

Types of Machine Learning

Supervised Learning:

In supervised learning problems, we start with a data set containing data (images, text, files, mails, etc.) with their correct labels. There are two main kinds of tasks in this area:

- **Classification:** Classifying the input into various categories
- **Regression:** By using the input data to predict some set of number, like the price of a house

For example, when learning to classify handwritten digits, a supervised learning algorithm takes thousands of pictures of handwritten digits along with labels containing the correct number each image represents. The algorithm will then learn the relationship between the images and their associated numbers, and apply that learned relationship to classify completely new images that the machine hasn't seen before.

To illustrate how supervised learning works, let's examine the problem of predicting annual income based on the number of years of higher education someone has completed. Expressed more formally, we'd like to build a model that approximates the relationship f between the number of years of higher education X and corresponding annual income Y .

Supervised machine learning solves this problem by getting the computer to do the work for you. By identifying patterns in the data, the machine can form heuristics.

In supervised learning, the machine attempts to learn the relationship between income and education from scratch, by running labeled training data through a learning algorithm. This learned function can be used to **estimate the income of people whose income Y is unknown, if we have years of education X as inputs**. In other words, we can apply our model to the unlabeled test data to estimate Y . The goal of supervised learning is to predict Y as accurately as possible when given new examples where X is known and Y is unknown.

Unsupervised Learning:

How do you find the underlying structure of a dataset? How do you summarize it and group it most usefully? How do you effectively represent data in a compressed format? These are the goals of unsupervised learning, which is called "unsupervised" because you start with **unlabeled data** (there's no Y).

The two unsupervised learning tasks we will explore are clustering the data into groups by similarity and reducing dimensionality to compress the data while maintaining its structure and usefulness.