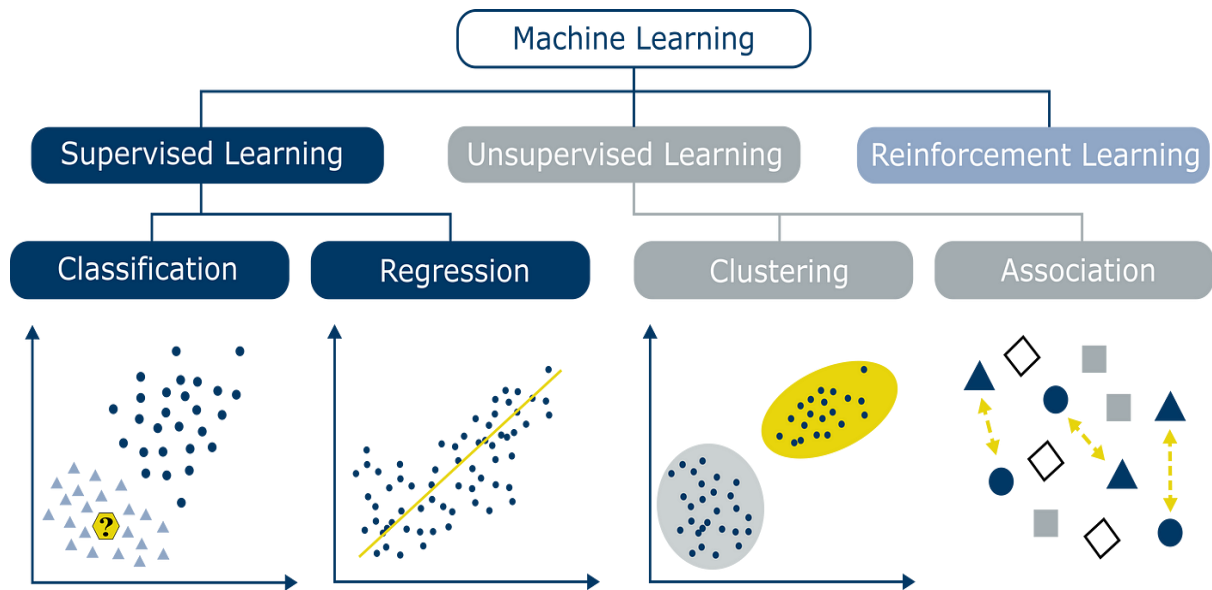


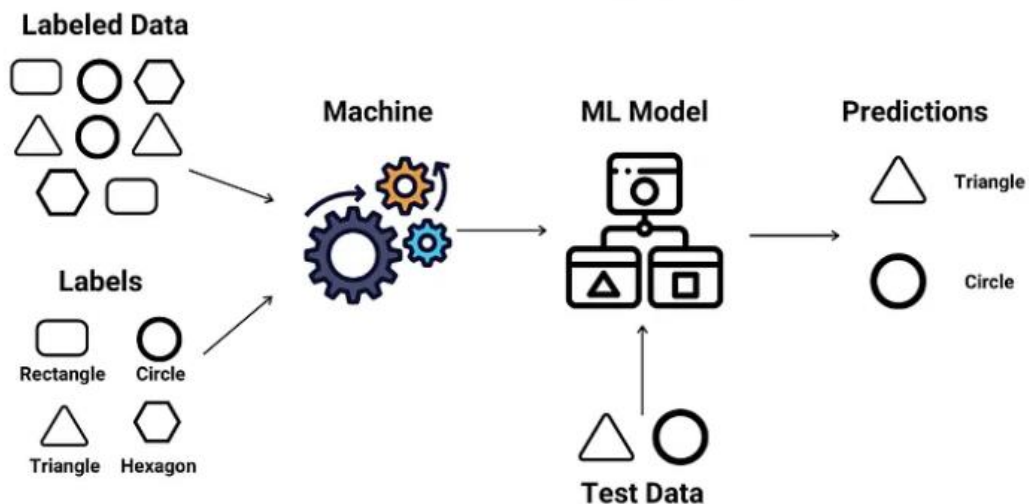
Supervised Learning-Regression

Introduction to Regression analysis



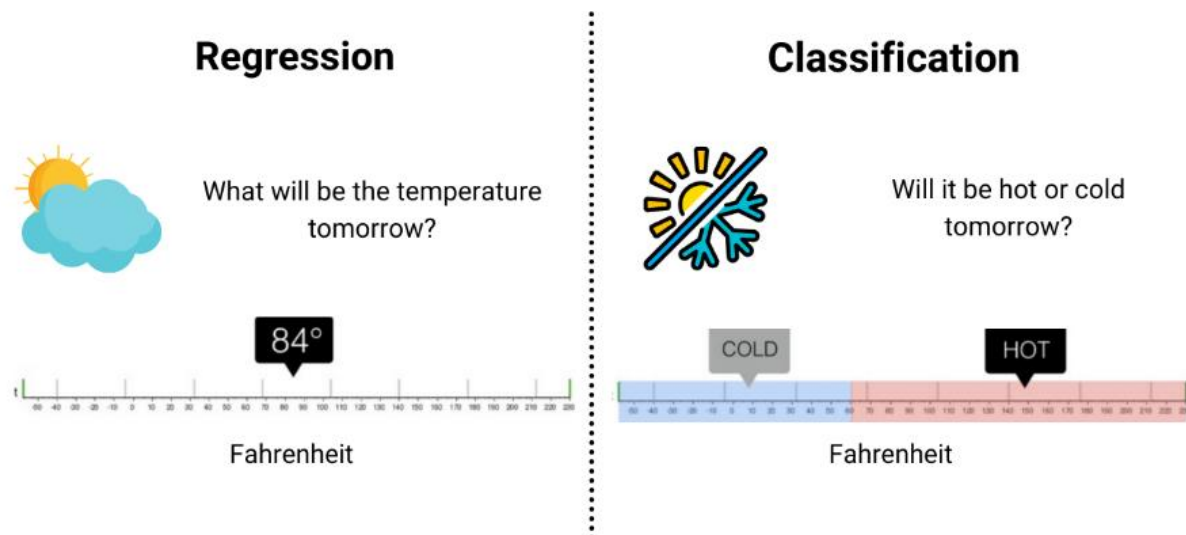
Supervised Learning

Supervised learning is a type of machine learning where an algorithm is trained on labeled data. The model learns from input-output pairs, where the correct output is already known. The goal is to predict the output for new, unseen data.



There are two main types of supervised learning:

1. **Classification:** Predicts discrete labels (e.g., spam or not spam).
2. **Regression:** Predicts continuous values (e.g., house prices).



Classification

Classification models classify our outputs to certain categories. If the number of categories are only two then it is specially called binary classification. For greater number of categories it is called multi-class classification.

Some examples are:

- *Whether a patient has cancer or not*

Regression

Regression models are for labeling outputs with continuous values.

- Predicting house prices, or
- How long is it gonna take you to get home

are both for regression models because the results are ever changing.

The key components of supervised learning include:

1. **Input Features (X):** These are the independent variables or attributes used by the model to make predictions. They could be anything from numerical data, such as age or income, to more complex data like images or text.
2. **Output Labels (Y):** These are the dependent variables or target outcomes that the model is trying to predict. In classification tasks, the outputs are discrete labels (e.g., "dog" or "cat"), while in regression tasks, the outputs are continuous values (e.g., house prices).

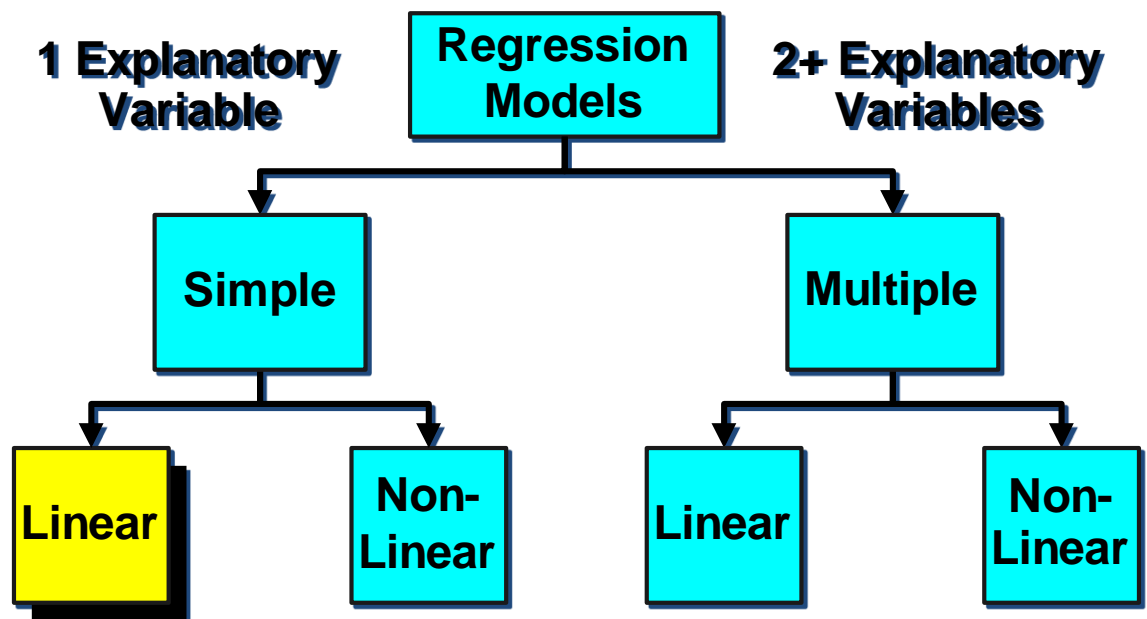
3. **Training Data:** A labeled dataset that the model uses to learn. Each example in the dataset consists of an input (X) and its corresponding correct output (Y).
4. **Learning Algorithm:** This is the method the model uses to learn the relationship between inputs and outputs.

Introduction to Regression Analysis:

Regression analysis is a key technique in supervised learning used to predict continuous outcomes. It helps in understanding the relationship between a dependent variable (the outcome) and one or more independent variables (the inputs or predictors). The primary goal of regression is to create a model that can predict the value of the dependent variable given the independent variables.

There are various types of regression techniques, with the most common being:

1. **Linear Regression:** Models the relationship between the dependent variable (Y) and one or more independent variables (X) by fitting a straight line ($Y = bX + c$).
2. **Multiple Regression:** Extends linear regression to multiple input variables.
3. **Polynomial Regression:** Fits a curve to the data instead of a straight line.
4. **Logistic Regression:** Though called regression, it's used for classification tasks, predicting the probability of an event.



Simple Linear Regression:

Simple linear regression is a model that assesses the relationship between a dependent variable and an independent variable.

The simple linear model is expressed using the following equation:

$$y=mx+b$$

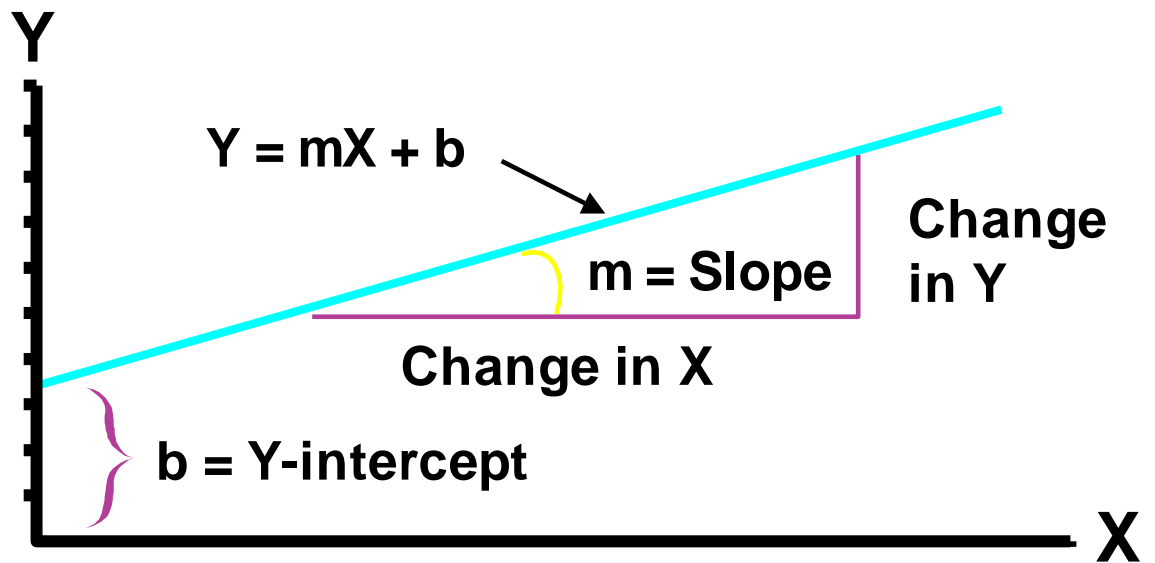
where,

y - Dependent variable

x - Independent variable

m - Slope

b - Intercept



Example: Suppose there is a marketing company A, who does various advertisement every year and get sales on that. The below list shows the advertisement made by the company in the last 5 years and the corresponding sales:

Advertisement	Sales
\$90	\$1000
\$120	\$1300
\$150	\$1800
\$100	\$1200
\$130	\$1380
\$200	??