTorch.

1.2. GIVING COMPUTERS THE ABILITY TO LEARN FROM DATA

So, basically in this scenario we will do the following:

- Implement machine learning algorithms using Python tools.
- Designing systems for supervised and unsupervised learning.
- Analyze data patterns to make future predictions.

1.3. INTRODUCTION

Definición 1.3.1 (Machine Learning)

Machine learning is the *application and science of algorithms that make sense of data*.

The goal of this section is to introduce some of the main concepts and different types of machine learning, together with a basic introduction to the relevant terminology. The goal is to establish the groundwork for successfully using machine learning techniques for practical problem solving.

1.3.1. THE THREE DIFFERENT TYPES OF MACHINE LEARNING

In this age, we have a large amount of structured and non-structured data.

Definición 1.3.2 (Structured and Non-structured Data)

Structured data refers to *information that is organized in a predefined format*, typically within a fixed schema, such as rows and columns in a database or spreadsheet

In contrast, **non-structured data** *lacks a predefined format or structure and exists in its native, raw state*. It is typically qualitative and encompasses a wide variety of formats such as text documents, emails, social media posts, images, videos, and audio files.

One of the main goals of machine learning is to extract knowledge from data in order to make predictions.

Observación 1.3.1 (Use of Machine Learning)

Machine Learning offers a more efficient alternative for capturing the knowledge in data to improve the performance of predictive models and make data-driven decisions.

Idea 1.3.1 (Note on the use of Machine Learning and some of its Applications)

Also, notable progress has been made in medical applications; for example, researchers demonstrated that deep learning models can detect skin cancer with near-human accuracy [link to article](#).

Another milestone was recently achieved by researchers at DeepMind, who used deep learning to predict 3D protein structures, outperforming physics-based approaches by a substantial margin [link to article](#).

There are three types of machine learning: **supervised learning**, **unsupervised learning**, and **reinforcement learning**. There are fundamental differences between these three types of machine learning and we will look at each of them in detail with some notes on its possible applications.

Three Types of Machine Learning	
Supervised learning	• Labeled data
	• Direct feedback
	• Predict outcome/future
Unsupervised learning	• No labels/targets
	• No feedback
	• Find hidden structure in data
Reinforcement learning	• Decision process
	• Reward system
	• Learn series of actions

Cuadro 1.1: Three Types of Machine Learning.

1.3.2. SUPERVISED LEARNING

The main goal of **supervised learning** is to *learn a model from labeled training data that allows us to make predicitions about unseen or future data*.

The term *supervised* refers to a set of training examples (data inputs), where the desired output signals (labels) are already known.

Definición 1.3.3 (Supervised Learning)

Supervised learning is the *process of modeling the relationship between data inputs and the labels*.

We can think of supervised learning as *label learning*.

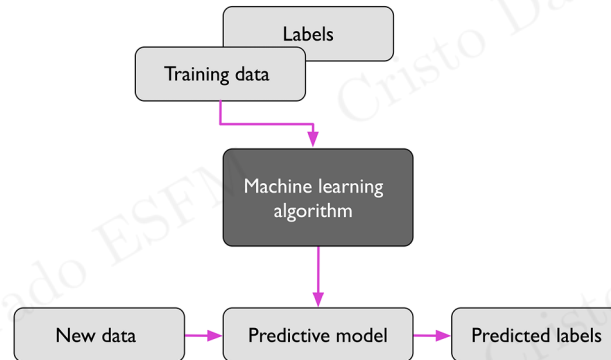


Figura 1.1: General Process of Supervised Learning

Ejemplo 1.3.1 (Example of Supervised Learning)

Lets consider the example of email spam filtering. We can train a model using a supervised machine learning algorithm on a corpus of labeled emails, which are correctly marked as spam or non-spam.

To predict whether a new email belongs to either of the two categories, a supervised learning task with discrete class labels, such as in the previous email spam filtering example.

This is called a **classification task**. Another subcategory of supervised learning is **regression**, where the outcome signal is a continous value.

1.3.3. CLASSIFICATION FOR PREDICTING CLASS LABELS

Definición 1.3.4 (Classification)

Classification is a *subcategory of supervised learning, where the goal is to predict the categorical class labels of new instances or data points based on past observations*.

Thos class labels are discrete, unordered values that can be understood as the group memberships of the data points. The previously mentioned example of email spam detection represents a typical example of a binary classification task, where the machine learning algorithm learns a set of rules to distinguish between two possible classes: spam and no-spam emails.

The next ilustration shows the concept of binary classification task given 30 training examples; 15 training are labeled as **class A** and the other 15 labeled as **class B**.

Our dataset is two-dimensional, meaning that each example has two values associated with it: x_1 and x_2 . We can use a supervised machine learning algorithm to to learn a rule (decision boundary represented as the dashed line) that can separate two classes and classify new data into each of those two categories given its x_1 and x_2 values.

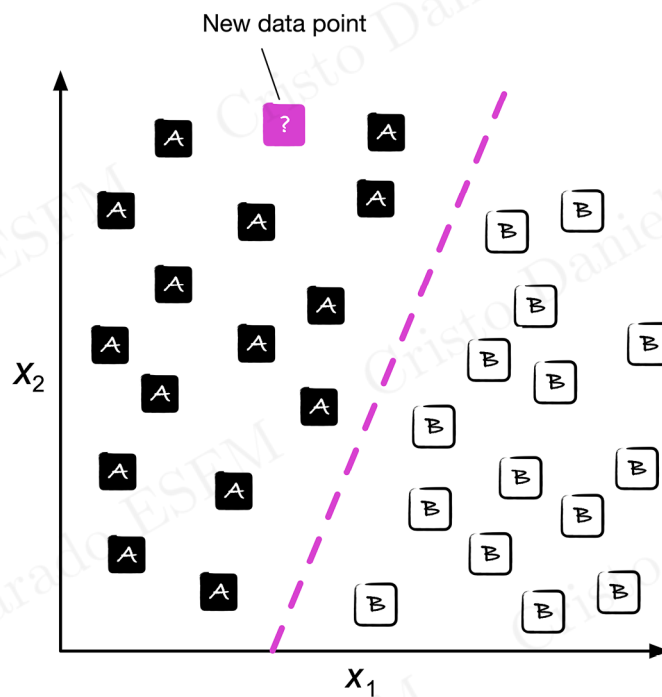


Figura 1.2: Classification in Machine Learning

Observación 1.3.2 (Nature of classification of class labels)

When a set of class labels does not have to be of a binary nature. The predictive model learned by a supervised learning algorithm can assign any class label that was presented in the training dataset to a new, unlabeled data point or instance.

Ejemplo 1.3.2

An example of a multiclass classification task is handwritten character recognition.

1.3.4. REGRESSION FOR PREDICTING CONTINUOUS OUTCOMES

Definición 1.3.5 (Regression Analysis)

In a **regression analysis**, we are given a number of predictor (**explanatory**) variables and a continuous response variable (**outcome**) and we try to find a relationship between those variables that allows us to predict an outcome.

Observación 1.3.3 (Feature and Target Variables)

In the field of machine learning, the predictor variables are commonly called *features*, and the response variables are referred to as *target variables*.

Ejemplo 1.3.3

Let's assume we want to predict the mat SAT scores of students. If there is a relationship between time spent studying for the test and the final scores, we could use it as training data to learn a model that uses the study time to predict the test scores of future students.

Given a feature variable x and a target variable y , we fit a straight line to the data points that minimizes the error in predicting y from x .

We can now use the intercept and slope learned from this data to predict the target variable of new data.

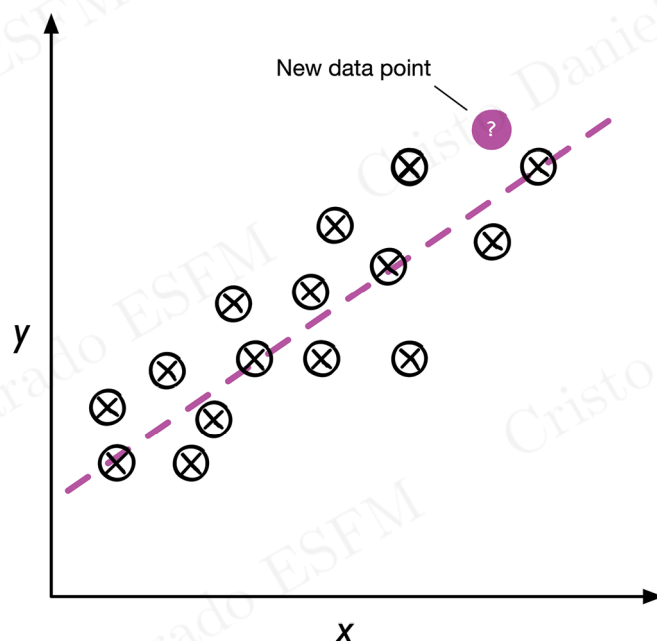


Figura 1.3: Linear Aproximation