

TYPES OF MACHINE LEARNING

1.1. Supervised Learning

Supervised learning is an algorithm that predicts the output of a new input based on given pairs of (input, output). These data pairs are also called (sample, label). Mathematically, supervised learning occurs when we have

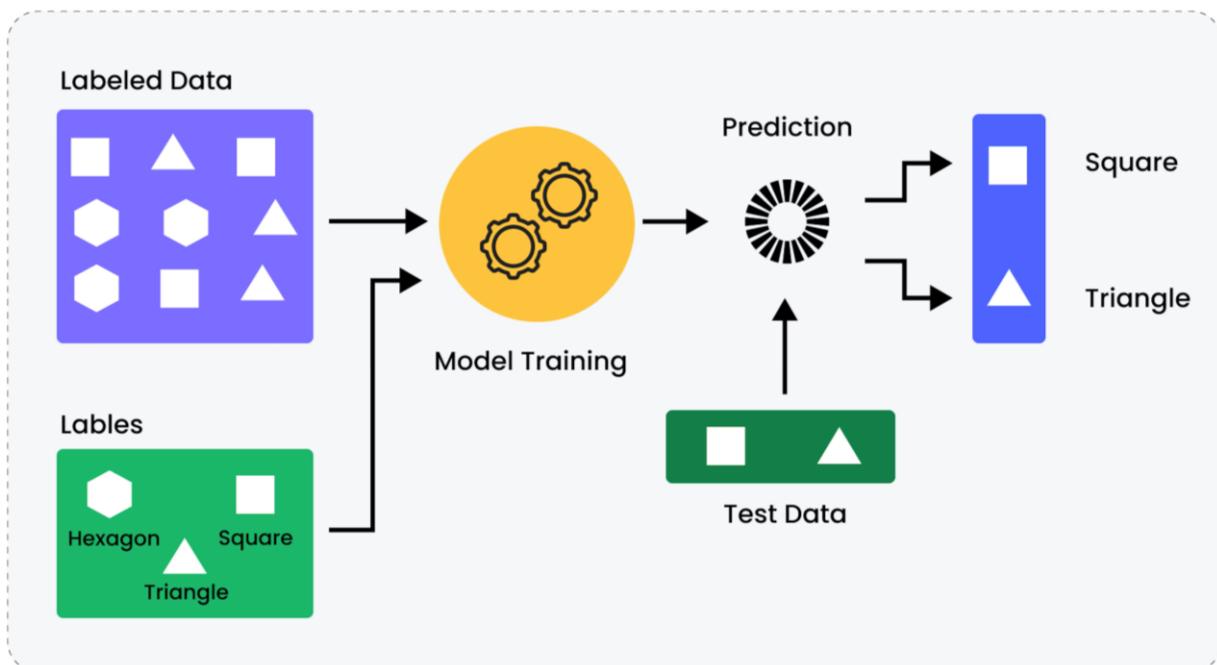
a set of input variables:

$$X = \{x_1, x_2, \dots, x_N\},$$

and a set of labels:

$$Y = \{y_1, y_2, \dots, y_N\},$$

The known data pairs $(x_i, y_i) \in X \times Y$ are called the **training data**. From this input, we need to create a function that maps each element from set X to an “correct” element in set Y . The goal is to normalize the function f well so that, given a new data point x , we can compute its corresponding label $y = f(x)$.



The supervised learning algorithm can be further divided into two main types:

1.1.1. Classification

A problem is called classification if the labels of the input data are divided into a finite number of groups. For example:

Problem: Predict whether an image contains a cat or a dog.

Input (features): Pixel image.

Output (label): “Cat” or “Dog”.

Description: The model is trained on a set of images, each labeled as a cat or a dog. During training, the model learns features in the images, such as shapes, textures, and colors, that help distinguish cats from dogs. After training, the model can look at a new image, recognize its features, and predict whether it is a cat or a dog. The prediction is a classification, because it chooses one label from the available labels.

1.1.2. Regression

In this case, the labels are not divided into groups but are specific real-valued numbers. For instance:

Problem: Predict a student’s exam score.

Input: Study time, IQ and EQ.

Output: Exam score (a continuous value, e.g., 0–100).

Description: The model is trained on data of old student and their corresponding exam scores. During training, it learns the relationship between the input's attributes and the score. Once trained, the model can take a new student’s **characteristics** and predict their exam score. The output can take any number from 0 to 10, means it’s a regression.