# Machine Learning

What is Machine Learning?

Field of study that gives computers the ability to learn without explicitly programmed *-Arthur Samuel (1959)*

Types of Machine Learning

Machine Learning can be broadly categorized into the following types:

1. Supervised Learning

- **Definition**: The model is trained on labeled data (input → output).
- **Goal**: Learn the mapping between inputs and outputs.
- **Key Characteristic:** You give your ML algorithm input with correct output and by seeing correct pairs of input and output, eventually algorithm learns to take just input without output and give right or close to right answer or guess.
- **Examples**:
    - Predicting house prices (Regression)
    - Spam detection (Classification) ![[Pasted image 20250829222847.png]]

Regression:

Regression is a **Supervised Learning** technique used to predict a **continuous numerical value** (not categories).
It learns the relationship between input variables (X) and output variable (Y).

- Here, we want to know the price of a house whose area is $750 ft^2$. ![[Pasted image 20250829230102.png]]

Classification:

Classification is a **Supervised Learning** technique used to predict a **discrete class label** (categories).
The model learns to map input variables (X) to output classes (Y).
Breast Cancer Detection: Using a machine learning system tries to figure out if a tumor (a lump) that is malignant (cancerous) or benign.

- Malignant -> 0
- Benign -> 1 This different from regression where we are predicting any number but here we are predicting small range of numbers (category) like in this case (0/1).
- We can also plot it on a line, there is only two categories. ![[Pasted image 20250829232315.png]]

In classification, we have more than two output categories.(*May be ML system can produce multiple types of cancer as output if it finds it malignant.*) Let's say there are two types of malignant cancers.

- Malignant Type 1
- Malignant Type 2 ![[Pasted image 20250829233257.png]]

In classification, class and category is same thing. Classification doesn't only predict only numeric categories but also non-numeric. like cat and dog.

Just only knowing output, we can have multiple inputs like tumor size and age. Here, Algorithm finds a boundary that separates both malignant and benign. ![[Pasted image 20250830000735.png]] Here, in this example, it is less likely for patient to have breast cancer.

In practical problems, more inputs is used as input for the ML algorithm.

Conclusion:

- **Supervised Learning** -> learning from the right answers.
- **Regression** -> Predicts a number infinitely many possible outputs.
- **Classification** -> Predict small numbers of possible outputs.

---

2. Unsupervised Learning

- **Definition**: The model is trained on unlabeled data, finding hidden patterns or structures.
- **Goal**: Discover relationships or groupings in data.
- **Examples**:
    - Customer segmentation (Clustering)
    - Market basket analysis (Association) **We call it unsupervised learning because we are not trying to supervise the algorithm to give some quote right answers for some input instead, we ask the system to figure out all by itself what interesting patterns or structures might be in the data**

Take the example of Breast Cancer Dataset.

- Here, the unsupervised algorithm might decide that the data can be assigned to two different groups or clusters. This is a particular type of unsupervised learning called **Clustering**. ![[Pasted image 20250830121429.png]]

Google news is an example of clustering.

- For instance, consider the news about the birth of twin panda cubs at a zoo. Google News shows us different articles from various sources, but it groups them together because they share similar words and themes such as *panda*, *twin*, and *zoo*.
- There is not a person in google who is doing this clusters because it is impossible because there is thousands news everyday that's why the algorithm has to figure it out on its own about clusters of news articles. ![[Pasted image 20250830122343.png]]

Let's take another example of clustering, DNA microarray.

- Here, each row shows a gene and each column is a person.
- For instance, a gene row might dictate in a person that which kind of eye color he has, how much tall he is, which foods likes or doesn't like and many more.

- Green and red colors shows that how much that gene is active in a person.
- Algorithm will group individuals with same genes without any supervision. ![[Pasted image 20250830123812.png]]

Here's another example of grouping customers.

- Grouping customers of a course to find why they are doing the course. ![[Pasted image 20250830124420.png]]

Let's look another types of unsupervised learning algorithms.

- **Anomaly Detection: Anomaly detection** means finding data points or events that are unusual compared to the normal pattern. *In a bank, if someone usually withdraws small amounts of money but suddenly withdraws a very large amount at midnight, the system may flag it as an anomaly (possible fraud).*
- **Dimensionality reduction: Dimensionality reduction** is the process of reducing the number of features while keeping as much important information as possible. *Suppose you have 100 measurements of a face (eye size, nose length, jaw width, etc.). Dimension reduction might compress them into just 2–3 features while still distinguishing between different faces.*

---

3. Semi-Supervised Learning

- **Definition**: Uses a mix of labeled and unlabeled data.
- **Goal**: Improve learning accuracy when labeled data is scarce.
- **Examples**:
    - Fraud detection
    - Medical diagnosis with limited labeled cases

---

4. Reinforcement Learning

- **Definition**: The model learns by interacting with an environment and receiving rewards or penalties.
- **Goal**: Maximize cumulative reward.
- **Examples**:
    - Game-playing AI (e.g., Chess, Go)
    - Self-driving cars

---

5. Self-Supervised Learning (Emerging)

- **Definition**: The model generates its own labels from input data.
- **Goal**: Learn representations without explicit labeling.
- **Examples**:
    - Large language models (e.g., GPT)
    - Vision models (e.g., contrastive learning)