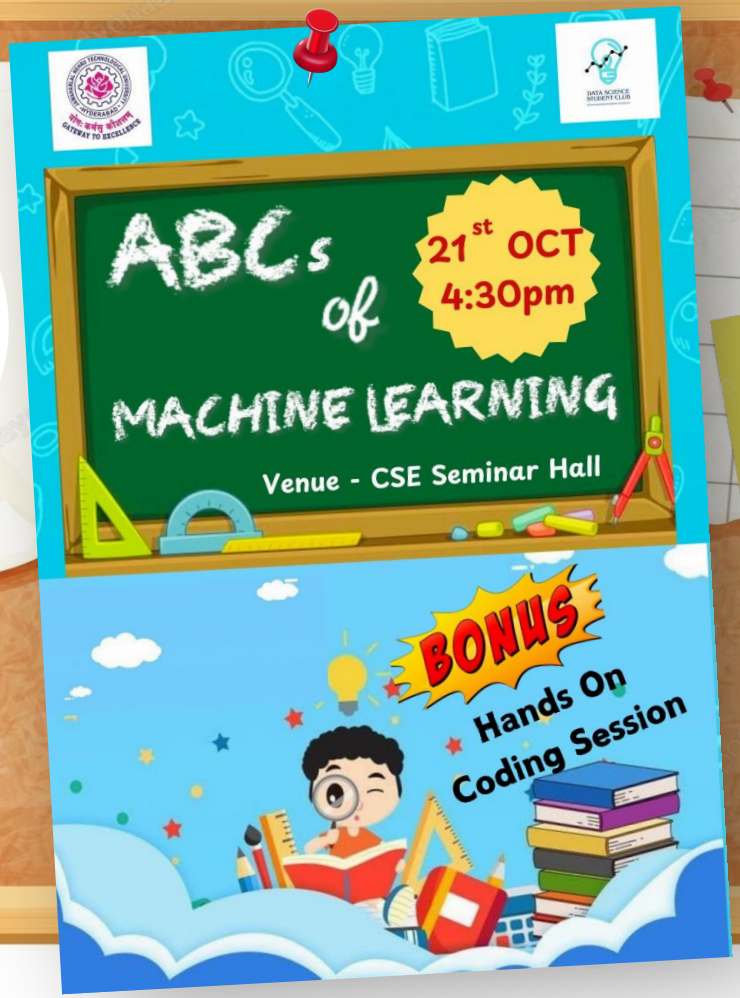


DATA SCIENCE STUDENT CLUB

Where Information Ignites Innovation

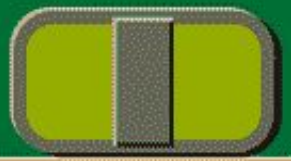
JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY • HYDERABAD •

योगः कर्मसु कौशलम्

GATEWAY TO EXCELLENCE

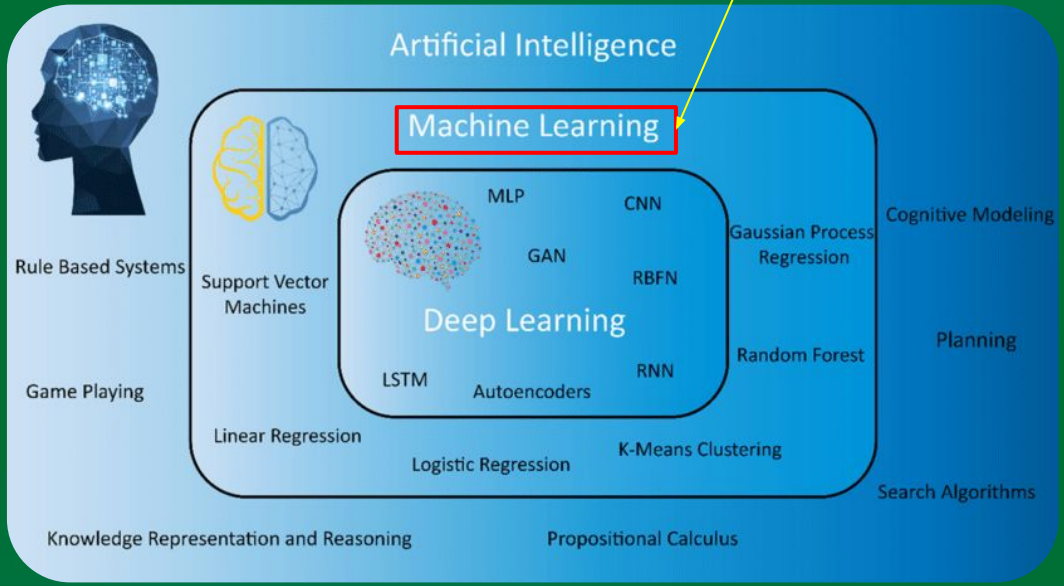
Nandini Maharaj
Manasvi G
R Yogender Reddy

ABCs of Machine Learning



What is Artificial Intelligence?

Artificial intelligence (AI) refers to computer systems capable of performing complex tasks that historically only a human could do, such as reasoning, making decisions, or solving problems.



https://www.researchgate.net/figure/Domains-of-AI-ML-DL-and-widely-used-algorithms_fig1_361501987

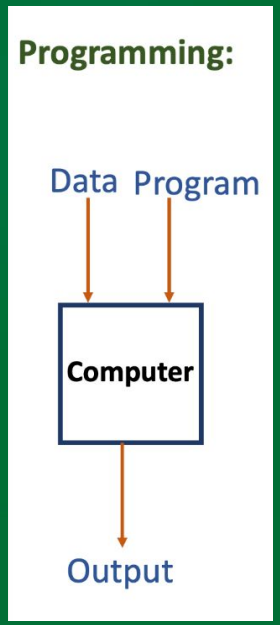
What is Machine Learning?

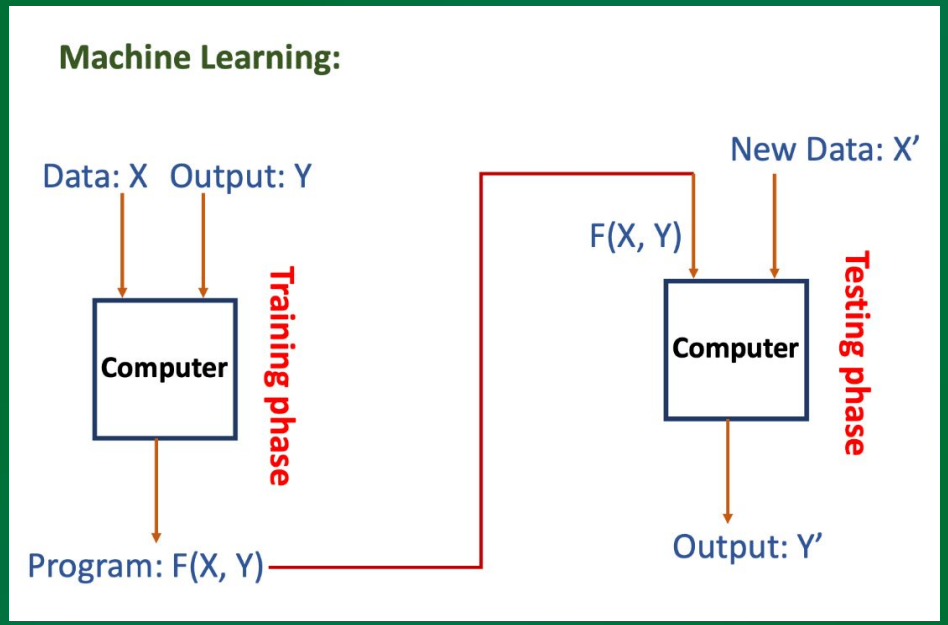
"The activity or process of gaining knowledge or skill by studying, practicing, being taught, or experiencing something."

Machine Learning is a set of algorithms or programs that enable systems to learn from data or experience and improve their performance over time without being explicitly programmed.

Usual Programming v/s Machine Learning

"Machine learning is a field of study that gives computers the ability to learn without being explicitly programmed." - Arthur Samuel





What is Machine Learning?

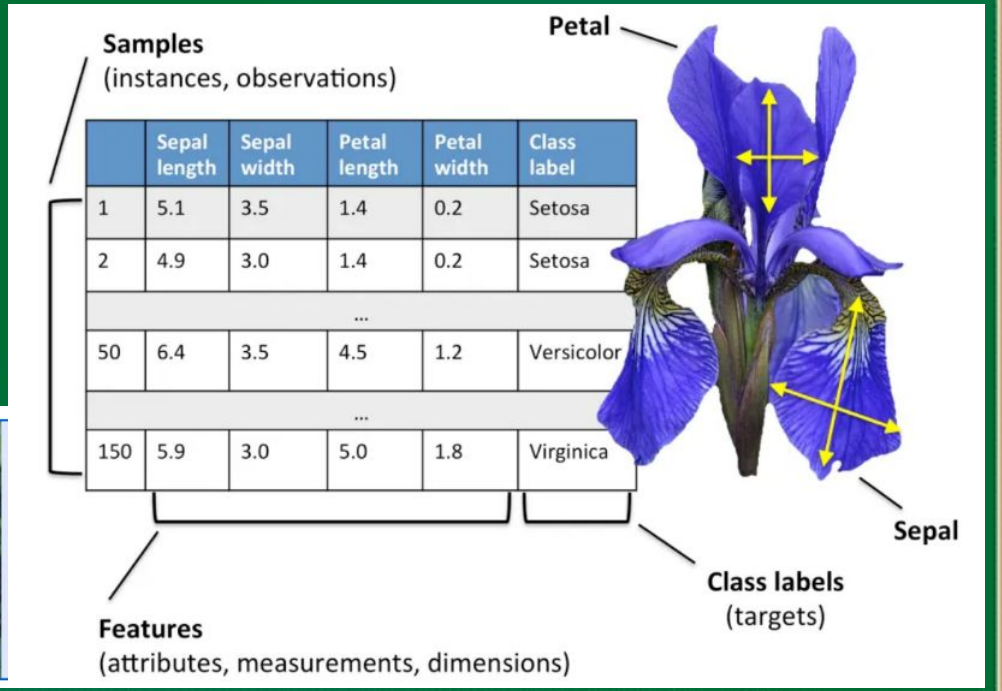
The function of a machine learning system can be:

- **Descriptive -** A telecom company analyzes customer call data to identify common reasons for service complaints.
- **Predictive -** An e-commerce platform predicts which products a customer is likely to purchase next based on browsing and purchase history.
- **Prescriptive -** A logistics company suggests optimal routes for drivers based on traffic, weather, and delivery schedules.

Data in Machine Learning

Dataset in Machine Learning

Problem Statement:
Create the model that can classify the different species of the Iris flower.



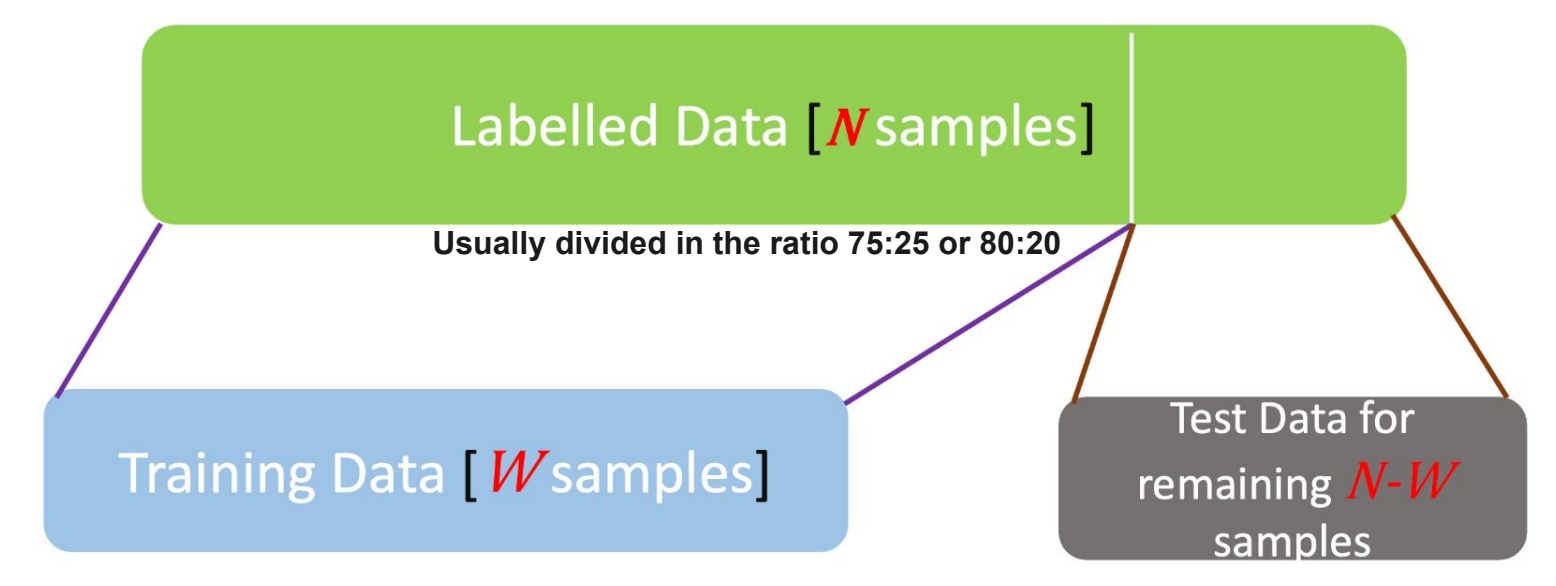
	Sepal length	Sepal width	Petal length	Petal width	Class label
1	5.1	3.5	1.4	0.2	Setosa
2	4.9	3.0	1.4	0.2	Setosa
...					
50	6.4	3.5	4.5	1.2	Versicolor
...					
150	5.9	3.0	5.0	1.8	Virginica

https://medium.com/@jebaseelanravi96/machine-learning-iris-classification-33aa18a4a983

Training & Testing Data

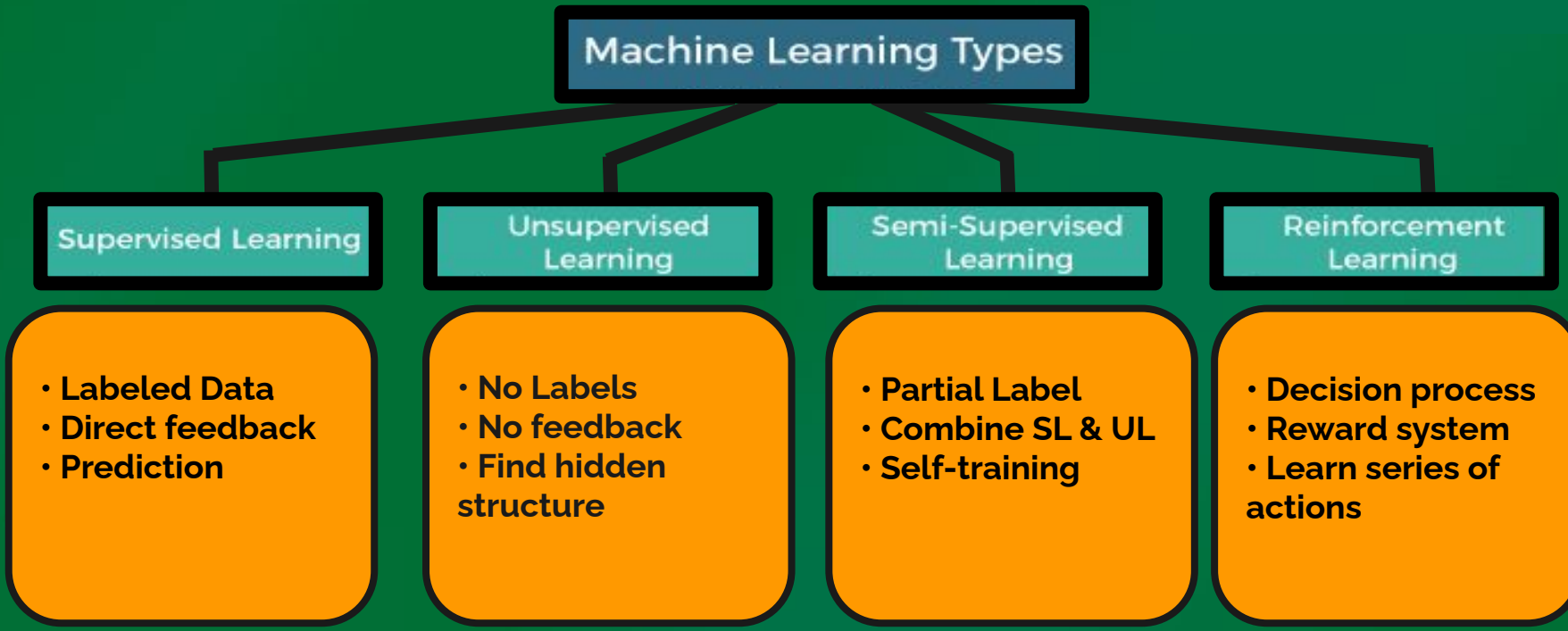
You should split the original dataset into two subsets:

1) A training set that the model trains on.
2) A test set for evaluation of the trained model.



Learning Techniques in Machine Learning

Learning Techniques in ML



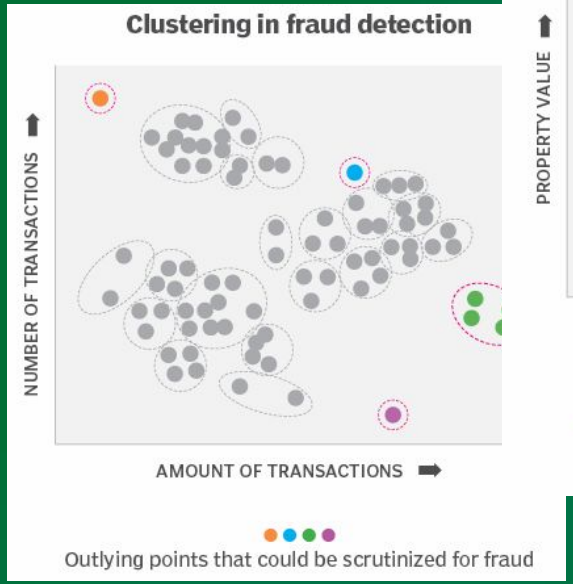
Supervised Learning

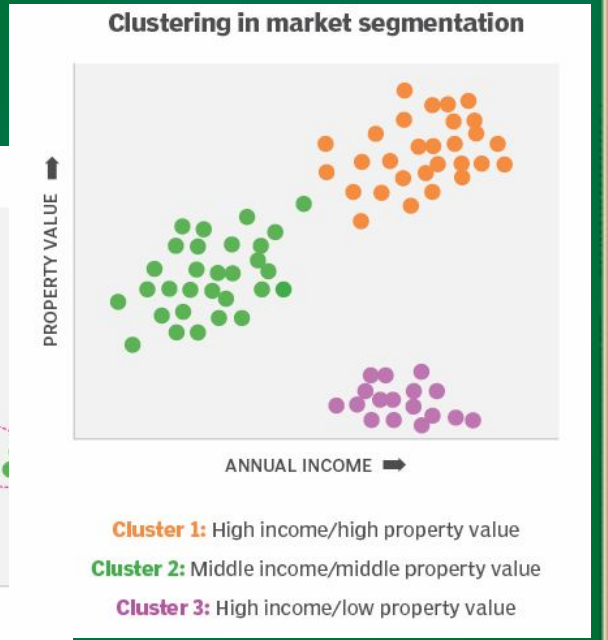
- Machine is taught by example.
- Dataset includes desired inputs and outputs
- Algorithm must determine the output given a new input.

https://c3.ai/introduction-what-is-machine-learning/supervised-learning/

Unsupervised Learning

- Models are trained using unlabeled dataset
- Learn interesting structures in the data without any supervision.
- Data is categorized based on their similarities and dissimilarities.

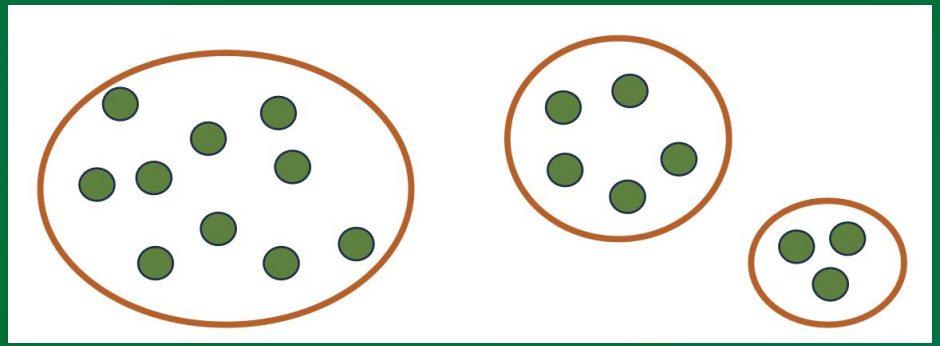




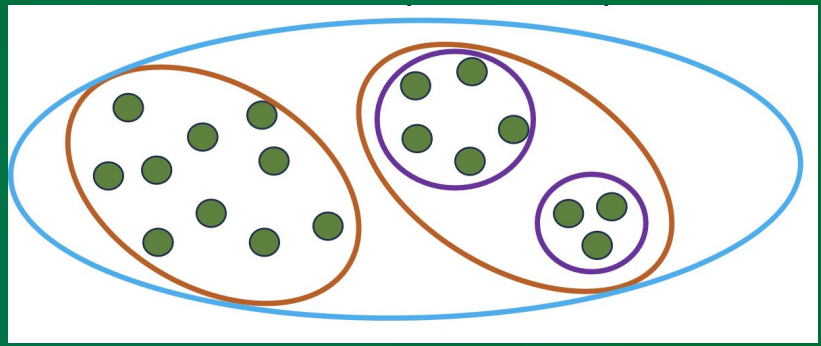
https://www.techtarget.com/searchenterpriseai/definition/clustering-in-machine-learning

Clustering

Partitioning Clustering:
Divides the dataset into a set number of non-overlapping clusters

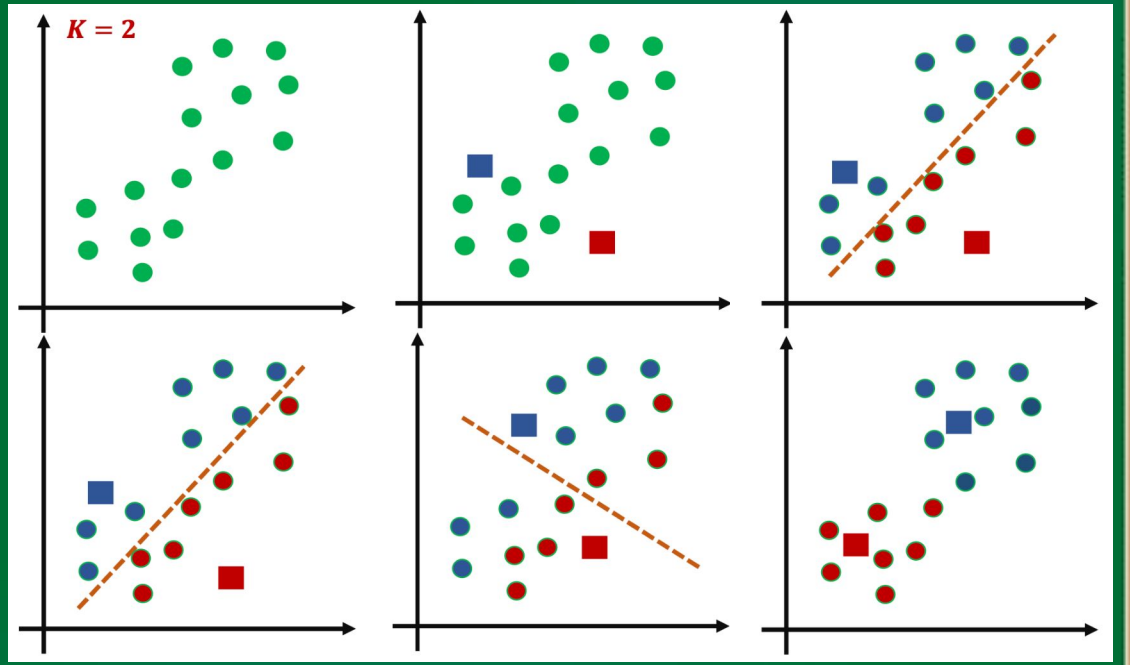


Hierarchical Clustering:
Divides the dataset into a tree-like structure of clusters. Begins with individual data points as separate clusters and merges or splits them iteratively.



K-Means Clustering

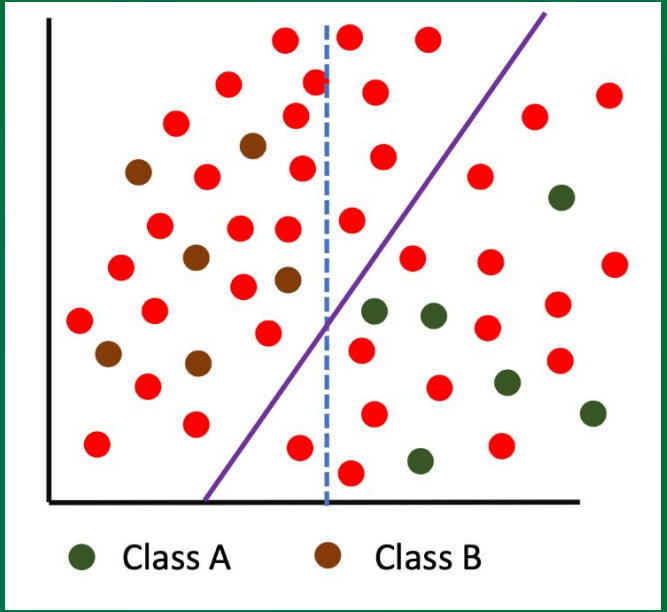
Algorithm organizes data into clusters such that there is high intra-cluster similarity and low inter-cluster similarity.



Semi-supervised Learning

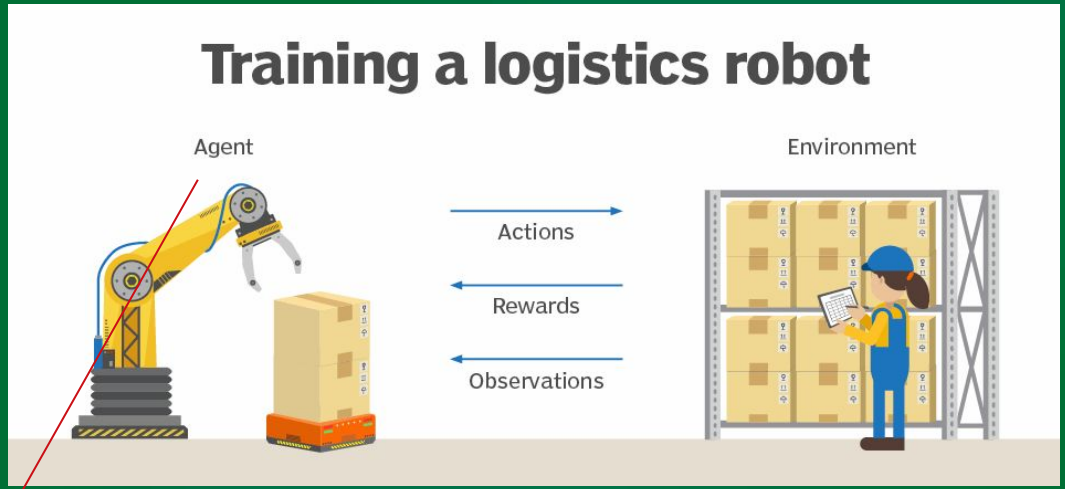
Can we use unlabeled data to augment a small labeled sample to improve learning?

1) Train a model using a small set of known examples.
2) Use the model to predict answers for unknown examples.
3) Add confident predictions to the training set and retrain.
4) Repeat until the model stops improving.



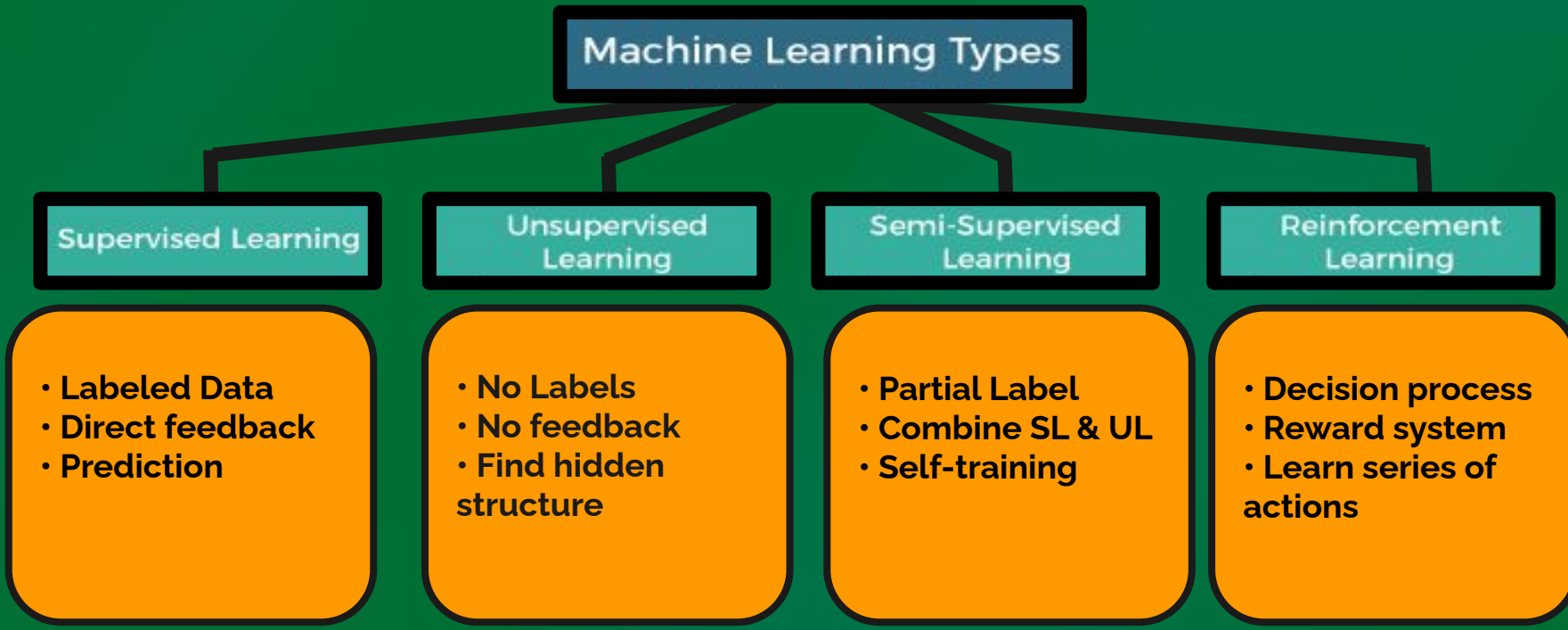
Reinforcement Learning

- Learning by doing
- We give rewards of feedback to the algorithm to learn from and improve future results.

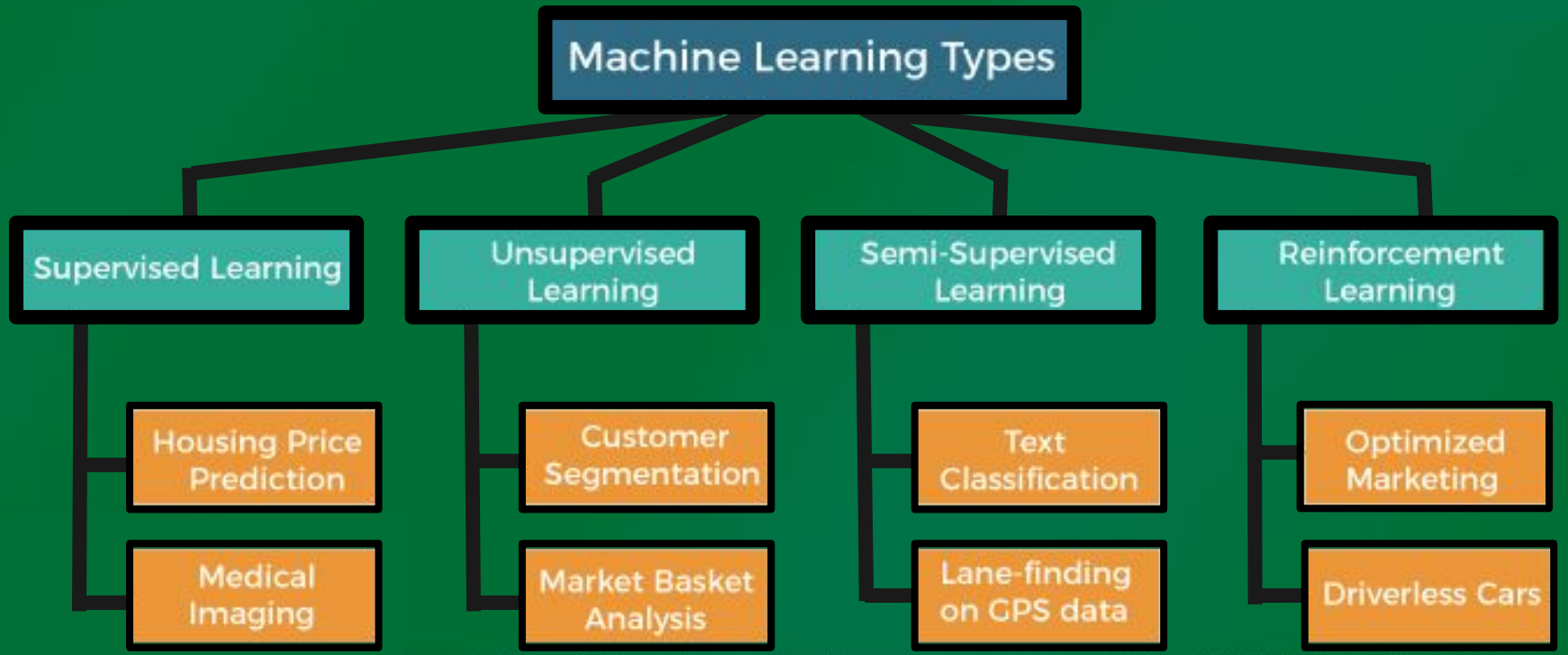


https://www.techtarget.com/searchenterpriseai/definition/reinforcement-learning

Learning Techniques in ML



Learning Techniques in ML



https://www.javatpoint.com/types-of-machine-learning

Supervised Learning

Supervised Learning

Classification

Age	Income	Loan Status
21	20000	Rejected
37	55000	Approved
29	35000	Approved
23	17000	Rejected
34	70000	Approved
47	84000	Rejected
25	30000	Approved

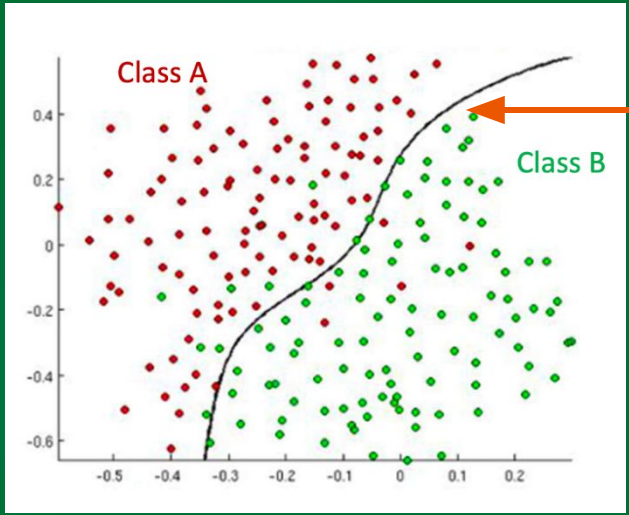
In Classification problem, output is a discrete value/class label

Regression

Age	Income	Loan Amount
21	20000	0
37	55000	150000
29	35000	120000
23	17000	550000
34	70000	250000
47	84000	0
25	30000	90000

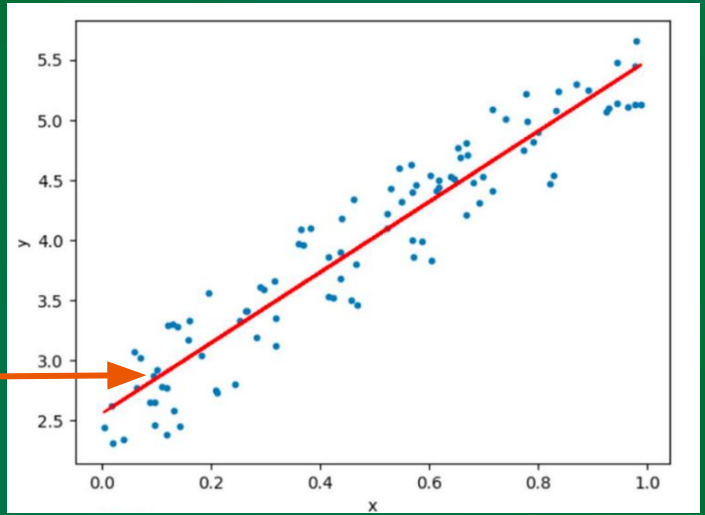
In Regression problem, output is a range of numeric values

Classification v/s Regression



Decision Boundary

Best Fit line



Examples:

- Spam Filtering
- Image classification
- Text classification

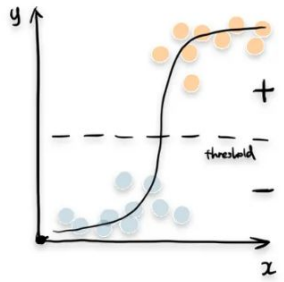
Examples:

- Predicting the price of a house
- Predicting the number of sales
- Predicting the risk of a disease

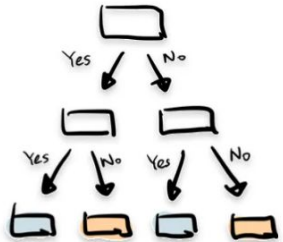
Classification in Supervised Learning

Classification

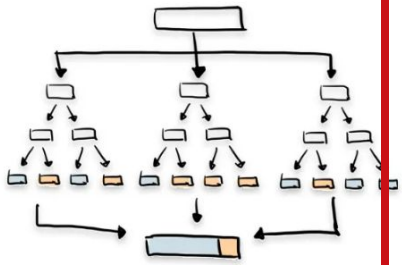
Logistic Regression



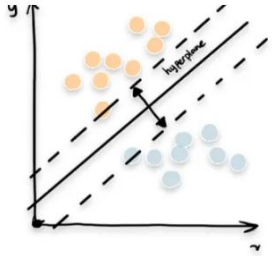
Decision Tree



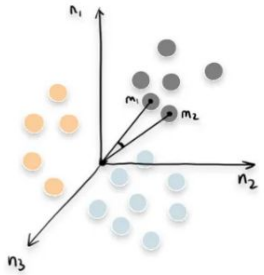
Random Forest



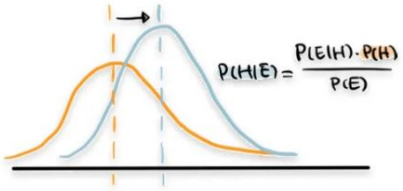
Support Vector Machine



K Nearest Neighbour



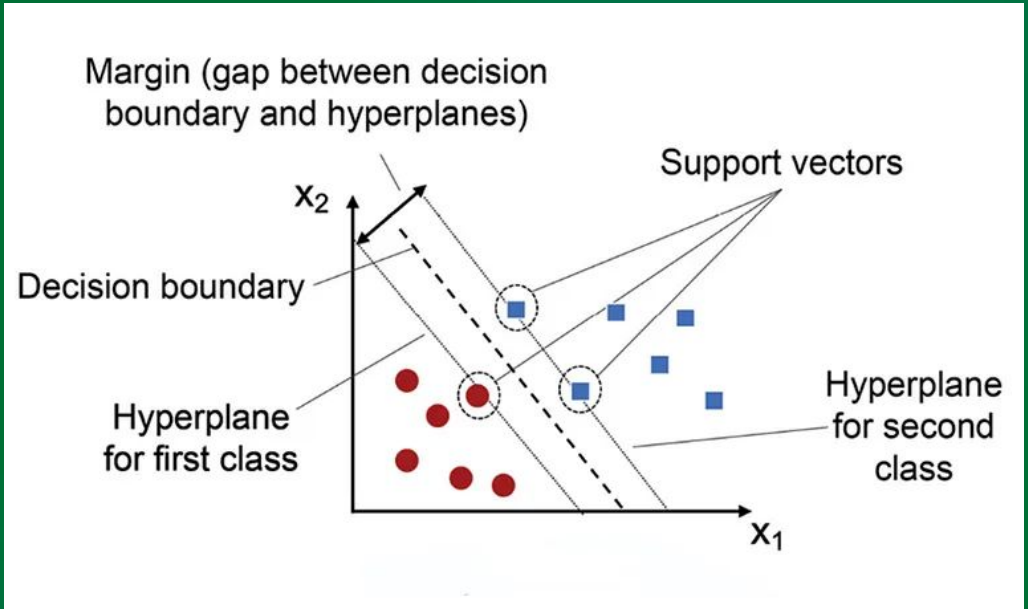
Naive Bayes



https://www.visual-design.net/post/top-machine-learning-algorithms-classification

Support Vector Machines

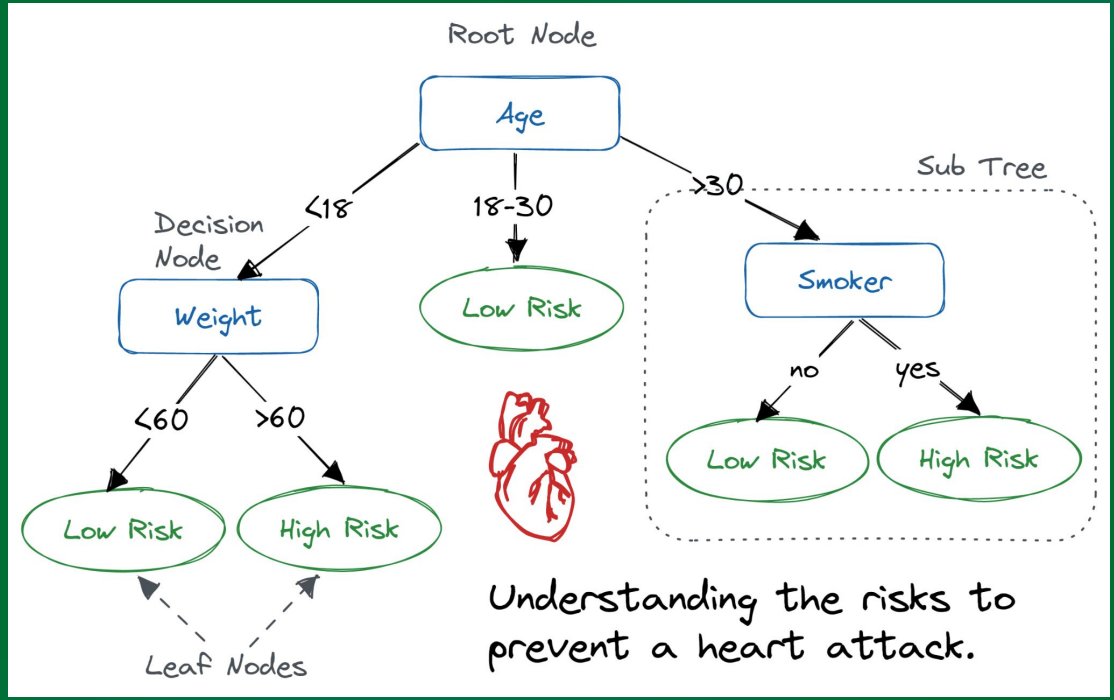
SVM finds the optimal line that separates data points of different classes, by maximizing the margin.



https://vitalflux.com/classification-model-svm-classifier-python-example/

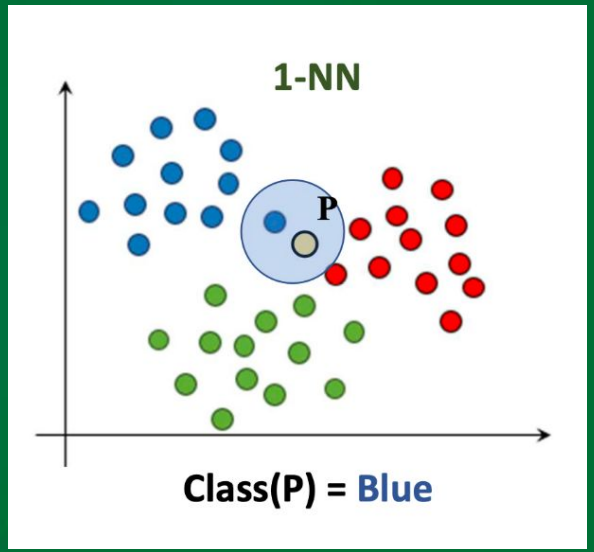
Decision Trees Classifier

- flowchart-like tree structure
- internal node represents a feature
- the branch represents a decision rule
- each leaf node represents the outcome.

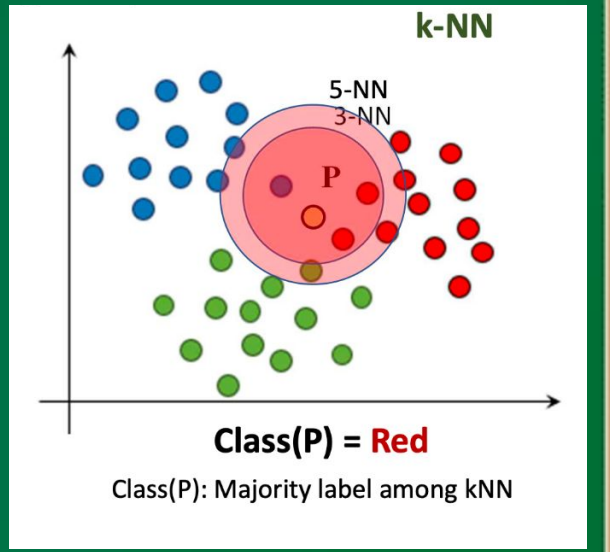


https://www.datacamp.com/tutorial/decision-tree-classification-python

K Nearest Neighbours Algorithm

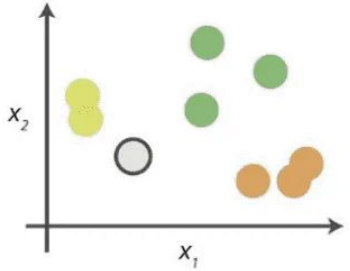


- Given training data and a test point.
- Prediction: Look at The Most Similar Training Example(s)



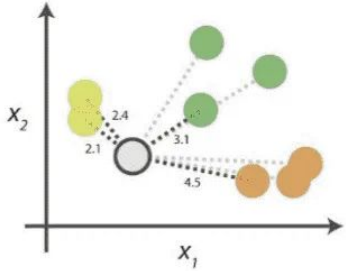
K Nearest Neighbours Algorithm

0. Look at the data



Say you want to classify the grey point into a class. Here, there are three potential classes - lime green, green and orange.

1. Calculate distances

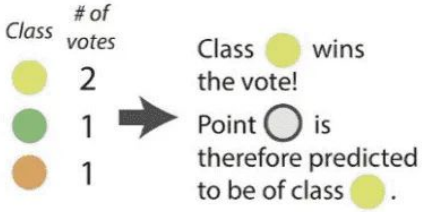


Start by calculating the distances between the grey point and all other points.

2. Find neighbours

Next, find the nearest neighbours by ranking points by increasing distance. The nearest neighbours (NNs) of the grey point are the ones closest in dataspace.

3. Vote on labels



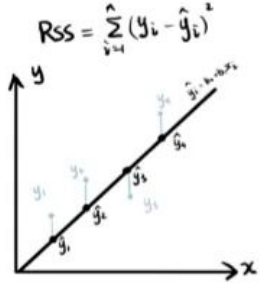
Vote on the predicted class labels based on the classes of the k nearest neighbours. Here, the labels were predicted based on the k=3 nearest neighbours.

https://medium.com/swlh/k-nearest-neighbor-ca2593d7a3c4

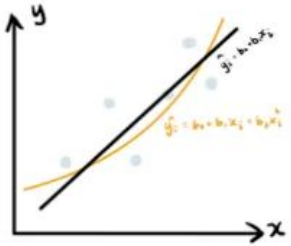
Regression in Supervised Learning

Regression

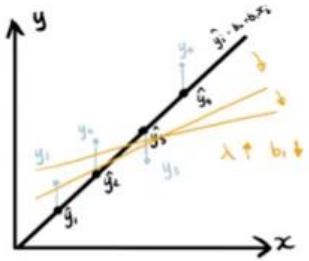
Linear Regression



Polynomial Regression



Regression with Regularization Techniques



Lasso Regression

$$\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \lambda |b_1|$$

Ridge Regression

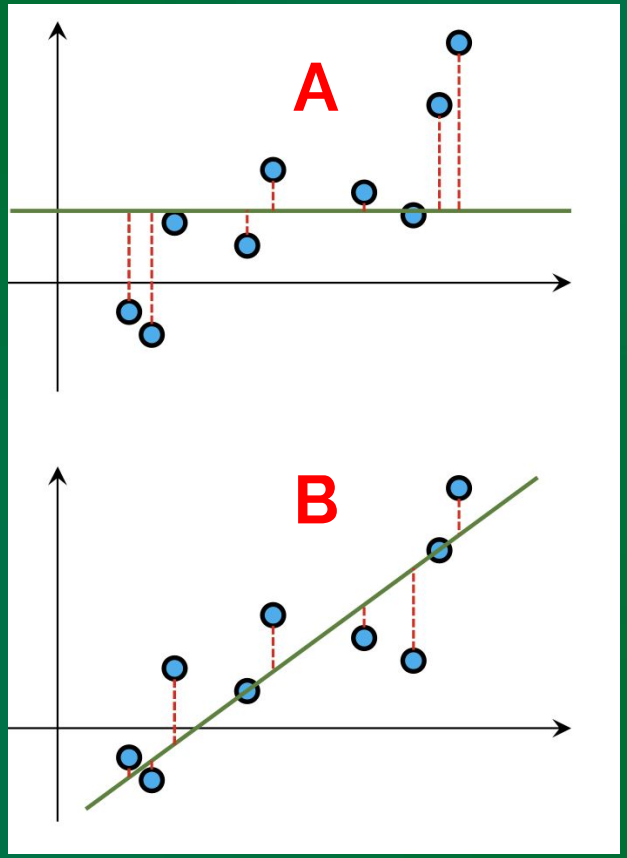
$$\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \lambda (b_1)^2$$

https://www.visual-design.net/post/top-machine-learning-algorithms-classification

Linear Regression

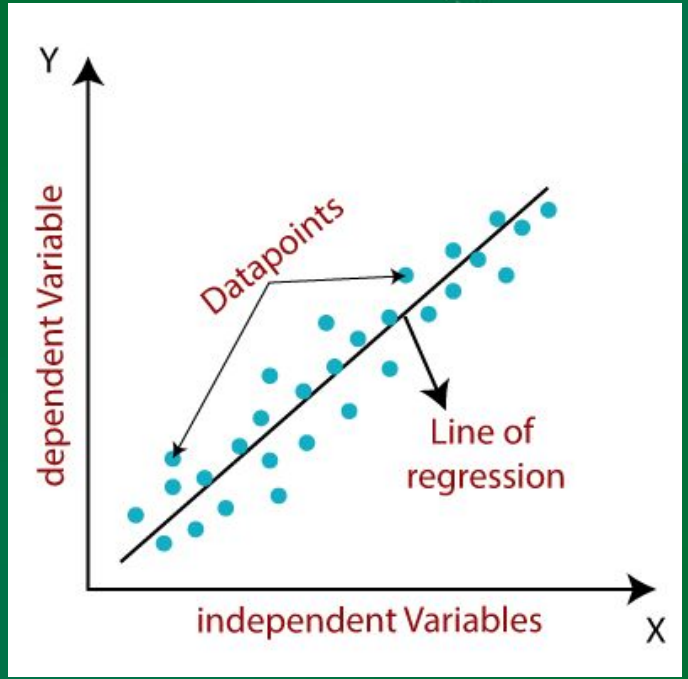
The line for which the error between the predicted values and the observed values is minimum is called the best fit line.

According to you which one of the two is the Best Fit Line - A or B?



Linear Regression

Linear Regression is a predictive model used for finding the linear relationship between a dependent variable and one or more independent variables.



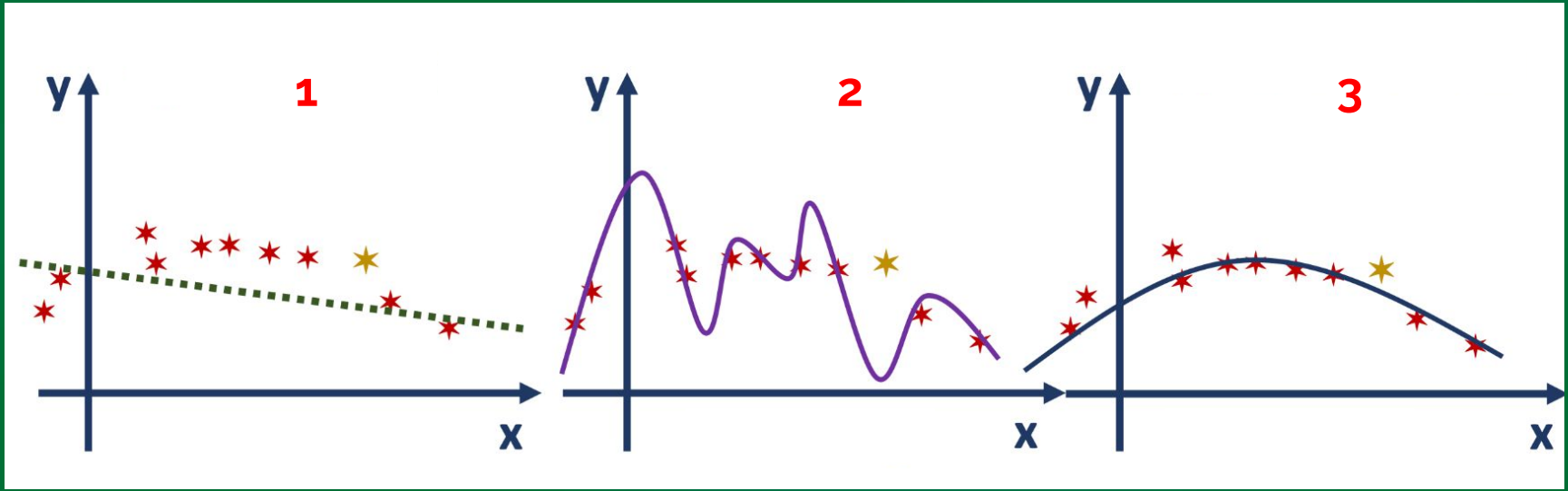
$$Y = mX + b$$

- Y = dependent variable,
- X = independent variable
- m = slope (or Gradient, determines change in Y, per unit change in X),
- b = Y-intercept

Evaluating Machine Learning Models

Underfitting & Overfitting

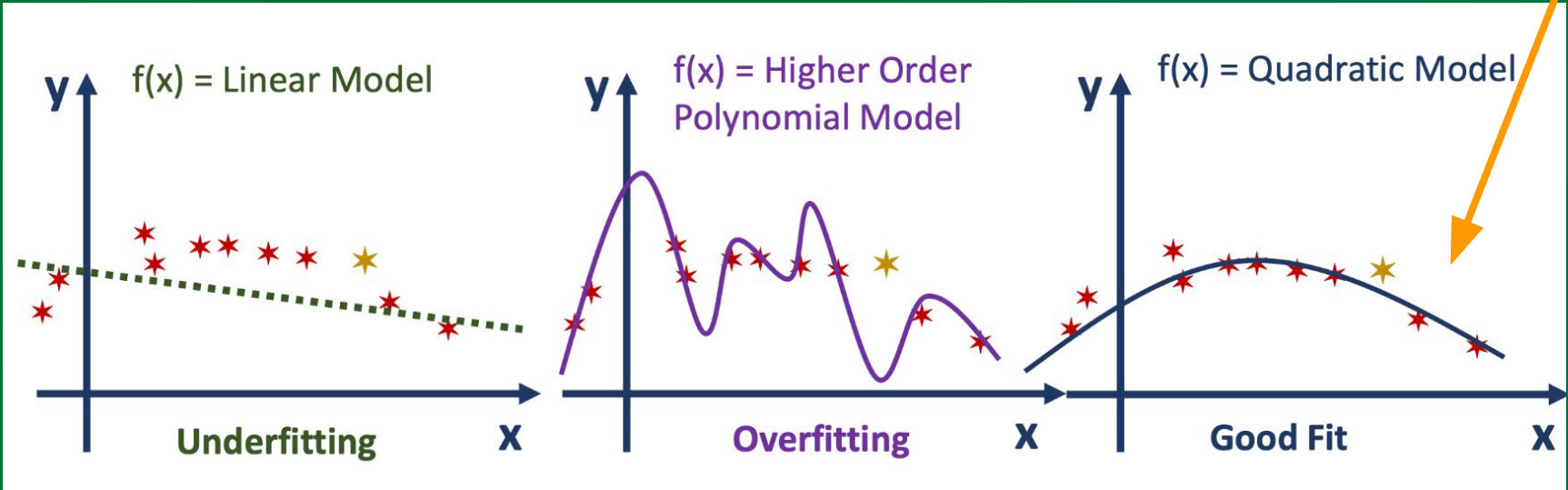
Which of the models will perform best on the unseen test data?



Underfitting & Overfitting

Model 3 will perform the best!

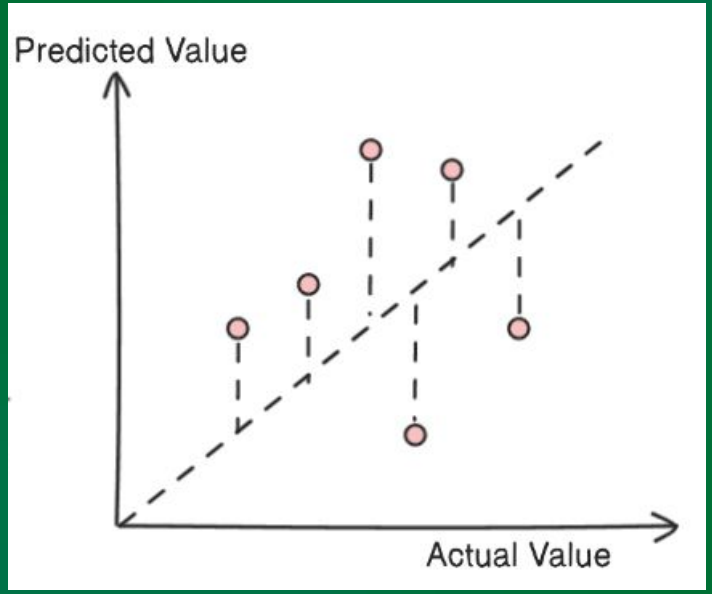
Good Generalization!



Evaluating ML Models - Regression

MSE - Mean Squared Error is used to find the squared difference between actual and predicted value.

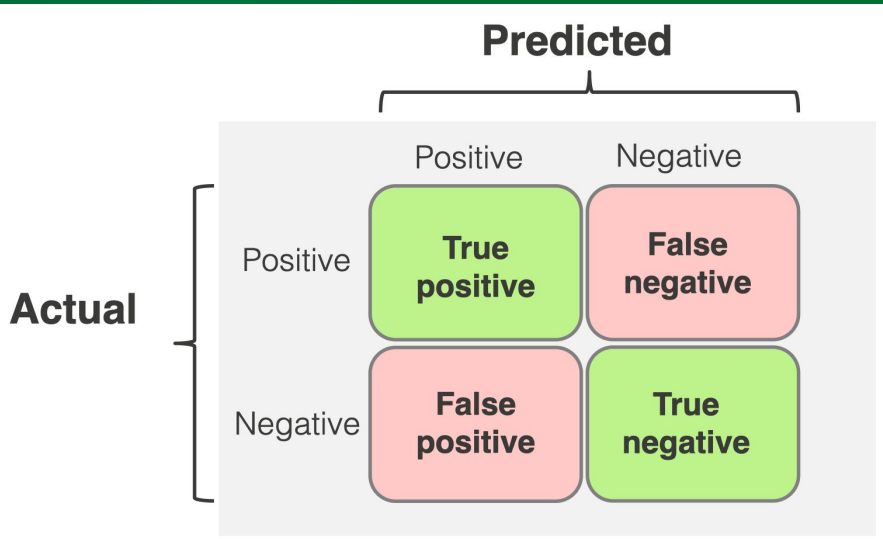
$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

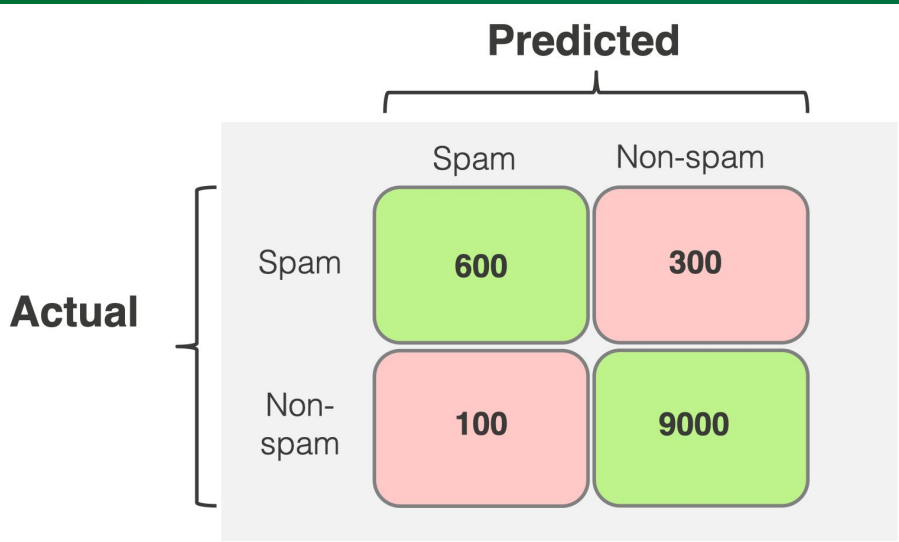


https://www.analyticsvidhya.com/blog/2021/05/know-the-best-evaluation-metrics-for-your-regression-model/

Evaluating ML Models - Classification

Confusion Matrix:



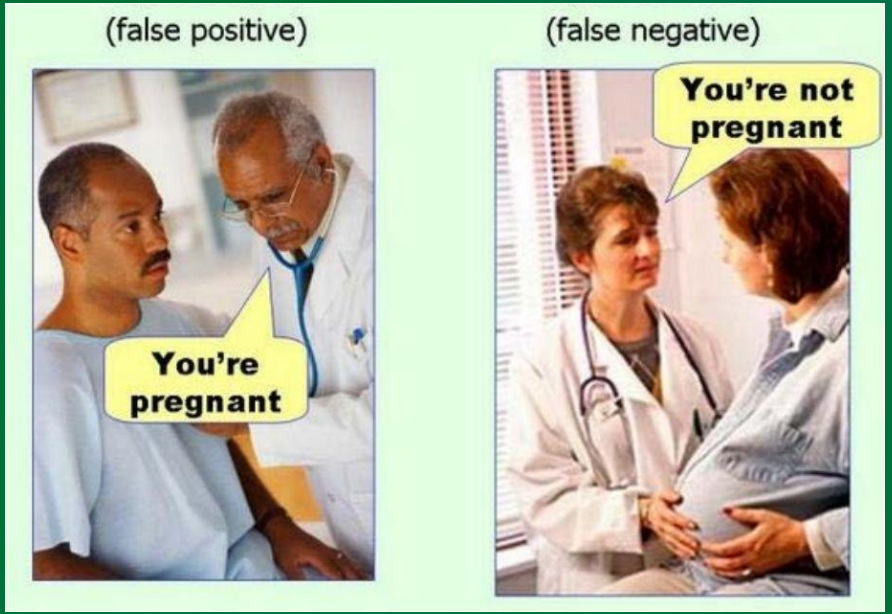


https://www.evidentlyai.com/classification-metrics/confusion-matrix

Evaluating ML Models - Classification

False Positives & False Negatives

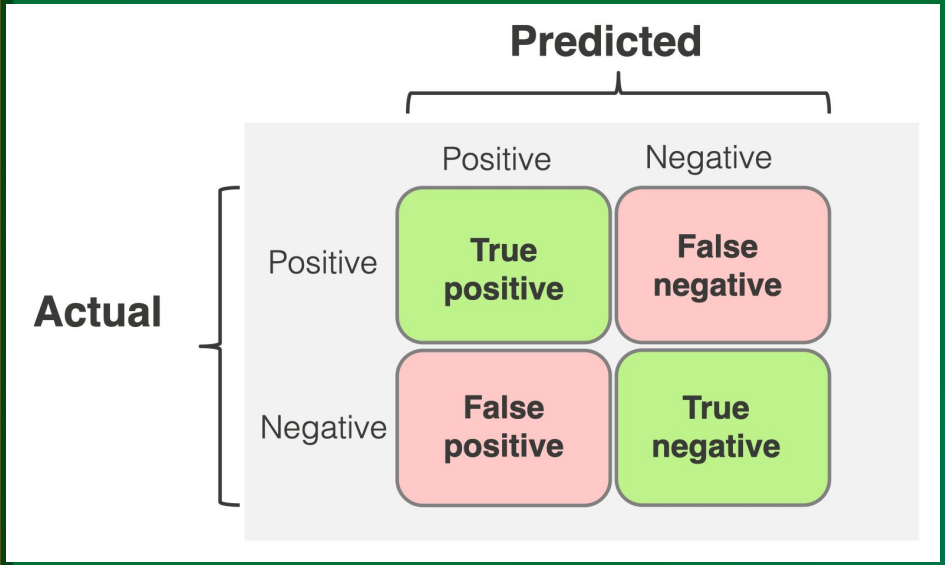
Is the patient pregnant or not?



https://cs229.stanford.edu/section/evaluation_metrics_spring2020.pdf

Evaluating ML Models - Classification

Confusion Matrix: - Question for You!

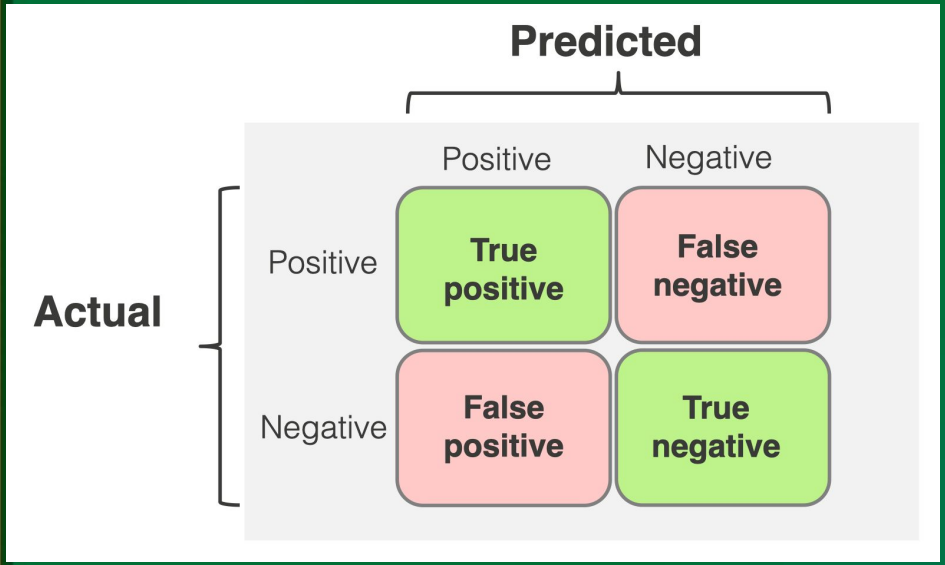


Which rate should be low?

1) Screening For A Terminal Disease.
2) Automatic bombing on detecting a target from a drone
3) Giving Access To A Secure Installation

Evaluating ML Models - Classification

Confusion Matrix: - Answers

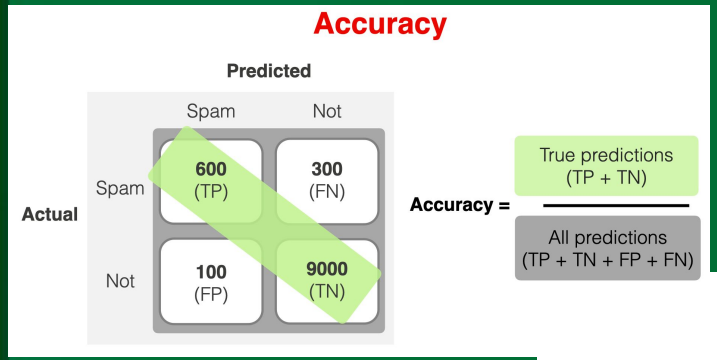


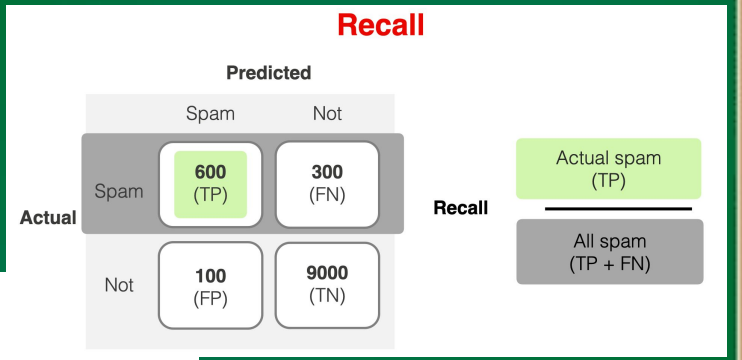
Which rate should be low?

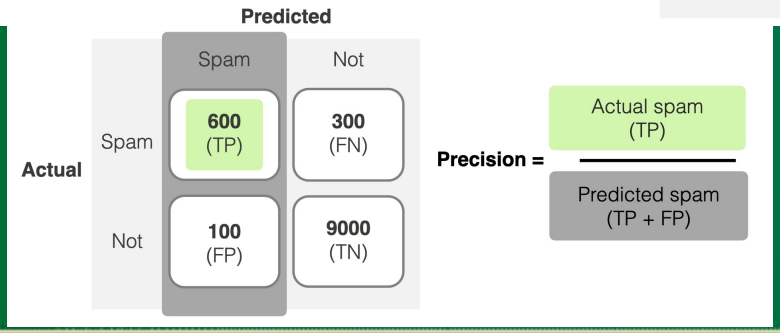
1) Screening For A Terminal Disease. - FALSE NEGATIVE SHOULD BE LOW
2) Automatic bombing on detecting a target from a drone - FALSE POSITIVE LOW/NULL
3) Giving Access To A Secure Installation - FALSE POSITIVE LOW

Evaluating ML Models - Classification

Confusion Matrix: - Accuracy, Precision & Recall







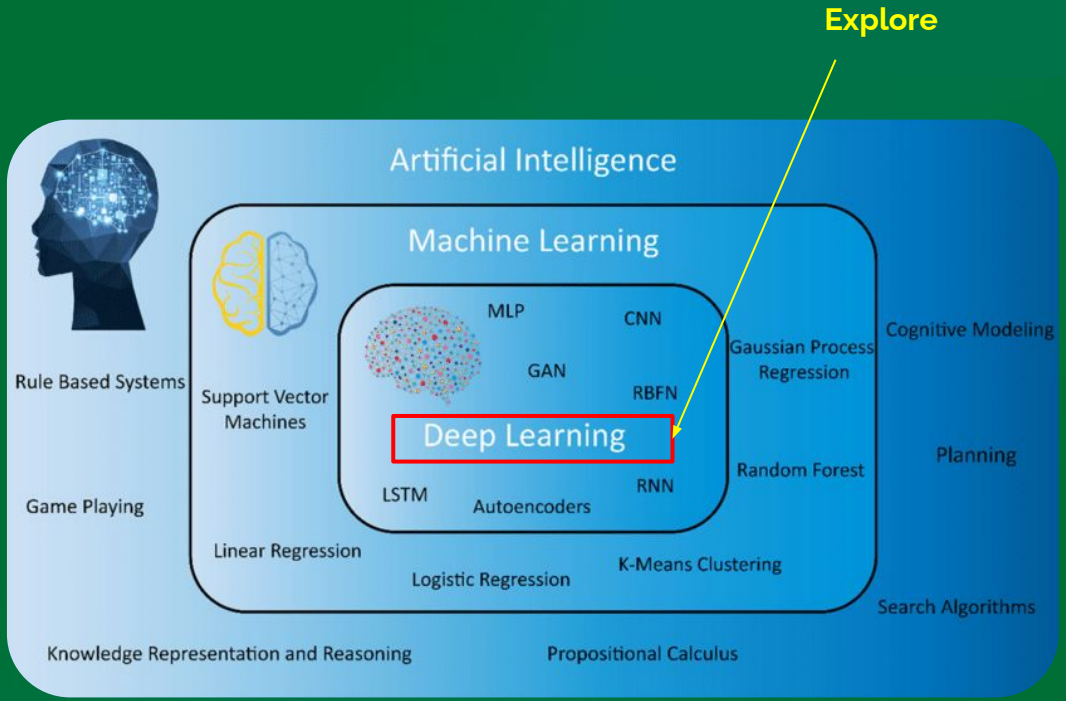
https://www.evidentlyai.com/classification-metrics/confusion-matrix

These were the ABCs of Machine Learning!

What after Machine Learning?

Machine Learning works well for Structured Data. (data stored in tables)

Deep Learning is used for unstructured data (rich media, text, social media activity, video files, audio files, surveillance imagery)



https://www.researchgate.net/figure/Domains-of-AI-ML-DL-and-widely-used-algorithms_fig1_361501987

How you can join us & connect with us

DATA SCIENCE STUDENT CLUB-JNTUH

WhatsApp group

@DATASCIENCESTUDENTCLUBJNTUH

Make a mini-project with us now!

Scan for Notebook 1:

Scan for Notebook 2: