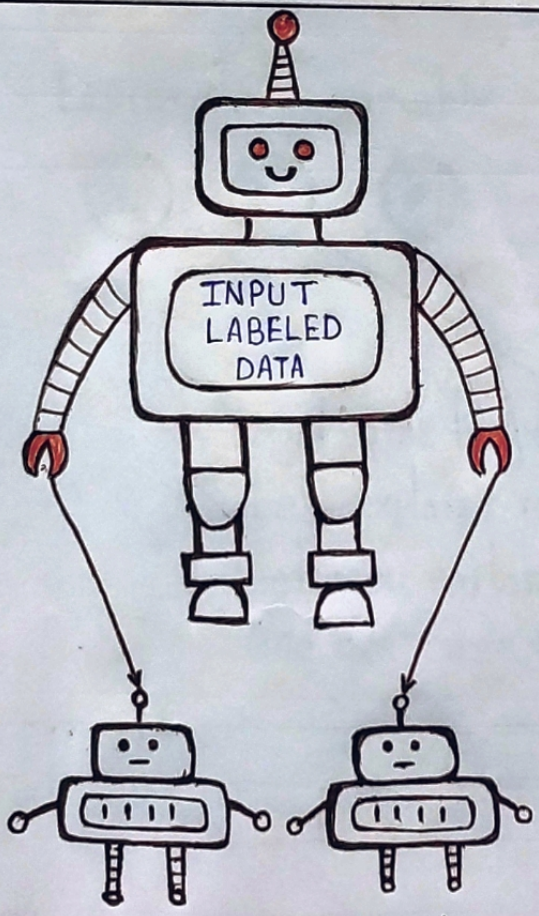
# Basic Concept About Supervised vs. Unsupervised Learning

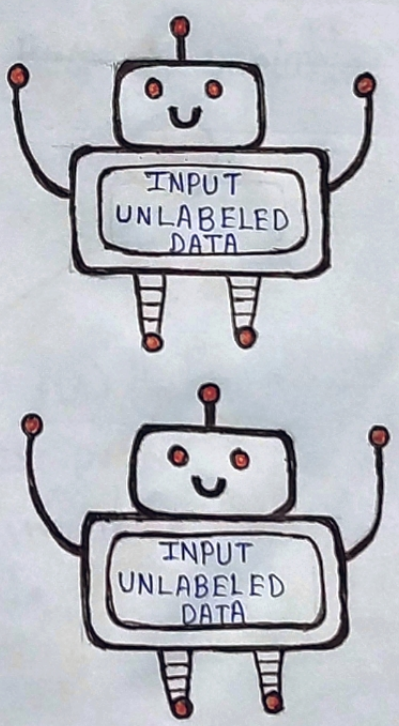
The field of machine learning contains a massive set of algorithms that can be used for understanding data. These algorithms can be classified into one of two categories:

1. Supervised Learning Algorithms: Involves building a model to estimate or predict an output based on one or more inputs.

2. Unsupervised Learning Algorithms: Involves finding structure and relationship from inputs. There is no "supervising" output.




Supervised Robots



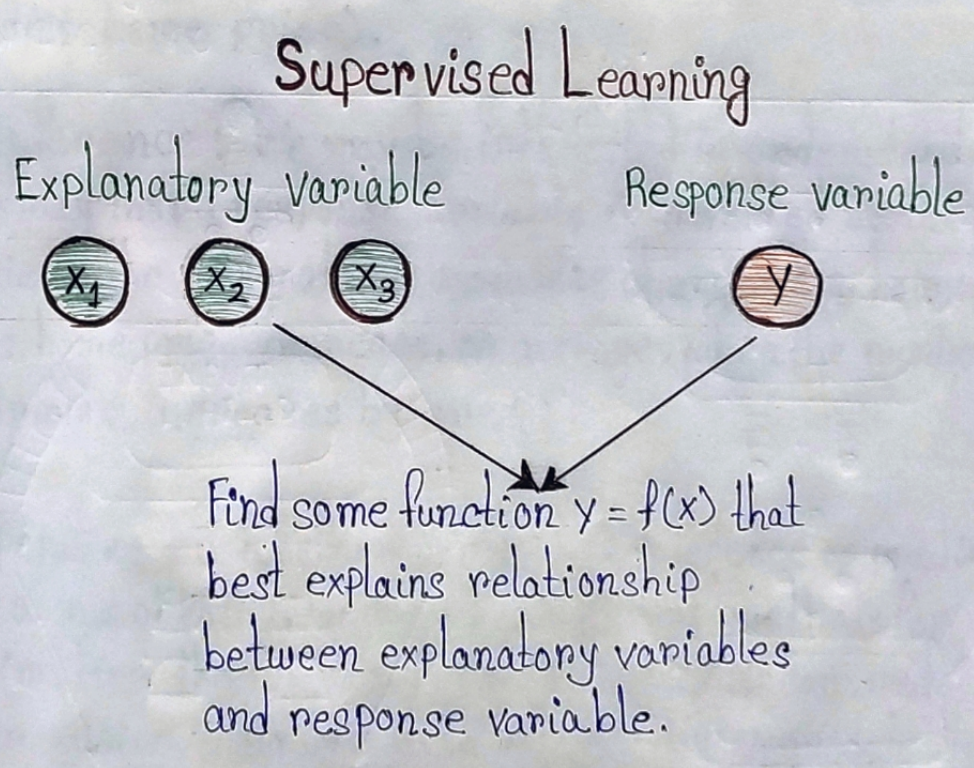

Unsupervised Robots

# Supervised Learning Algorithms

A supervised learning algorithm can be used when we have one or more explanatory variables ($X_1, X_2, X_3, \ldots, X_p$) and a response variable ($Y$) and we would like to find some function that describes the relationship between the explanatory variables and the response variable:

$$Y = f(X) + \varepsilon$$

Where $f$ represents systematic information that $X$ provides about $Y$ and where $\varepsilon$ is a random error term independent of $X$ with a mean of Zero.



There are two main types of supervised learning algorithms:

1. Regression: The output variable is continuous (e.g. weight, height, time, etc.)

2. Classification: The output variable is categorical (e.g. male or female, pass or fail, benign or malignant, etc.)

There are two main reasons that we use supervised learning algorithms:

1. Prediction: We often use a set of explanatory variables to predict the value of some response variable (e.g. using square footage and number of bedrooms to predict home price).

2. Inference: We may be interested in understanding the way that a response variable is affected as the value of the explanatory variables change (e.g. how much does home price increase, on average, when the number of bedrooms increases by one?).

Depending on whether our goal is inference or prediction (or a mix of both), we may use different methods for estimating the function $f$. For example, linear models offer easier interpretation but non-linear models that are difficult to interpret may offer more accurate prediction.

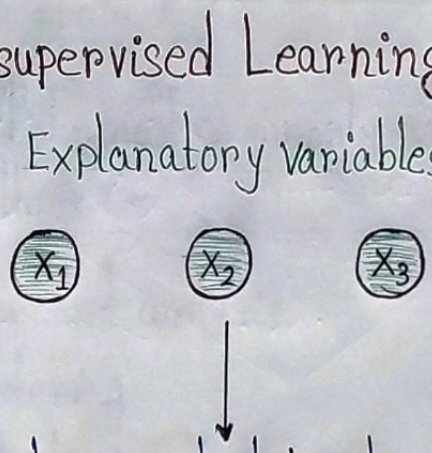
Here is a list of the most commonly used supervised learning algorithms:

- Linear regression

- Logistic regression
- Linear discriminant analysis
- Quadratic discriminant analysis
- Decision trees
- Naive bayes
- Support vector machines
- Neural networks

## Unsupervised Learning Algorithms

An unsupervised learning algorithm can be used when we have a list of variables $(X_1, X_2, X_3, \ldots, X_p)$ and we would simply like to find underlying structure or patterns within the data.



There are two main types of unsupervised learning algorithms :

1. Clustering: Using these types of algorithms, we attempt to find "clusters" of observations in dataset that are similar to each other. This is often used in retail when a company would like to identify clusters of customers who have similar shopping habits so that they can create specific marketing strategies that target certain clusters of customers.
2. Association: Using these types of algorithms, we attempt to find "rules" that can be used to draw associations. For example, retailers may develop an association algorithm that says "if a customer buys product X they are highly likely to also buy product Y."

Here is a list of the most commonly used unsupervised learning algorithms :

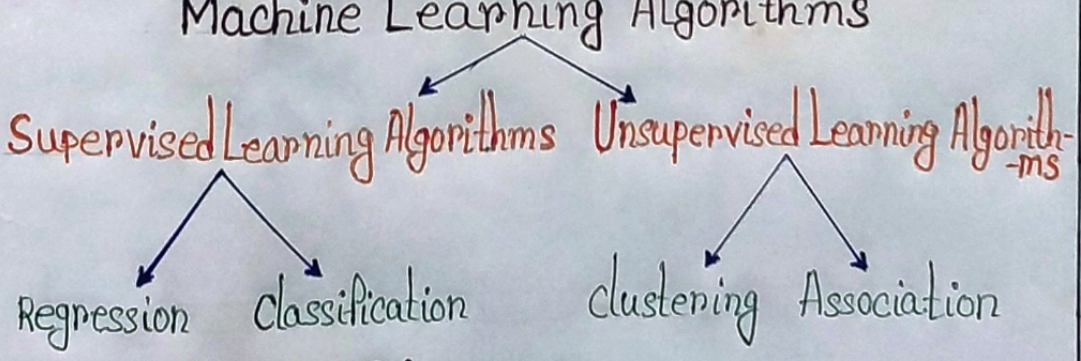
- Principal component analysis
- K-means clustering
- K-medoids clustering
- Hierarchical clustering
- Apriori algorithm

# Summary: Supervised vs. Unsupervised Learning

The following table summarizes the differences between Supervised and unsupervised Learning algorithms:

## TABLE

| | Supervised Learning | Unsupervised Learning |
|---|---|---|
| Description | Involves building a model to estimate or predict an output based on one or more inputs. | Involves finding structure and relationships from inputs. There is no "Supervising" output. |
| Variables | Explanatory and Response variables. | Explanatory variables only. |
| End goal | Develop model to (1) predict new values or (2) understa--nd existing relationship between explanatory and response variables. | Develop model to (1) place observations from a dataset into a specific clust--er or to (2) create rules to identify associations between variables. |
| Types of algorithms | (1) Regression and (2) classification. | (1) Clustering and (2) Association |




## Diagram