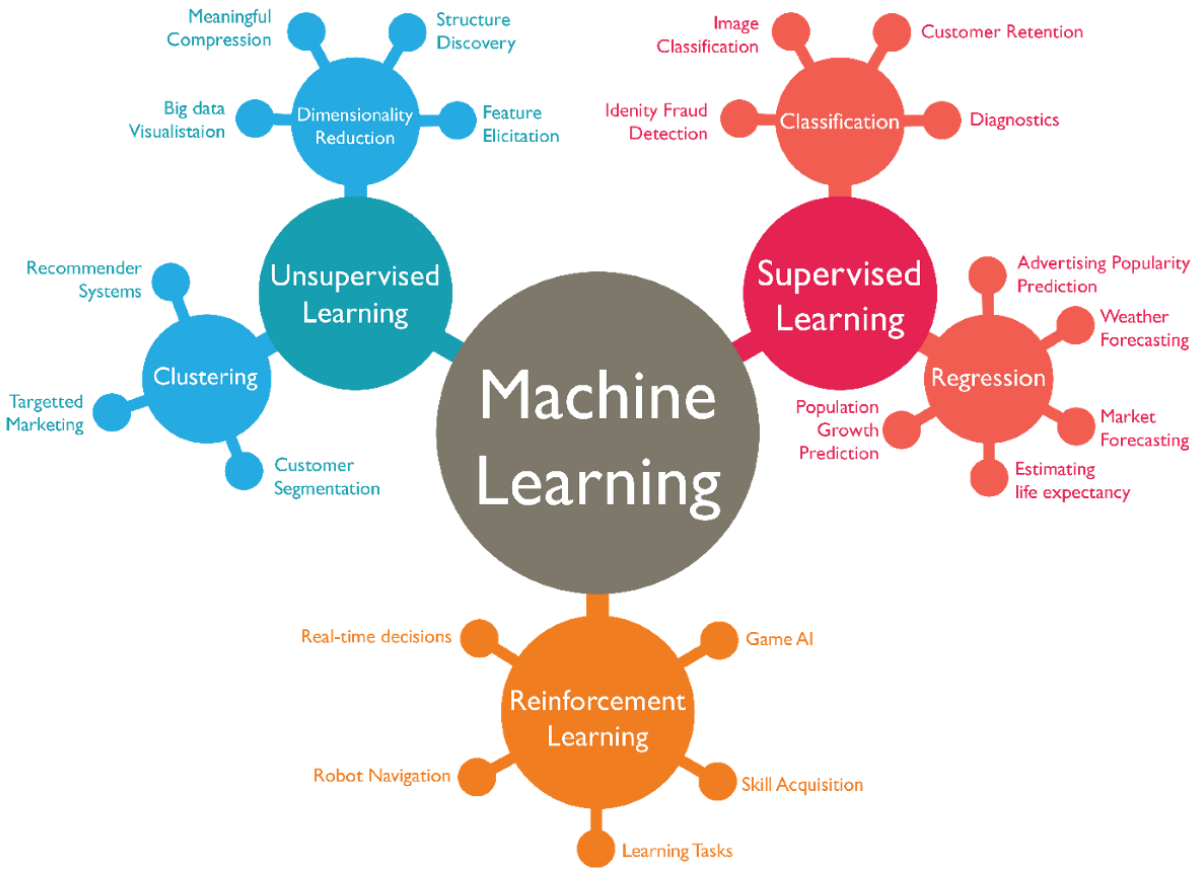
TYPES OF MACHINE LEARNING

ML (Machine Learning) is a interdisciplinary mix of statistics and computer science. ML is extremely powerful because it gives computers the ability to learn without being explicitly programmed to do so. Meaning, the computer can learn without step-by-step instructions. Essentially, machine learning learns patterns from existing data and applies it to new data.

A ML model is a statistical representation of a real-world process, like how we recognize cats or hourly changes in traffic. A process is modeled using data. There are three types of machine learning. The most common types are supervised and unsupervised learning. Their main difference lies in their training data. Reinforcement learning, is used for deciding sequential actions, like a robot deciding its path or its next move in a chess game. Reinforcement learning is not as common as the others and uses complex mathematics, like game theory.



1. Supervised Learning

ML “learns” patterns from existing data which is called “training data”. When a model is being built and learning from training data, this is called as "training a model". Let's look at training data for a supervised learning model. We'd like to train a model to predict whether a patient has heart disease. We have existing records from patients who've experienced chest pains and been tested for heart disease.

Target Variable

Age	Sex	Cholesterol	Cigarettes per day	Family history of heart disease	Chest pain type	Blood sugar	Heart disease
55	M	221	5	True	typical angina	118	True
50	F	196	0	False	non-anginal pain	98	False
53	F	215	0	True	asymptomatic	110	True
62	M	245	3	False	typical angina	126	True
48	M	190	0	True	non-anginal pain	99	False
70	M	201	0	True	typical angina	105	False

Labels

Our target variable is "heart disease", because this what we want to predict. The values "True" and "False" are labels for the target variable, meaning whether it's true or false that a patient has heart disease. Labels don't have to come in this form - they can be numbers or categories.

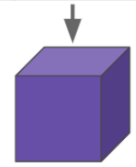
Features

Age	Sex	Cholesterol	Cigarettes per day	Family history of heart disease	Chest pain type	Blood sugar	Heart disease
55	M	221	5	True	typical angina	118	True
50	F	196	0	False	non-anginal pain	98	False
53	F	215	0	True	asymptomatic	110	True
62	M	245	3	False	typical angina	126	True
48	M	190	0	True	non-anginal pain	99	False
70	M	201	0	True	typical angina	105	False

Features are different pieces of information that might help predict the target. Age, cholesterol, and smoking habits are known factors of heart disease. The magic of machine learning is that we can analyze many features at once, even the ones we're unsure about, and find relationships between different features. We input labels and features as data to train the model.

Once training is done, we can give the model new input. In our case, a new patient. The features are inputted and, the model outputs its prediction.

Age	Sex	Cholesterol	Cigarettes per day	Family history of heart disease	Chest pain type	Blood sugar	Heart disease
65	F	208	2	False	typical angina	105	???



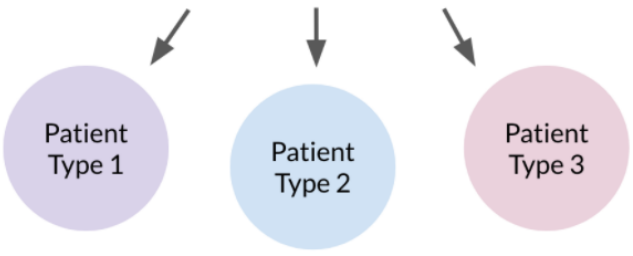
Heart disease
False

In supervised learning, the training data is "labeled", meaning the values of our target are known. For instance, we knew if previous patients had heart disease based on the labels "true" and "false". In unsupervised learning, we don't have labels, only features. What can we do with this? Usually tasks like anomaly detection and clustering, which divides data into groups based on similarity. Let's explore this with our dataset.

2. Unsupervised Learning

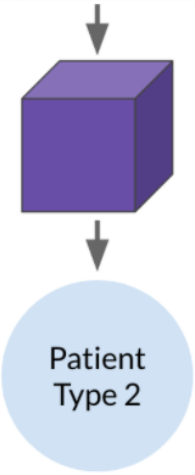
There are different treatments for heart disease. Different types of patients respond better or worse to certain treatments. We can use unsupervised learning to understand the different types of patients we have. Let's filter our dataset to only include patients with heart disease. We can pass it into a clustering model and get categories of patients based on feature similarity. For example, one category could be patients with high cholesterol and blood sugar level of a certain age range. Note, we didn't know these categories and, even, the number of categories before running this. With this output, we can group patients and research better treatments for each group.

Age	Sex	Cholesterol	Cigarettes per day	Family history of heart disease	Chest pain type	Blood sugar	Heart disease
55	M	221	5	True	typical angina	118	True
53	F	199	0	True	non-anginal pain	98	True
53	F	215	0	True	asymptomatic	110	True
62	M	245	3	False	typical angina	126	True
...	...	...	...	...	...	...	...



Now, with a new patient, we can input the features into the model and get which patient type they best fit into.

Age	Sex	Cholesterol	Cigarettes per day	Family history of heart disease	Chest pain type	Blood sugar	Heart disease
65	F	208	2	False	typical angina	105	True



In reality, data doesn't always come with labels. Either it's too much manual work to label or we don't even know what the labels are. Think of the effort it would take to label millions of road images for self-driving cars. This is when unsupervised learning shines. In this sense, the model is unsupervised and finds it's own patterns.

3. Reinforcement Learning

The last listed, reinforcement learning, is used for deciding sequential actions, like a robot deciding its path or its next move in a chess game. Reinforcement learning is not as common as the others and uses complex mathematics, like game theory. Reinforcement learning is the training of machine learning models to make a sequence of decisions. The agent learns to achieve a goal in an uncertain, potentially complex environment. In reinforcement learning, an artificial intelligence faces a game-like situation. The computer employs trial and error to come up with a solution to the problem. To get the machine to do what the programmer wants, the artificial intelligence gets either rewards or penalties for the actions it performs. Its goal is to maximize the total reward.