Machine Learning
Lecture 2

Dr. Muhammad Kamran Malik

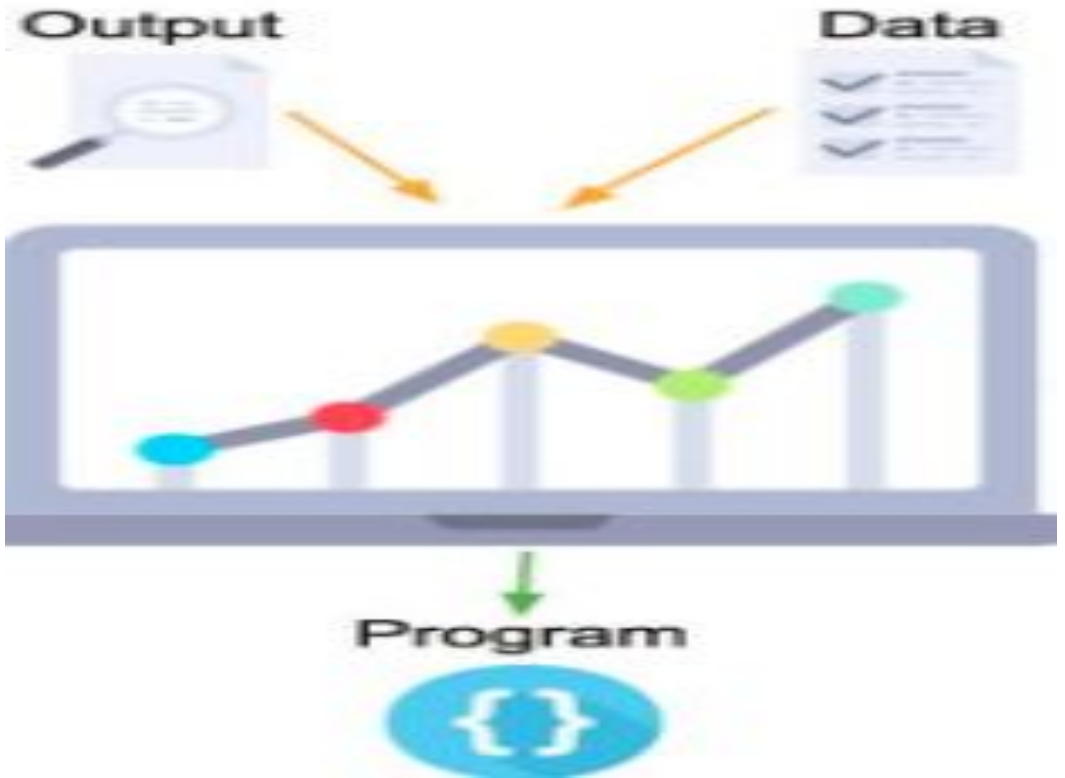
Supervised Machine Learning

What is Supervised Machine Learning

- Supervised learning is the types of machine learning in which machines are trained using well "labelled" training data, and on basis of that data, machines predict the output.
- The labelled data means some input data is already tagged with the correct output.
- In supervised learning, the training data provided to the machines work as the supervisor that teaches the machines to predict the output correctly.
- It applies the same concept as a student learns in the supervision of the teacher.

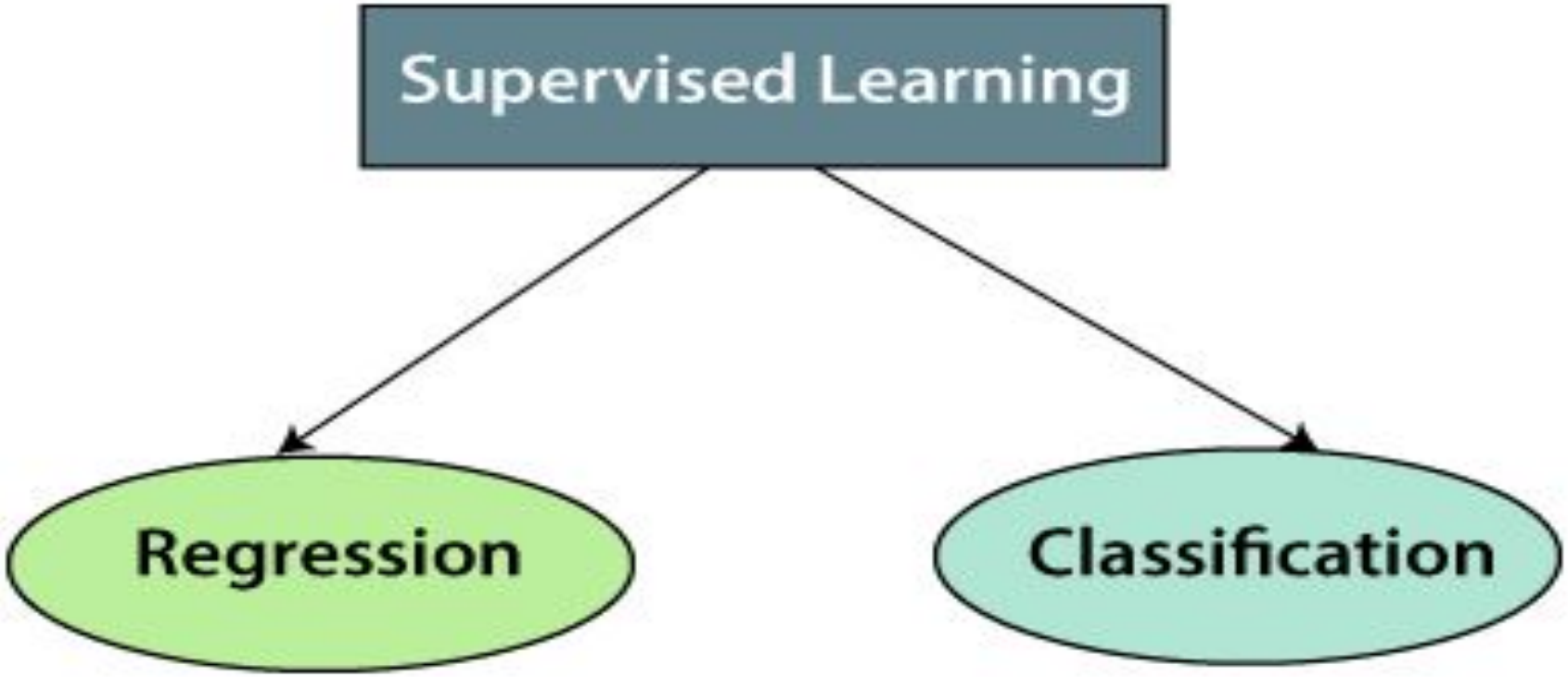
What is Supervised Machine Learning

- Supervised learning is a process of providing input data as well as correct output data to the machine learning model.
- The aim of a supervised learning algorithm is to **find a mapping function to map the input variable(x) with the output variable(y)**.



Types of Supervised Learning

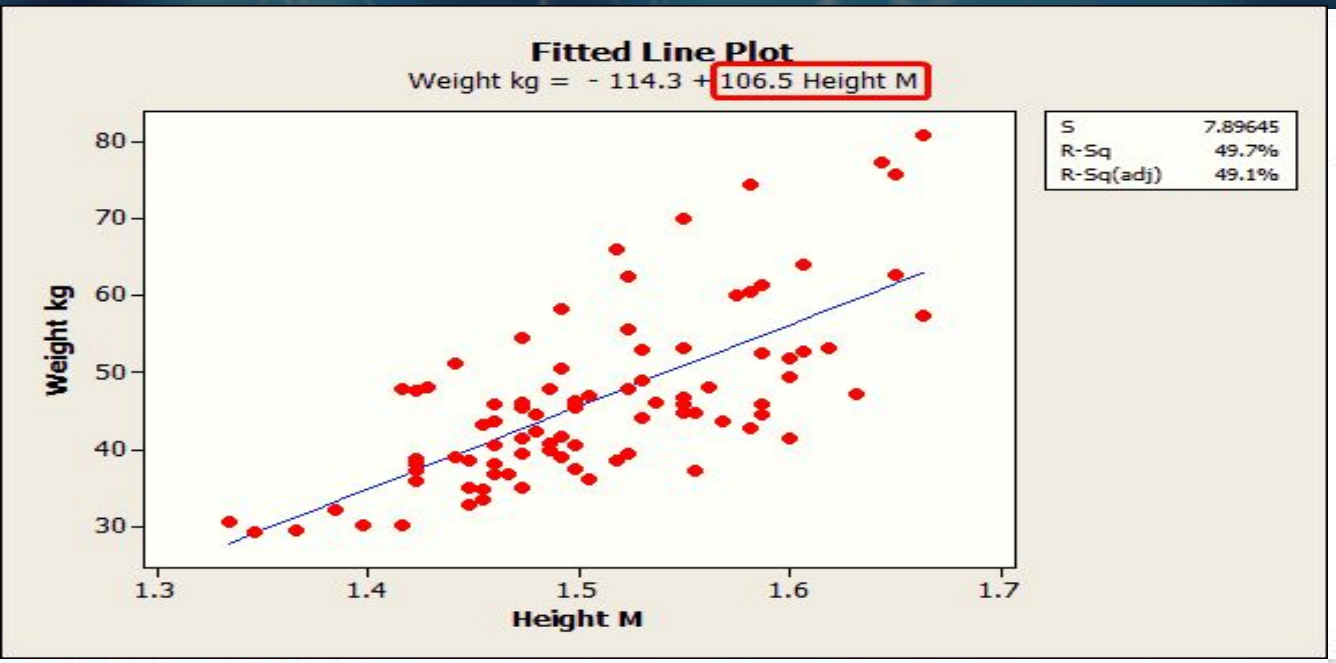
Types of Supervised Learning



Types of Supervised Learning (Regression)

- Regression algorithms are algorithms are used when the output variable is continuous variables, such as Weather forecasting, Market Trends, etc.
- Below are some popular Regression algorithms which come under supervised learning:
  - Linear Regression
  - Regression Trees
  - Non-Linear Regression
  - Bayesian Linear Regression
  - Polynomial Regression, etc.

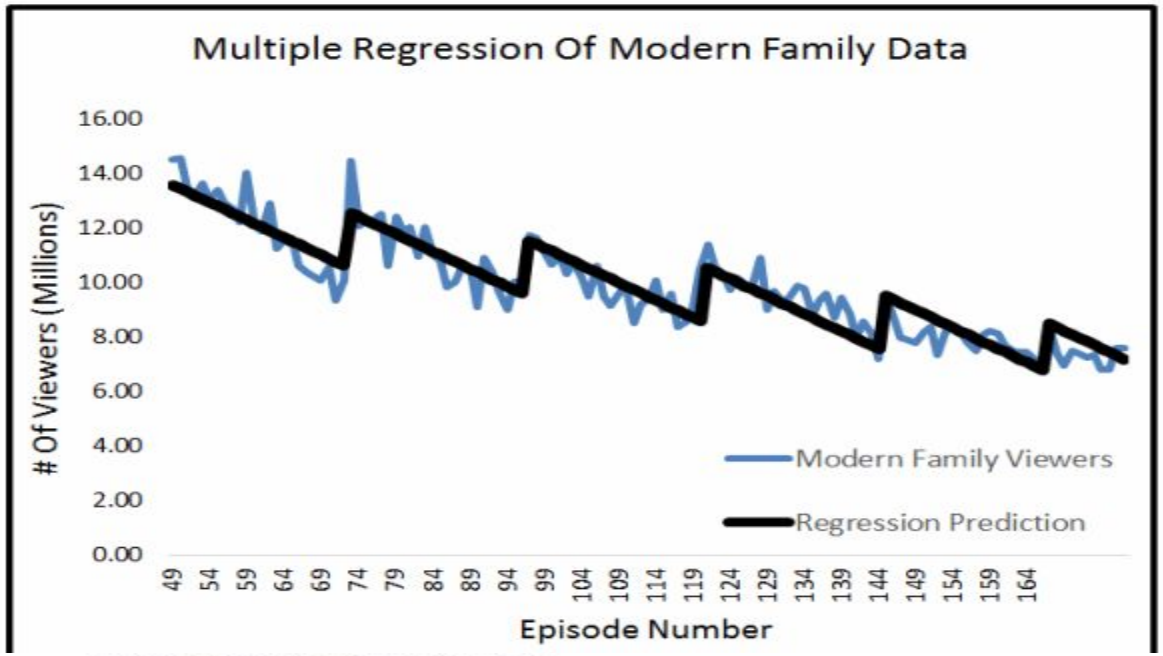
Regression Example



Year	Sales (Million Euro)	Advertising (Million Euro)
1	651	23
2	762	26
3	856	30
4	1,063	34
5	1,190	43
6	1,298	48
7	1,421	52
8	1,440	57
9	1,518	58

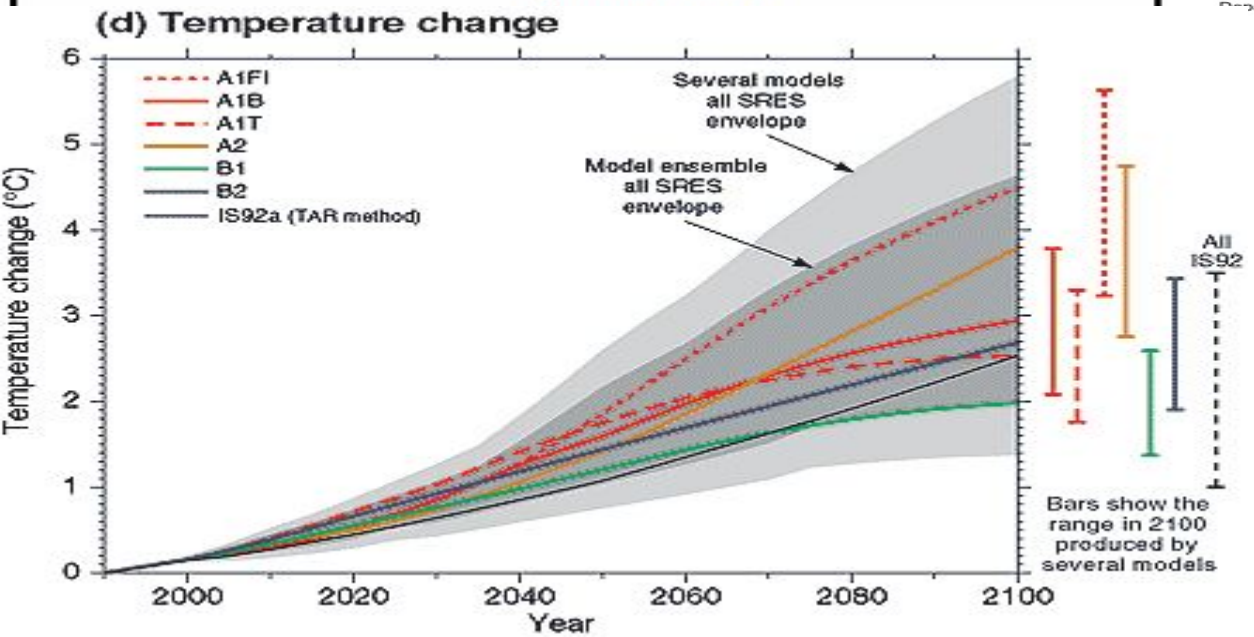
	A	B	C	D
1	Quantity Sold	Price	Advertising	
2	8500	$2	$2,800	
3	4700	$5	$200	
4	5800	$3	$400	
5	7400	$2	$500	
6	6200	$5	$3,200	
7	7300	$3	$1,800	
8	5600	$4	$900	
9				

Regression Example



Simple Linear Regression Example

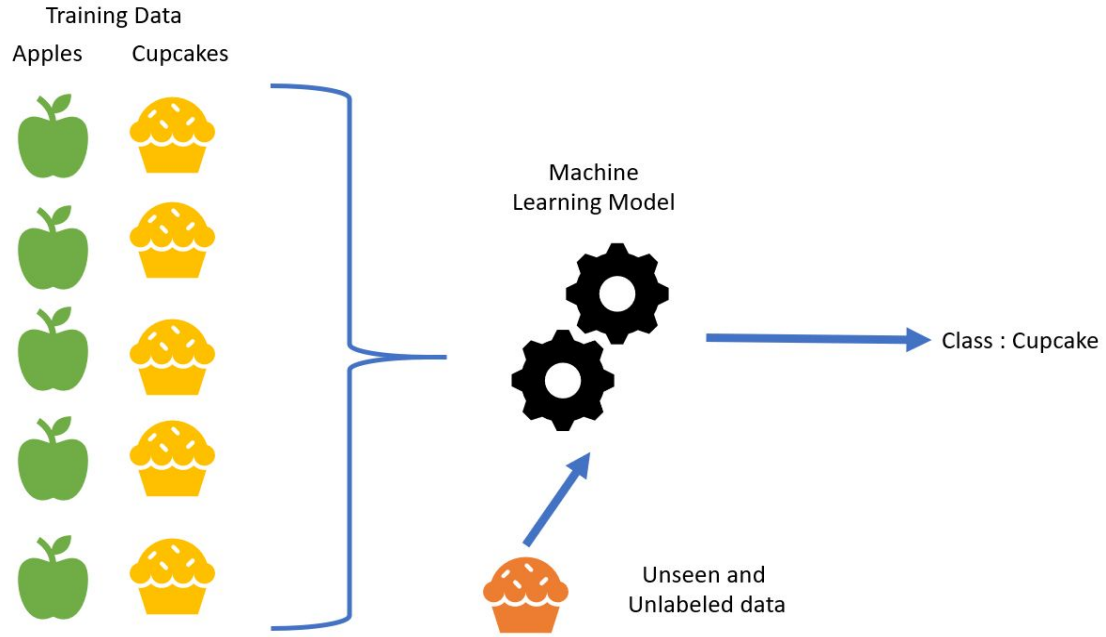
- A real estate agent wishes to examine the relationship between the selling price of a home and its size (measured in square feet)
- A random sample of 10 houses is selected
  - Dependent variable (Y) = house price in $1000s
  - Independent variable (X) = square feet

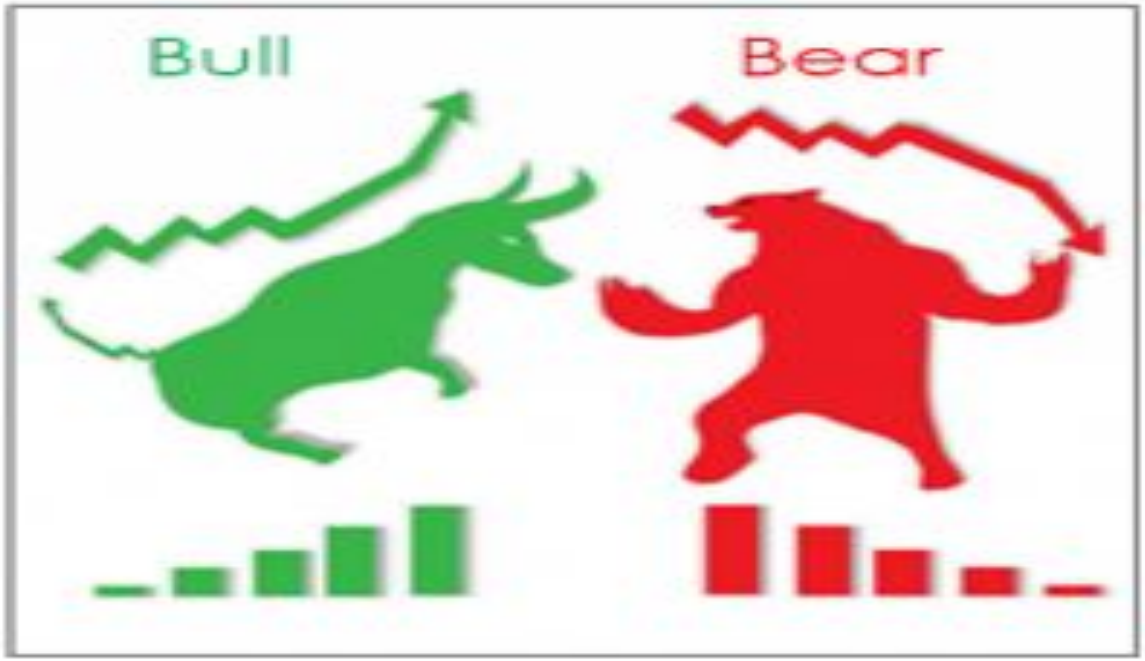


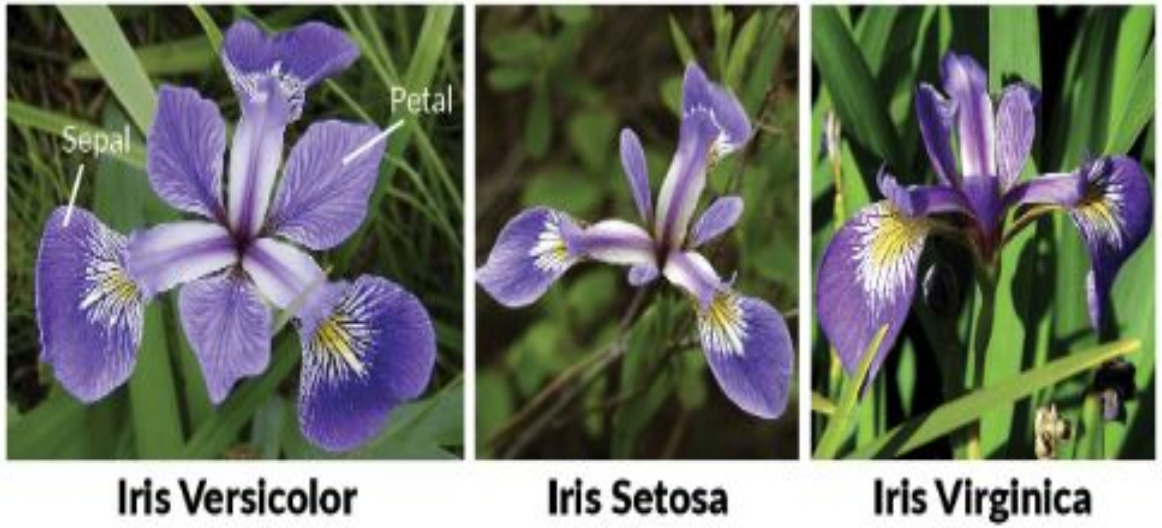
Types of Supervised Learning (Classification)

- Classification algorithms are used when the output variable is categorical, which means there are two classes such as Yes-No, Male-Female, True-false, etc.
  - Naïve Bayes
  - Random Forest
  - Decision Trees
  - Logistic Regression
  - Support vector Machines etc

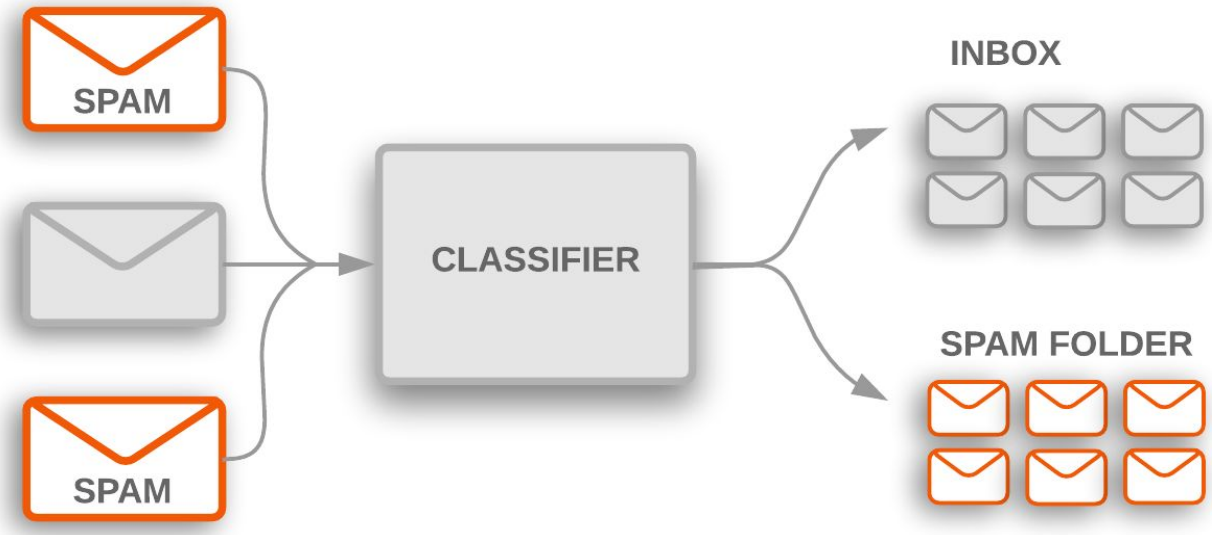
Classification Example

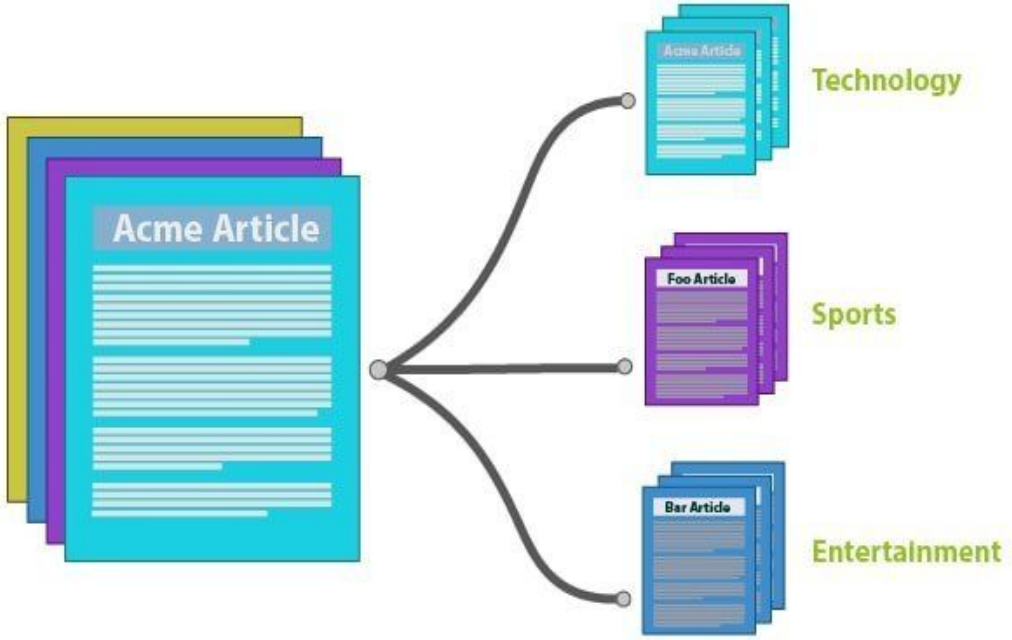


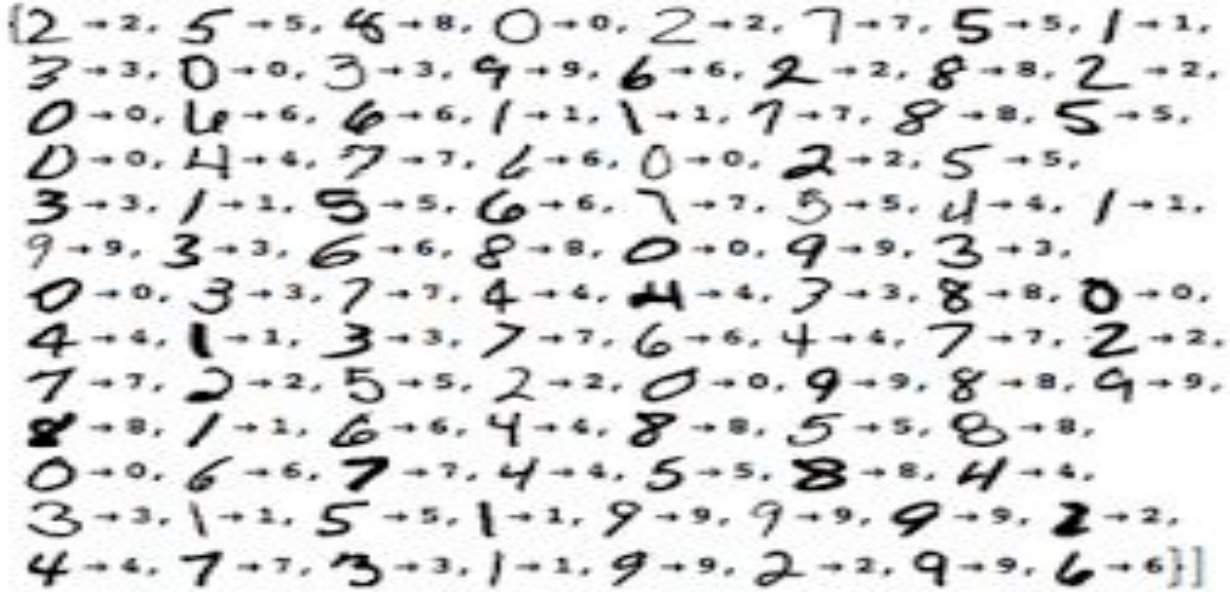




Classification Example







Q & A

How Supervised Learning Works

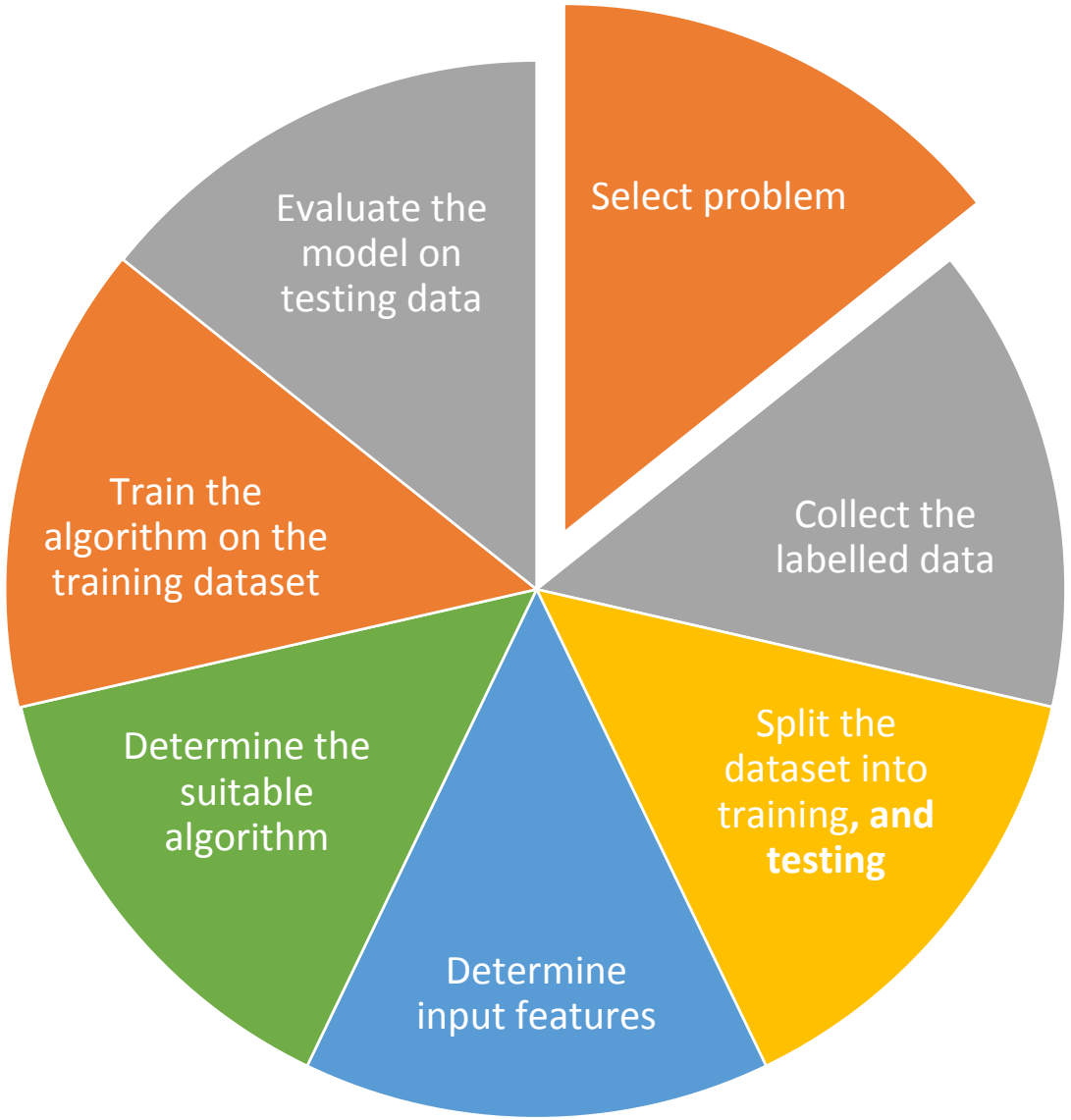
How Supervised Learning Works?

- First determine the type of problem
- Collect the labelled data.
- Split the dataset into **training (80%)** and **testing (20%).**
- Determine the input features of the training dataset, which should have enough knowledge so that the model can accurately predict the output.
- Determine the suitable algorithm for the model, such as support vector machine, decision tree, etc.
- Execute the algorithm on the training dataset.
- Evaluate the accuracy of the model by providing the test set. If the model predicts the correct output, which means our model is accurate.

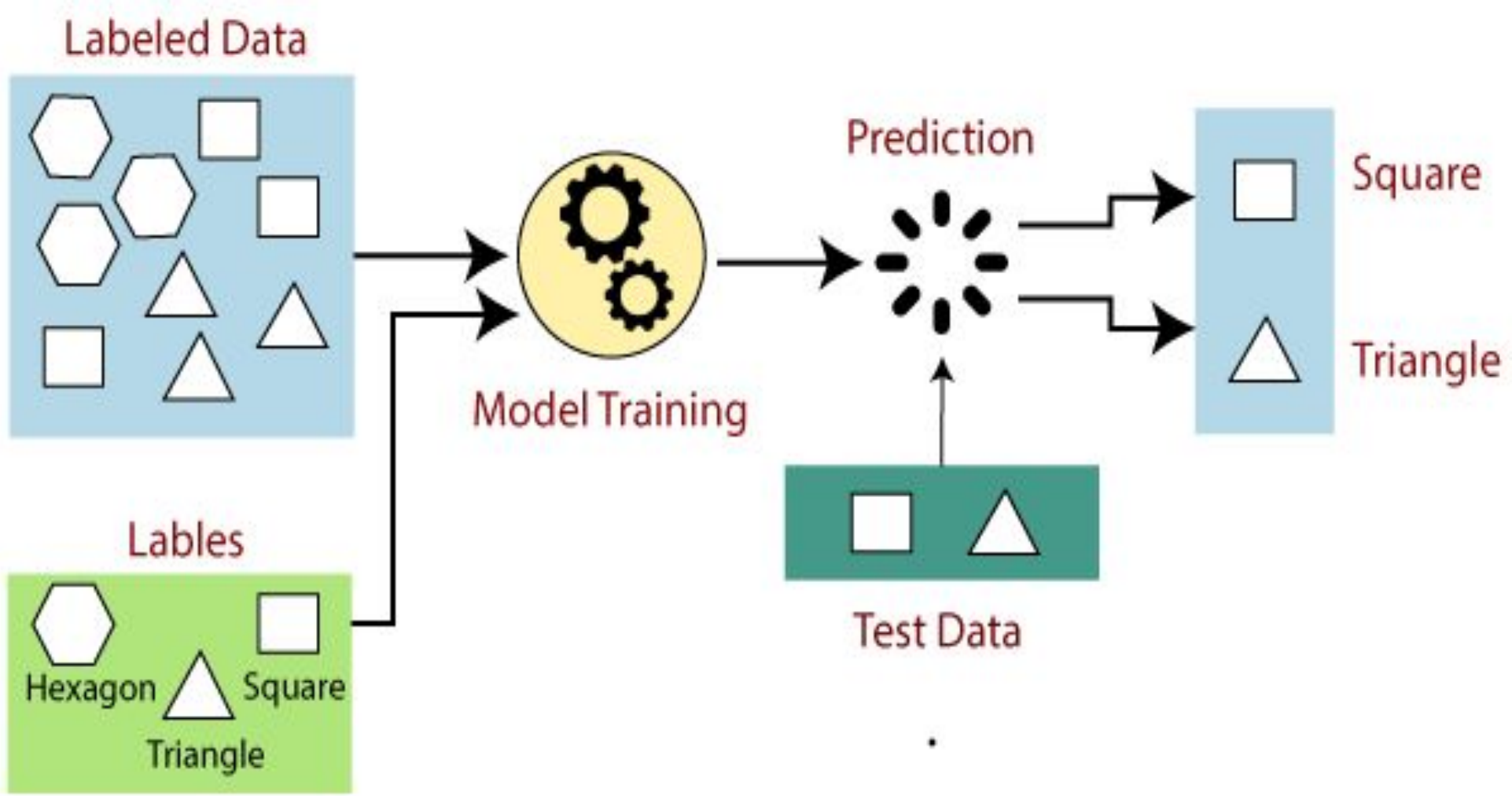
How Supervised Learning Works?

- First Determine the type of problem
- Collect the labelled data.
- Split the dataset into **training, testing, and validation dataset**.
- Determine the input features of the training dataset, which should have enough knowledge so that the model can accurately predict the output.
- Determine the suitable algorithm for the model, such as support vector machine, decision tree, etc.
- Execute the algorithm on the training dataset. Sometimes we need validation sets as the control parameters, which are the subset of training datasets.
- Evaluate the accuracy of the model by providing the test set. If the model predicts the correct output, which means our model is accurate.

How Supervised Learning Works?



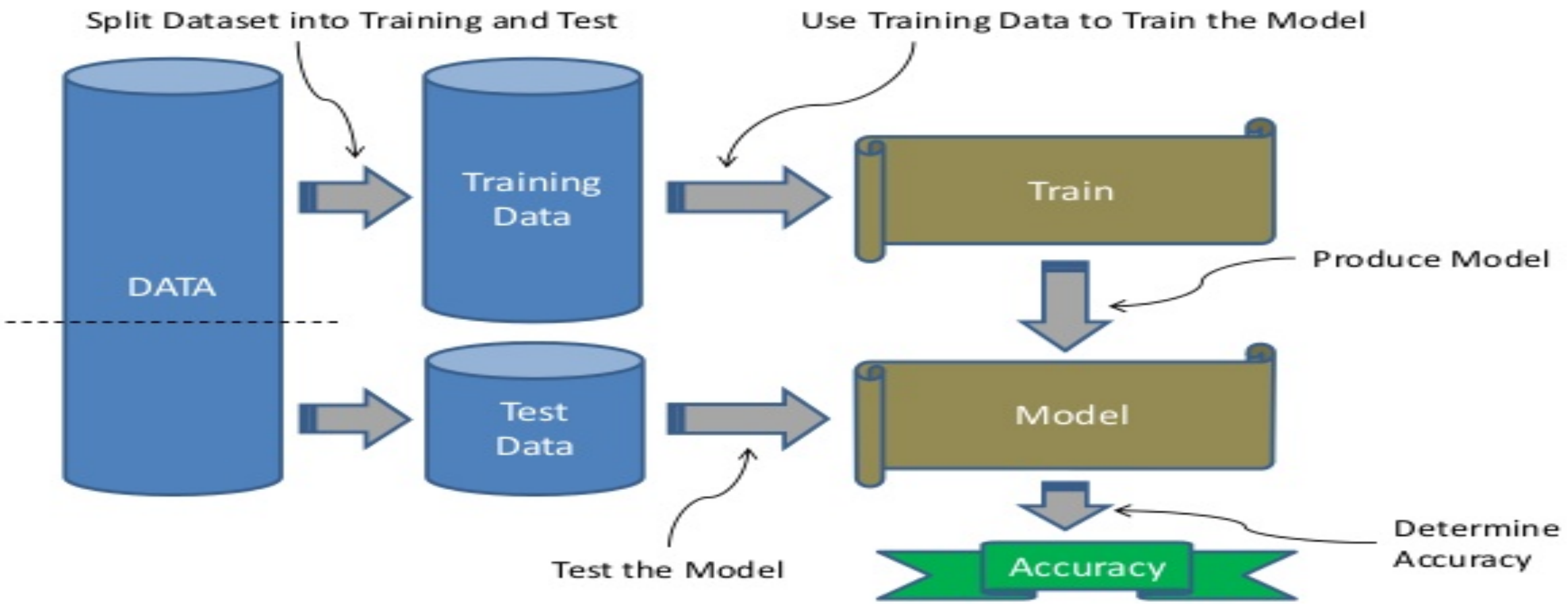
How Supervised Learning Works?



How Supervised Learning Works?

It's About Training

Machine Learning is about using data to train a model



How Supervised Learning Works?

Q & A

Model Evaluation Measure

Model Evaluation Measures

- **True positives (TP):** data points predicted/labeled as positive that are actually positive
- **False positives (FP):** data points predicted/labeled as positive that are actually negative
- **True negatives (TN):** data points predicted/labeled as negative that are actually negative
- **False negatives (TN):** data points predicted/labeled as negative that are actually positive

Model Evaluation Measures

- **Accuracy**: Closeness of a measured value to a standard or known value **((TP + TN)/(TP+TN+FN+FP))**
- **Recall:** ability of a classification model to identify all relevant instances **(TP / TP + FN)**
- **Precision:** ability of a classification model to return only relevant instances **(TP / TP + FP)**
- **F1 score:** single metric that combines recall and precision using the harmonic mean **(2 (Precision * Recall)/(Precision + Recall))**

Model Evaluation Measures (Classification)

Images	Actual Label	Predicted Label
Image1	Happy	Sad
Image2	Sad	Sad
Image3	Sad	Happy
Image4	Happy	Happy
Image5	Happy	Happy
Image6	Sad	Happy
Image7	Sad	Happy
Image8	Happy	Happy
Image9	Happy	Happy
Image10	Sad	Sad

		Actual	
		Positive	Negative
Predicted	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

Accuracy = (TP + TN)/(TP+TN+FN+FP)
Recall = (TP / TP + FN)
Precision = (TP / TP + FP)

Model Evaluation Measures (Classification)

Images	Actual Label	Model 1	Model 2
Patient1	Not Corona	Not Corona	Not Corona
Patient2	Not Corona	Not Corona	Not Corona
Patient3	Not Corona	Not Corona	Not Corona
Patient4	Not Corona	Not Corona	Not Corona
Patient5	Not Corona	Not Corona	Not Corona
Patient6	Not Corona	Not Corona	Corona
Patient7	Not Corona	Not Corona	Corona
Patient8	Not Corona	Not Corona	Corona
Patient9	Corona	Not Corona	Corona
Patient10	Corona	Not Corona	Corona

		Actual	
		Positive	Negative
Predicted	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

Accuracy = (TP + TN)/(TP+TN+FN+FP)
Recall = (TP / TP + FN)
Precision = (TP / TP + FP)

Model Evaluation Measures (Regression)

GPA	Salary (thousands) y	Model 1 y’	Model 2 y’
3	40	35	39
3.2	45	45	44
3.5	50	55	48
3.8	80	75	82

$$Mean\ Square\ Error\ (MSE) = \frac{1}{N}\sum_{i=1}^{n}(y^{(i)} - y'^{(i)})^2$$

Q & A

Basic Notations

	A	B	C	D	E
1	Sepal Length	Sepal Width	Petal Length	Petal Width	Class
2	5.1	3.5	1.4	0.2	Iris-setosa
3	4.9	3	1.4	0.2	Iris-setosa
4	4.7	3.2	1.3	0.2	Iris-setosa
5	4.6	3.1	1.5	0.2	Iris-setosa
6	5	3.6	1.4	0.2	Iris-setosa
7	5.4	3.9	1.7	0.4	Iris-setosa
8	4.6	3.4	1.4	0.3	Iris-setosa
9	5	3.4	1.5	0.2	Iris-setosa
10	4.4	2.9	1.4	0.2	Iris-setosa
11	4.9	3.1	1.5	0.1	Iris-setosa
12	5.4	3.7	1.5	0.2	Iris-setosa
13	4.8	3.4	1.6	0.2	Iris-setosa
14	4.8	3	1.4	0.1	Iris-setosa
15	4.3	3	1.1	0.1	Iris-setosa
16	5.8	4	1.2	0.2	Iris-setosa
17	5.7	4.4	1.5	0.4	Iris-setosa
18	5.4	3.9	1.3	0.4	Iris-setosa
19	5.1	3.5	1.4	0.3	Iris-setosa
20	5.7	3.8	1.7	0.3	Iris-setosa
21	5.1	3.8	1.5	0.3	Iris-setosa
22	5.4	3.4	1.7	0.2	Iris-setosa
23	5.1	3.7	1.5	0.4	Iris-setosa
24	4.6	3.6	1	0.2	Iris-setosa
25	5.1	3.3	1.7	0.5	Iris-setosa

- **D** = labeled training data (X,y)
- **m** = no of examples in the dataset
- **X** = set of training examples
- $\boldsymbol{n_x}$= no of features
- $\boldsymbol{x^i}$= i^{th} example point
- $\boldsymbol{x_j^i}$= j^{th} feature value of i^{th} example
- $\boldsymbol{y^i}$= label of i^{th} example

Advantages and Disadvantages of Supervised Machine Learning

Advantages of Supervised Learning

- With the help of supervised learning, the model can predict the output on the basis of prior experiences.
- In supervised learning, we can have an exact idea about the classes of objects.
- Supervised learning model helps us to solve various real-world problems such as **fraud detection, spam filtering**, etc.

Disadvantages of Supervised Learning

- Requires a lot of effort in the development of labeled data.
- Supervised learning cannot predict the correct output if the test data is different from the training dataset.
- In supervised learning, we need enough knowledge about the classes of object.

Q & A