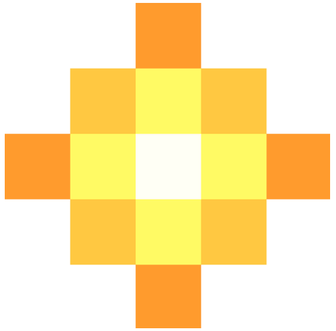


Introduction to ML

Lecture 1

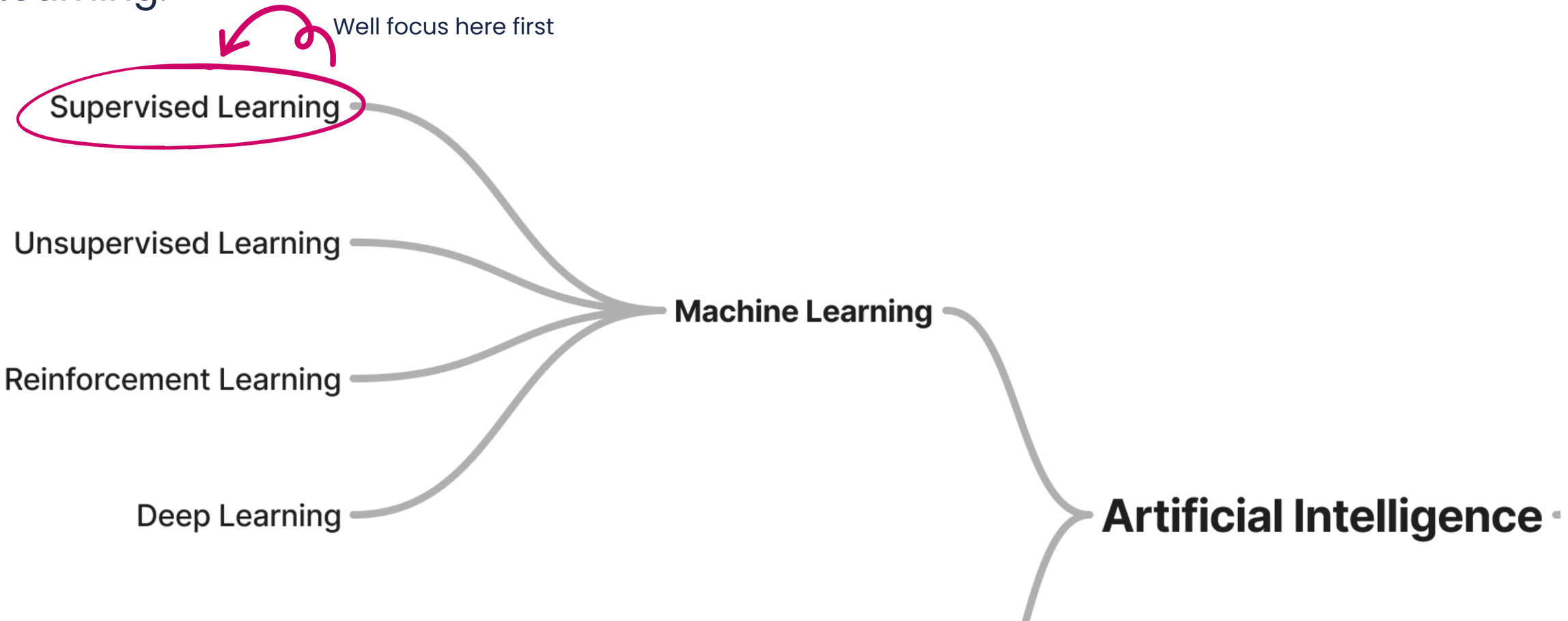
neumentora

What is Machine Learning?

Machine learning is a subdivision of **artificial intelligence** based on the biological learning process, dealing with algorithms to learn from **machine-readable data** for various applications in science and engineering.

ML Branches

Our conversation will cover supervised learning, unsupervised learning, and reinforcement learning. In the final part, we will focus specifically on deep learning.



Supervised Learning

This week, we will delve deeper into supervised learning and explore several algorithms, including:

- **Linear Regression**
- **Logistic Regression**
- **Decision Trees**

Next week, our topics will include:

- Support Vector Machines (SVMs)
- Random Forests
- Neural Networks

Linear Regression

What is Linear Regression?

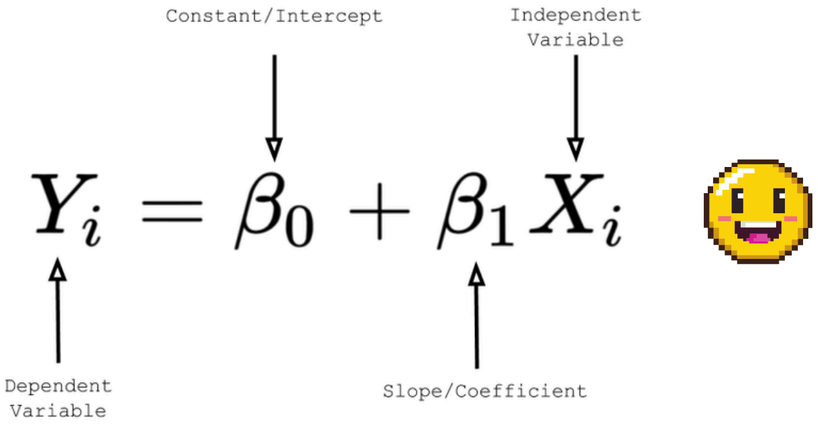
- Linear Regression is a fundamental statistical technique.
- It models the relationship between a **dependent variable** and **one or more independent variables**.
- Assumes a linear relationship between the variables.
- Changes in independent variable(s) correspond to proportional changes in the dependent variable.
- Aims to find the **"best fit" straight line** representing the relationship.

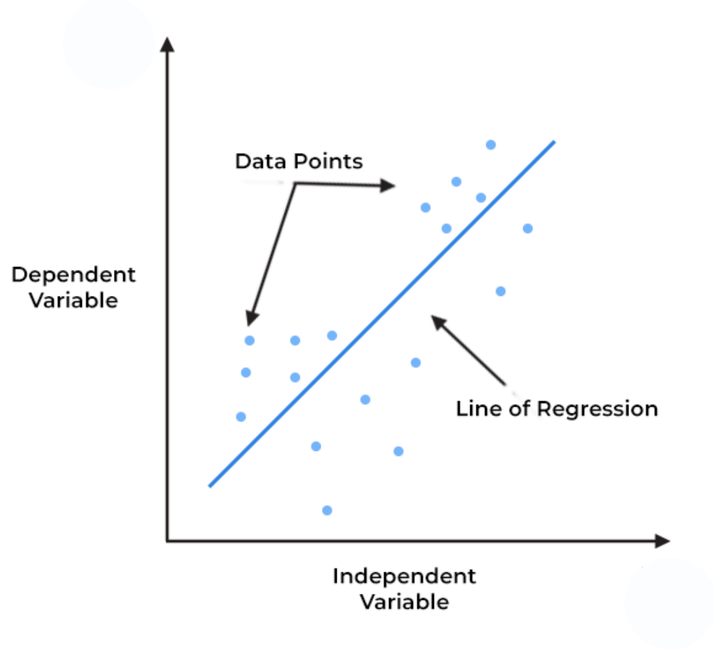
Example: Predicting house prices based on features like **size** and **location**. Helps determine the contribution of each feature to the **overall price**.

Linear Regression

Types of Linear Regression

- **Simple Linear Regression**
  - One dependent variable (interval or ratio)
  - One independent variable (interval or ratio or dichotomous)



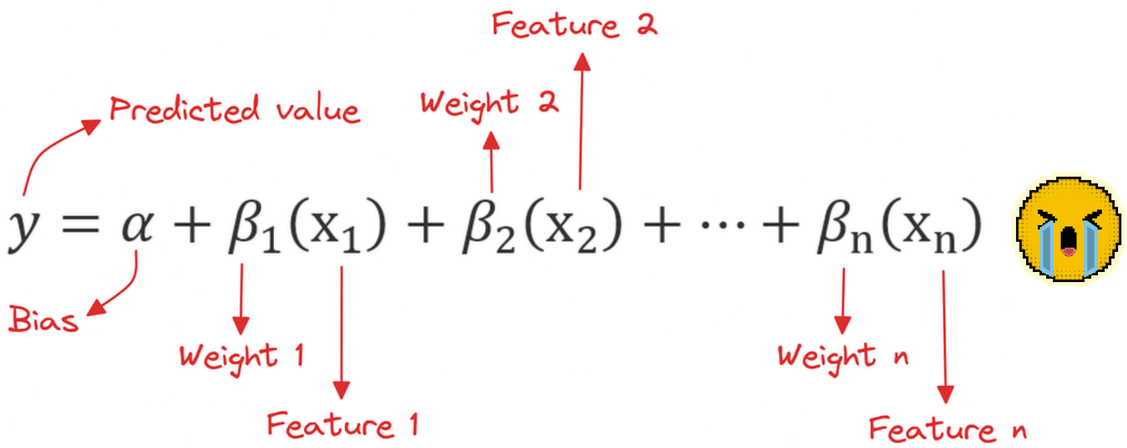


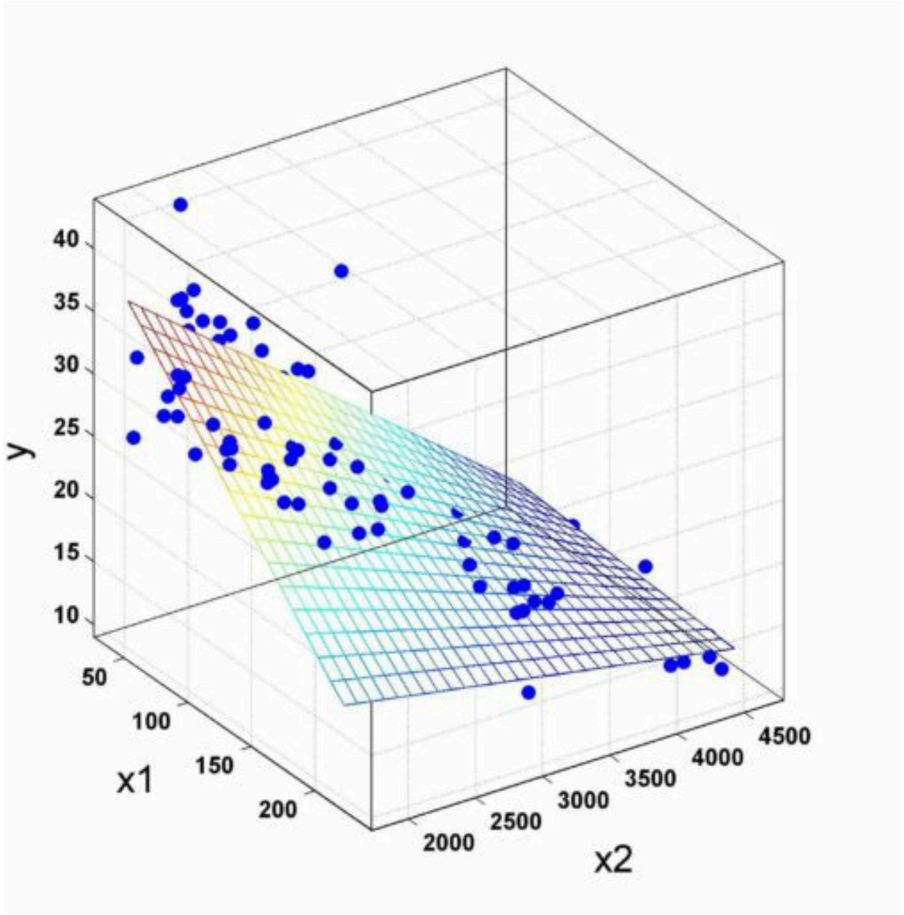
All images are collected from multiple online sources

Linear Regression

Types of Linear Regression

- **Multiple Linear Regression**
  - One dependent variable (interval or ratio)
  - Two or more independent variables (interval or ratio or dichotomous)





All images are collected from multiple online sources

Linear Regression

Evaluation Metrics for Linear Regression

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y})^2$$

MSE

Mean Squared Error (MSE) measures how **far predictions are from actual values**, but it gives more weight to bigger errors by squaring them. This means larger mistakes hurt the score more than smaller ones.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}|$$

MAE

Mean Absolute Error (MAE) tells you how much, **on average, your predictions are off from the actual values**. It simply calculates the average of all the absolute differences between predicted and actual values. Smaller values mean better predictions.

$$RMSE = \sqrt{MSE}$$

RMSE

RMSE is just the square **root of MSE**. It's useful because it brings the error back to the same scale as the target variable (like house prices or sales).

$$R^2 = 1 - \frac{\sum(y_i - \hat{y})^2}{\sum(y_i - \bar{y})^2}$$

R-Sqrd

This metric shows how well the model **explains the variation in the dependent variable.** If is 1, it means the model explains all the variation; if it's 0, it explains none.

All the formulas are explained in code

Logistic Regression

What is Logistic Regression?

- Logistic Regression is a statistical method used for **classification.**
- It deals with **categorical variables** that have two outcomes (e.g., yes/no, true/false, 0/1).
- The primary goal is to predict the probability of a given input belonging to a specific class.
- It differs from linear regression, which predicts **continuous values.**
- Logistic regression fits data to a **logistic function** to model the likelihood of an event occurring.

Logistic Regression

Types of Logistic Regression

Binomial

Used when the target variable has **two** possible outcomes (e.g., yes/no, pass/fail).

Multinomial

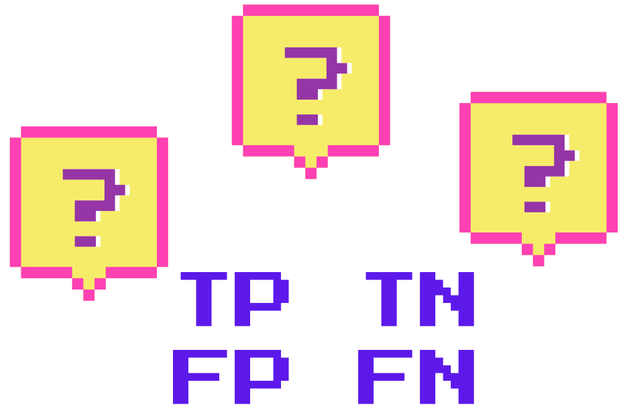
Used when the target variable has **three or more unordered categories** (e.g., predicting the type of fruit: apple, orange, or banana).

Ordinal

Used when the target variable has **three or more ordered categories** (e.g., customer satisfaction levels: low, medium, high).

Logistic Regression

Evaluation of Linear Regression Model



Accuracy

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision

$$Precision = \frac{TP}{TP + FP}$$

Recall

$$Recall = \frac{TP}{TP + FN}$$

F1-Score

$$F1\text{-}score = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

AUC-ROC

Logistic Regression

AUC-ROC?

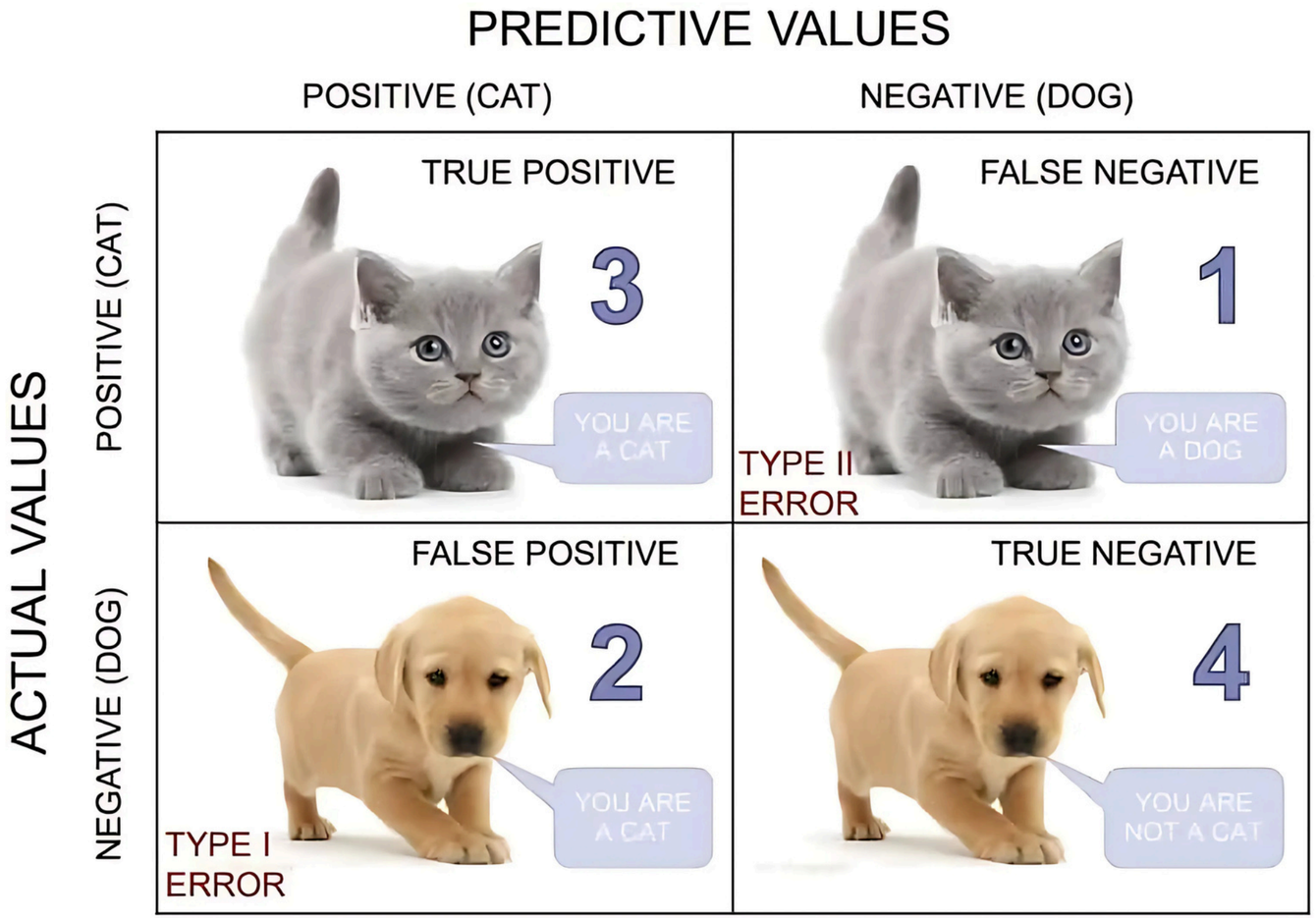
Evaluation of Linear Regression Model

TP = TRUE POSITIVE

TN = TRUE NEGATIVE

FP = FALSE POSITIVE

FN = FALSE NEGATIVE



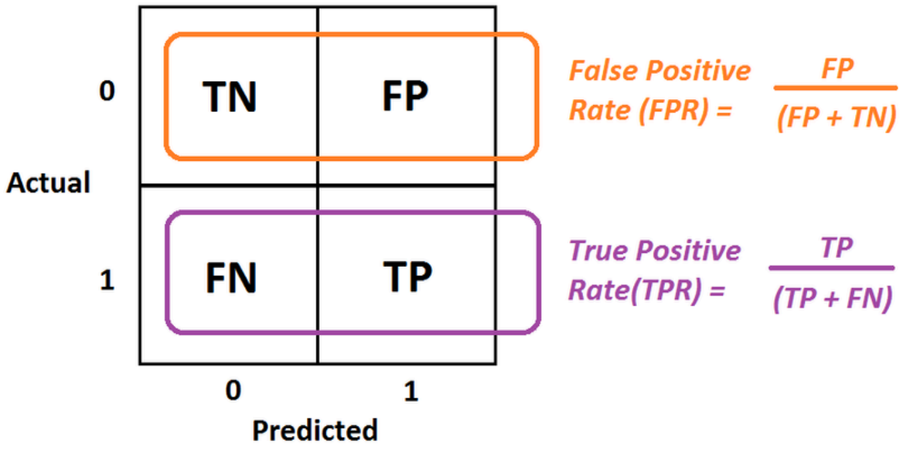
This is called confusion matrix

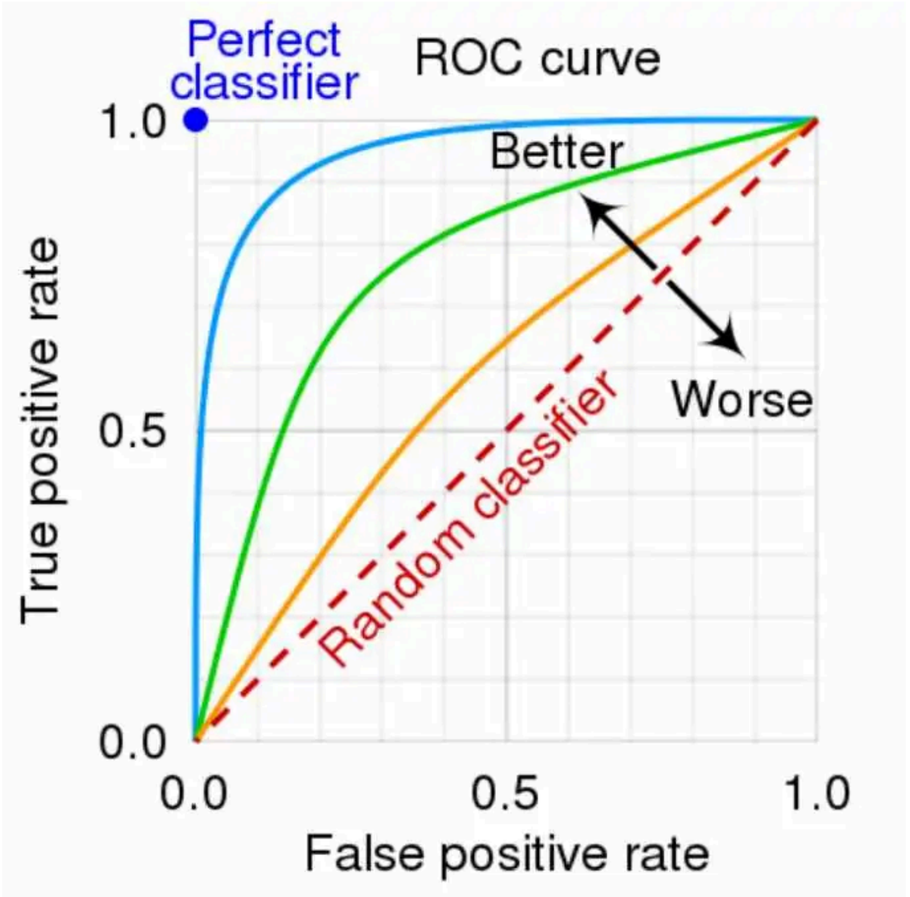
Logistic Regression

Evaluation of Linear Regression Model

AUC-ROC?

$$AUC = \int_0^1 TPR d(FPR)$$





Decision Trees

Watch this video along with the code in the following one to grasp the concept of the decision tree.