3. Types of Machine Learning

Thursday, 8 January 2026 1:20 PM

Types of Machine Learning:

1. Supervised Machine Learning
2. Unsupervised Machine Learning
3. Semi supervised Machine Learning
4. Reinforcement

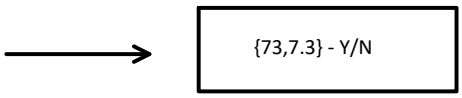
1. Supervised Machine Learning:

- Regression
- Classification

1. **Supervised Machine Learning: Supervised Machine Learning** is a subcategory of AI where an algorithm is trained using a **labeled dataset** containing both **input** and **output** data. The primary goal is to establish a mathematical **relationship** (or mapping) between these variables. Once the model understands this relationship, it can accurately predict outcomes or take decisions when presented with new, unseen data in the future.

Example:

IQ	CGPA	Y/N Placement
87	7.1	Y
111	8.9	Y
75	6.3	N



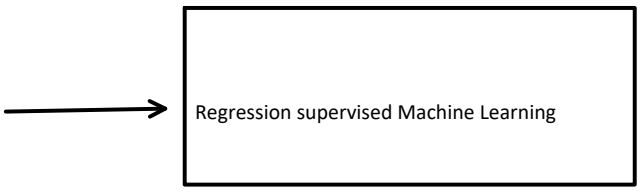
- *Regression: If your output or target column is numerical than this type of Machine Learning is Regression.*
- *Classification*

Before learning Regression and Classification we need to understand data types.
Data can be categorized into two parts.

1. *Numerical - Age, Weight, CGPA, IQ*
2. *Categorical - Gender, Nation*

Example:

IQ	CGPA	Package
8.7	*8.9*	*6*
150	*9.1*	*4*



2. Essential Prerequisites: Understanding Data Types

Before deciding which type of Supervised Learning to use, we must categorize our data:

1. **Numerical Data:** Continuous values used for measurement (e.g., Age, Weight, CGPA, IQ, or Salary Package).
2. **Categorical Data:** Data that represents groups or labels (e.g., Gender, Nationality, "Spam" vs. "Not Spam").

3. Two Main Types of Supervised Learning

Supervised learning is divided based on the nature of the **target (output)** column:

A. Regression

- **Definition:** Used when the output or target column is **numerical**.
- **Example:** Predicting the **price** of a house based on its square footage or location.

B. Classification

- **Definition:** Used when the output is a **category** or discrete label.
- **Examples:**
  - **Email Filtering:** Deciding if an email is "SPAM" or "Not SPAM".
  - **Weather Prediction:** Determining if it will "Rain" or "Not Rain" today.

Summary

Feature	Regression	Classification
Output Type	Numerical / Continuous	Categorical / Discrete
Example Goal	Predict a specific value (e.g., Price)	Assign a label (e.g., Yes/No)

Unsupervised Machine Learning

1. Definition

Unsupervised Machine Learning is a type of algorithm that learns from **unlabeled data**. Unlike supervised learning, the system is not given a "teacher" or an "answer key".

- **The Goal:** To find **hidden patterns**, structures, or clusters within the data on its own.
- **The Process:** The model looks for similarities or differences in the input features to organize the data into meaningful groups.

2. Key Sub-types of Unsupervised Learning

As noted in your initial chart, there are four main categories:

A. Clustering

- **Definition:** Grouping data points together so that items in the same group are more similar to each other than to those in other groups.
- **Example:** A brand grouping customers into "High Spenders" and "Budget Shoppers" based on their purchase history.

B. Dimensionality Reduction

- **Definition:** The process of reducing the number of variables (features) under consideration by obtaining a set of principal variables.
- **Purpose:** It helps simplify complex data without losing the important information, making it easier to visualize.
- Reducing number of columns. Let's say 1000 input columns. PCA is algorithms. Feature extraction we call it. Graphs we can't plot everything, we need to reduce some data.
- MNIST data

C. Anomaly Detection

- **Definition:** Identifying rare items, events, or observations which raise suspicions by differing significantly from the majority of the data.
- **Example:** Detecting **fraudulent transactions** in a credit card statement where one purchase looks completely different from your usual spending habits.

D. Association

- **Definition:** Discovering rules that describe your data, such as "People who buy X also tend to buy Y".
- **Example:** Market Basket Analysis—realizing that customers who buy bread and milk often buy eggs as well.

3. Comparison: Supervised vs. Unsupervised

Feature	Supervised Learning	Unsupervised Learning
Input Data	Labeled (Input + Output)	Unlabeled (Input only)
Feedback	Direct (Teacher provides answers)	No feedback (System finds patterns)
Task	Prediction / Classification	Pattern discovery / Clustering

Test:

"If I give you 1,000 photos of random animals but don't tell you what they are, and you group them by 'has wings' and 'has four legs,' are you doing Supervised or Unsupervised learning?" (Answer: Unsupervised, because you weren't given labels; you found the patterns yourself!)

Semi-supervised Learning is an approach where the algorithm is trained on a small set of labeled data and a significantly larger set of unlabeled data. The model uses the labeled data to learn the basic categories and the unlabeled data to understand the overall shape and distribution of the data.

- **The Problem:** Labeling data is expensive and time-consuming (requires humans).
- **The Solution:** Use what little labeled data you have to "guide" the interpretation of the vast amount of unlabeled data.

2. How It Works (The Process)

1. **Initial Training:** The model is trained on the small portion of **labeled data**.
2. **Pseudo-Labeling:** The model then looks at the **unlabeled data** and "guesses" their labels based on what it just learned.
3. **Refinement:** The model treats its own best guesses as real labels and retrains itself on the newly expanded dataset to improve accuracy.

3. Why Use Semi-Supervised Learning?

In the real world, data is often "partially labeled":

- **Cost Efficiency:** You don't have to pay experts to label millions of images; you only need them to label a few thousand.
- **Improved Accuracy:** Even though most data is unlabeled, the *volume* of that data helps the model understand the patterns better than it could with just a tiny labeled set.

4. Real-World Examples

- **Medical Imaging:** A hospital might have 100,000 X-rays (unlabeled), but only 500 have been analyzed and labeled by a radiologist. Semi-supervised learning uses the 500 to help "understand" the other 99,500.
- **Speech Analysis:** Learning to recognize a specific voice. Labeling every single word in a 10-hour recording is hard, but labeling just a few minutes helps the computer understand the rest of the audio.
- **Web Content Classification:** Grouping billions of webpages into categories (News, Sports, Tech) using only a few websites that have already been categorized by hand.

Summary Comparison

Type	Data Used	Teacher?
Supervised	100% Labeled	Full Supervision
Unsupervised	100% Unlabeled	No Supervision
Semi-Supervised	Small Labeled + Large Unlabeled	**Partial Supervision**

Reinforcement Learning (RL)

1. Definition

Reinforcement Learning is a type of machine learning where an **Agent** learns to make decisions by performing actions in an **Environment** to achieve a goal. It learns through a system of **rewards** and **punishments**.

- **The Core Concept:** It is "Trial and Error" learning.
- **The Goal:** To maximize the total "reward" over time.

2. The Five Key Components

To explain this to your class, use the analogy of training a dog or playing a video game:

1. **Agent:** The "student" or the AI (e.g., the dog).
2. **Environment:** The world the agent lives in (e.g., the park).
3. **Action:** What the agent does (e.g., "Sit" or "Bark").
4. **State:** The current situation (e.g., the dog is standing).
5. **Reward:** Feedback from the environment (e.g., a treat for sitting, or no treat for barking).

4. Real-World Applications

- **Gaming:** This is how AI like *AlphaGo* or *Stockfish* (chess) became world champions. They played millions of games against themselves to learn winning moves.
- **Robotics:** Teaching a robot to walk. Every time it takes a successful step, it gets a "point"; every time it falls, it loses a "point."
- **Self-Driving Cars:** Learning when to brake or accelerate based on the safety and speed of the ride.
- **Recommendation Engines:** Netflix or YouTube "learn" to show you videos that keep you watching (the "reward" is your watch time).

The "Big Four"

Type	Data Type	Key Learning Method
Supervised	Labeled	Learns from an answer key (Teacher).
Unsupervised	Unlabeled	Finds hidden patterns (Self-discovery).
Semi-Supervised	Mixed	Uses a few clues to solve a big puzzle.
Reinforcement	Interactive	Learns by trial and error (Rewards).