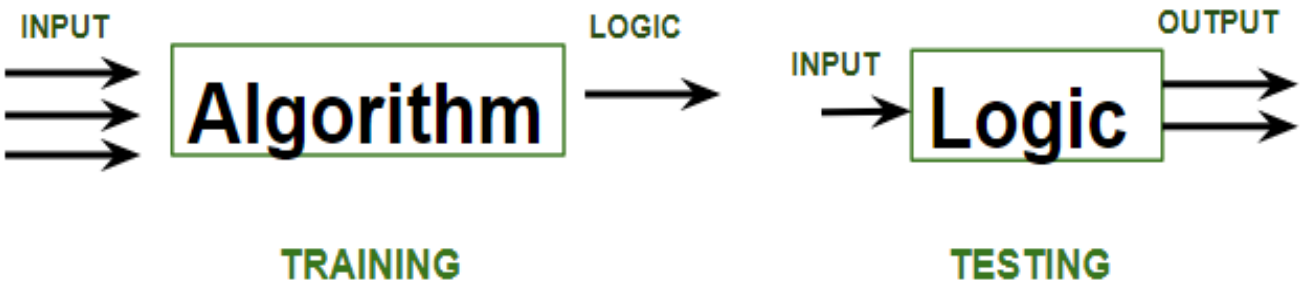
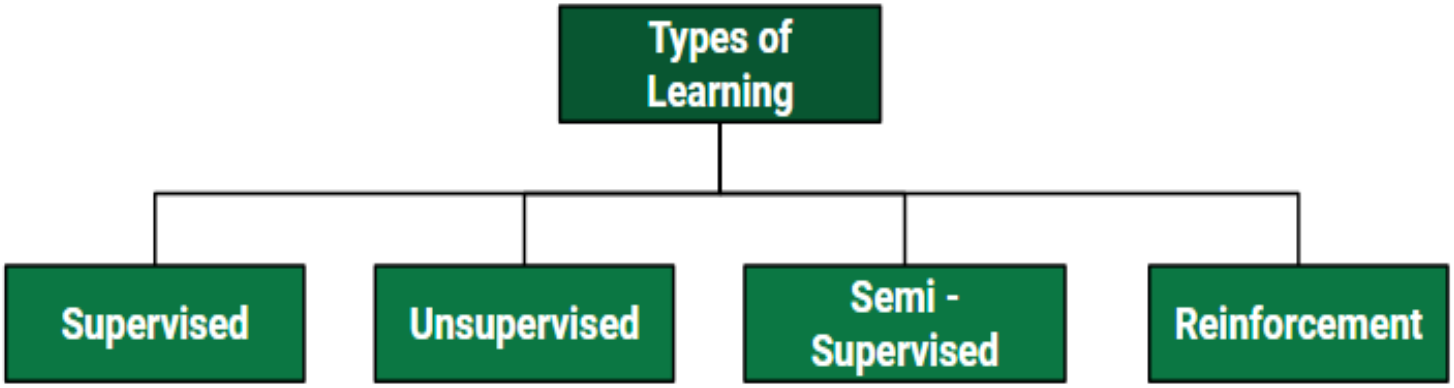
Machine Learning

Types of Learning



What is Learning for a machine?

A machine is said to be learning from **past Experiences** (data feed in) with respect to some class of **Tasks**, if it's **Performance** in a given Task improves with the Experience.



Supervised Learning :

Supervised learning is when the model is getting trained on a labeled dataset.

Labeled dataset is one which have both input and output parameters.

In this type of learning both training and validation datasets are labeled as shown in the figures below.

User ID	Gender	Age	Salary	Purchased
15624510	Male	19	19000	0
15810944	Male	35	20000	1
15668575	Female	26	43000	0
15603246	Female	27	57000	0
15804002	Male	19	76000	1
15728773	Male	27	58000	1
15598044	Female	27	84000	0
15694829	Female	32	150000	1
15600575	Male	25	33000	1
15727311	Female	35	65000	0
15570769	Female	26	80000	1
15606274	Female	26	52000	0
15746139	Male	20	86000	1
15704987	Male	32	18000	0
15628972	Male	18	82000	0
15697686	Male	29	80000	0
15733883	Male	47	25000	1

Figure A: CLASSIFICATION

Temperature	Pressure	Relative Humidity	Wind Direction	Wind Speed
10.69261758	986.882019	54.19337313	195.7150879	3.278597116
13.59184184	987.8729248	48.0648859	189.2951202	2.909167767
17.70494885	988.1119385	39.11965597	192.9273834	2.973036289
20.95430404	987.8500366	30.66273218	202.0752869	2.965289593
22.9278274	987.2833862	26.06723423	210.6589203	2.798230886
24.04233986	986.2907104	23.46918024	221.1188507	2.627005816
24.41475295	985.2338867	22.25082295	233.7911987	2.448749781
23.93361956	984.8914795	22.35178837	244.3504333	2.454271793
22.68800023	984.8461304	23.7538641	253.0864716	2.418341875
20.56425726	984.8380737	27.07867944	264.5071106	2.318677425
17.76400389	985.4262085	33.54900114	280.7827454	2.343950987
11.25680746	988.9386597	53.74139903	68.15406036	1.650191426
14.37810685	989.6819458	40.70884681	72.62069702	1.553469896
18.45114201	990.2960205	30.85038484	71.70604706	1.005017161
22.54895853	989.9562988	22.81738811	44.66042709	0.264133632
24.23155922	988.796875	19.74790765	318.3214111	0.329656571

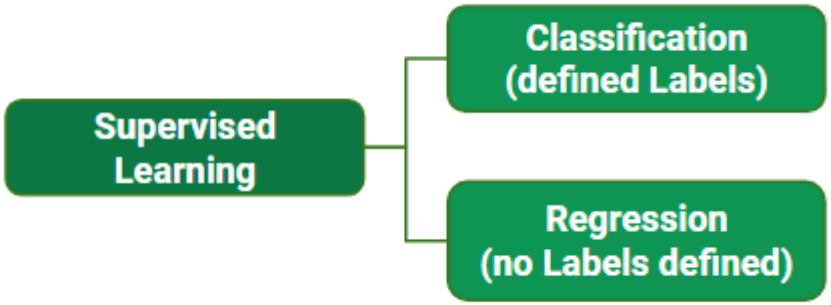
Figure B: REGRESSION

- **Figure A:** It is a dataset of a shopping store which is useful in predicting whether a customer will purchase a particular product under consideration or not based on his/ her gender, age and salary.
  Input : Gender, Age, Salary
  Ouput : Purchased i.e. 0 or 1 ; 1 means yes the customer will purchase and 0 means that customer won't purchase it.
- **Figure B:** It is a Meteorological dataset which serves the purpose of predicting wind speed based on different parameters.
  Input : Dew Point, Temperature, Pressure, Relative Humidity, Wind Direction
  Output : Wind Speed

- T**raining the system:**
  While training the model, data is usually split in the ratio of 80:20
- 80% as training data and rest as testing data.
- In training data, feed input as well as output for 80% data.
- The model learns from training data only.
- Use different machine learning algorithms to build a model.
- By learning, it means that the model will build some logic of its own.

- Once the model is ready then it is good to be tested.
- At the time of testing, input is fed from remaining 20% data which the model has never seen before
- The model will predict some value
- Compare it with actual output and calculate the accuracy.

Types of Supervised Learning:



Classification : It is a Supervised Learning task where output is having defined labels (discrete value).

It can be either binary or multi class classification. In **binary** classification, model predicts either 0 or 1 ; yes or no but in case of **multi class** classification, model predicts more than one class.

Example: Gmail classifies mails in more than one classes like social, promotions, updates, forum.

Regression

- It is a Supervised Learning task where output is having continuous value.

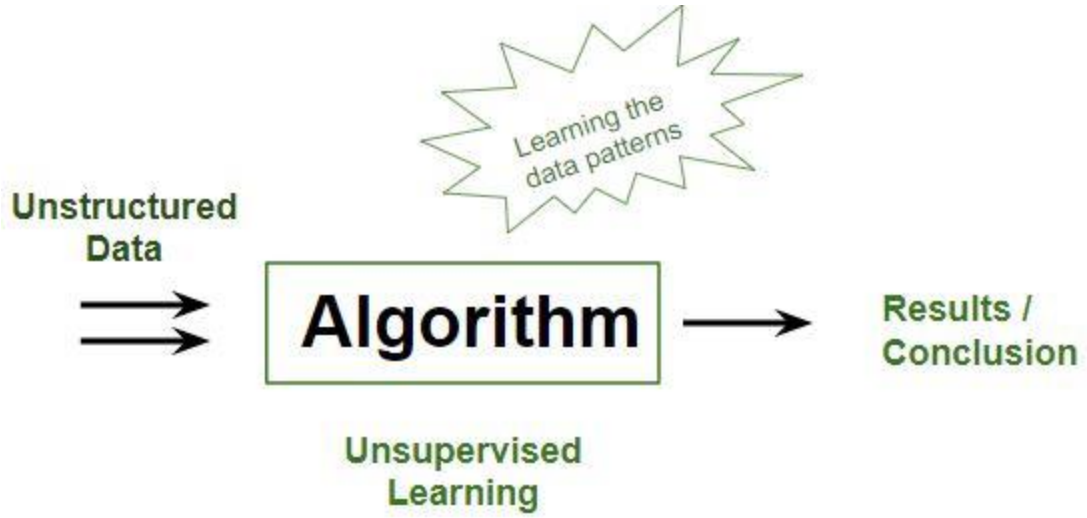
- Example

Output – Wind Speed is not having any discrete value but is continuous in the particular range.

The goal here is to predict a value as much closer to actual output value as our model can and then evaluation is done by calculating error value.

The smaller the error the greater the accuracy of our regression model.

Unsupervised Learning :



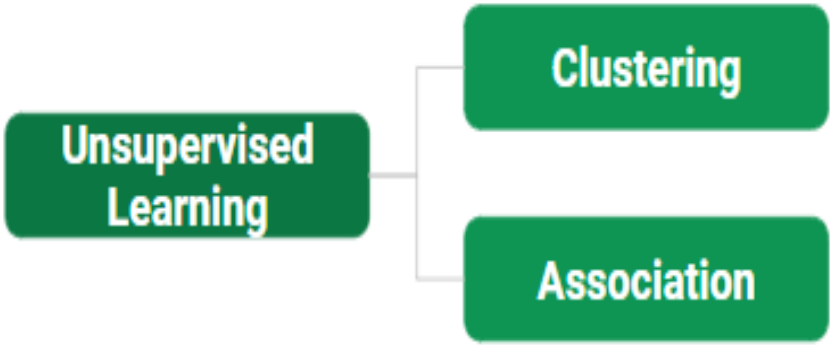
It’s a type of learning where we don’t give target to our model while training i.e. training model has only input parameter values.

CustomerID	Genre	Age	Annual Income (k$)	Spending Score (1-100)
1	Male	19	15	39
2	Male	21	15	81
3	Female	20	16	6
4	Female	23	16	77
5	Female	31	17	40
6	Female	22	17	76
7	Female	35	18	6
8	Female	23	18	94
9	Male	64	19	3
10	Female	30	19	72
11	Male	67	19	14
12	Female	35	19	99
13	Female	58	20	15
14	Female	24	20	77
15	Male	37	20	13
16	Male	22	20	79
17	Female	35	21	35

Figure A

Data-set in Figure A is mall data that contains information of its clients that subscribe to them. Once subscribed they are provided a membership card and so the mall has complete information about customer and his/her every purchase. Now using this data and unsupervised learning techniques, mall can easily group clients based on the parameters we are feeding in.

- Training data we are feeding is –
- **Unstructured data**: May contain noisy (meaningless) data, missing values or unknown data
- **Unlabeled data** : Data only contains value for input parameters, there is no targeted value (output). It is easy to collect as compared to labeled one in Supervised approach.



Semi-supervised Learning:

working lies between Supervised and Unsupervised techniques

These techniques are used when dealing with a data which is labeled to a smaller extent and rest large portion of it is unlabeled

Reinforcement Learning:

In this technique, model keeps on increasing its performance using a Reward Feedback to learn the behavior or pattern