Module 1

Machine Learning

→ is a type of Artificial Intelligence that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so.
→ Use historical data as input to predict new output values.

Difference b/w ML & AI

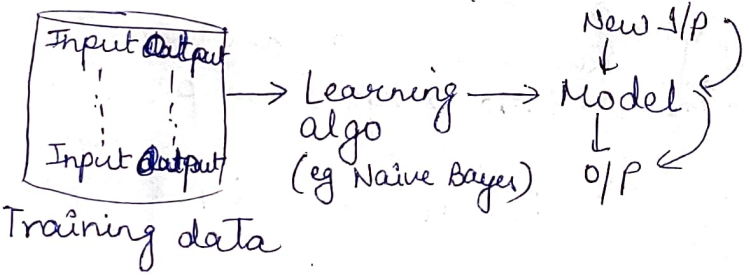
AI is a technology that enables a machine to simulate human behavior. ML is a subset of AI which allows a m/c to automatically learn from past data without programming explicitly.

Types of ML

- Supervised learning
- Unsupervised learning
- Semi-Supervised learning
- Reinforcement learning

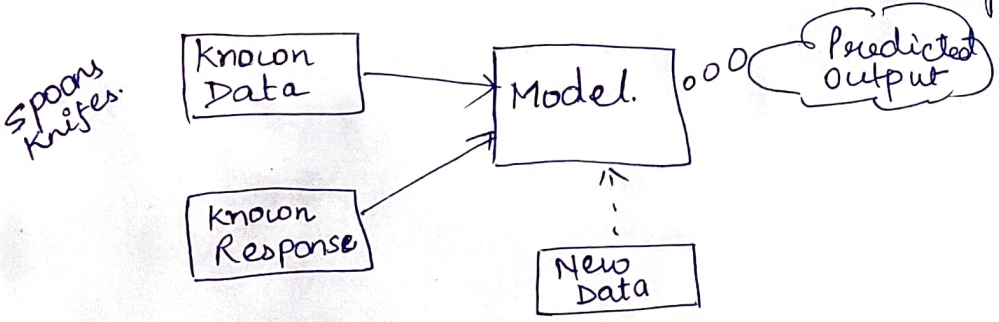
1) Supervised learning

As the name says we have a supervisor who is giving us instructions, here the supervisor is the training data (means we have i/p as well as o/p)



ex exit polls.
→ take data from diff category of people & gets o/p who will win our lose. Based on the labeled data the model is trained. When actual voting is done, the model based on the training data makes prediction as who will win or lose.

→ m/c is trained using "labelled" dataset
Labelled data means o/p is already known to us.
→ M/c is trained with the i/p & corresponding o/p.



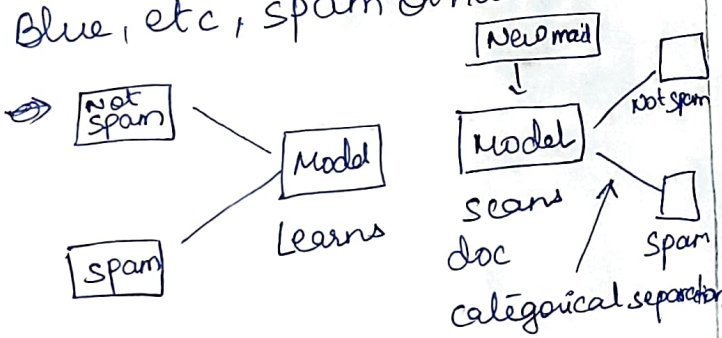
Suppose we have an input dataset of cats & dogs images. First we provide training to the m/c to understand the images such as shape & size of the tail of cat & dog, shape of eyes, colour, height etc. After completion of training, we i/p the picture of a cat & ask the m/c to identify the object and predict the output. Now m/c is well trained, so it will check the features of the object such as height, shape, colour, eyes, etc & find that it is a cat. So it will put it in cat category.

→ The main goal of supervised learning is to map the i/p variable (x) with the o/p variable (y).

SL

Classification Algo

- classification algorithms are used to solve classification problems in which o/p variable is categorical such as Yes or No, Male or Female, Red or Blue, etc, spam or ham



Regression Algo

→ we want to predict the score of students of a particular subj based on last year scores

→ o/p variable to be predicted is a continuous variable / numeric variable. Like in scores ex the scores can be b/w 0-100.

→ Regression algorithms are used to solve regression problems in which there is a linear relationship b/w i/p & o/p variable

* If I say will I get a salary hike that is classification but if I say how much salary rise will I get is Regression.

Applications of SL

Risk Assessment
(to assess risk in financial magament, insurace domain to minimize risk portfolio of the company)

Image classification
image classification is performed on diff image data with pre-defined labels

Spam detection
Fraud detection (fraud transactions, fraud customers. by using historic data)

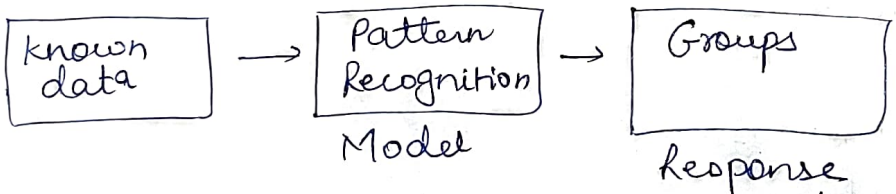
Supervised Learning Method

Label

	Label	Types of Lable
Regression ✓	score of a particular student	Continuous
classification ✓	spam or ham	categorical

Unsupervised Learning

→ As the name says, there is no need for supervision.
→ The m/c is trained using unlabeled dataset & the m/c predicts the o/p without any supervision.



ex suppose there is a basket of fruit images & we i/p it into the m/c learning model. The images are totally unknown to the model & the task of the m/c is to find the patterns & categories of the objects. So now the m/c will discover its patterns & differences such as colour difference, shape difference & predict the o/p when it is tested with test dataset.

UL

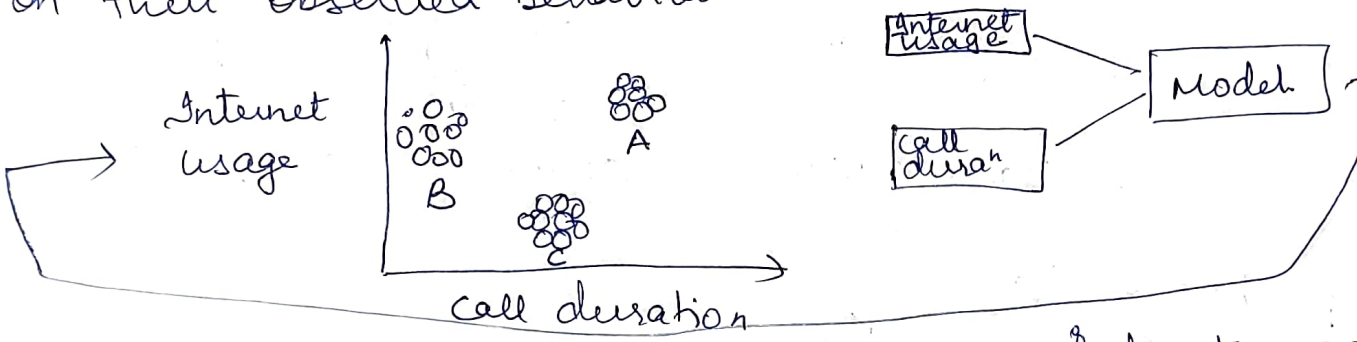
clustering
The method of dividing the objects into clusters which are similar b/w them

Association.
Discovering the probability of the co-occurence of items in a collection

ex which customer makes similar product purchasing this is clustering whereas association is which products are purchased together.

Example of Clustering

suppose a telecom company wants to reduce its customer churn rate. It takes total call dura^n & ~~inter~~ internet usage as an i/p data which is fed to the model. The model based on the phone usage & internet usage group customers. The customers are grouped based on their observed behaviour



As we can see here cust B has more internet usage so they will be given cheap internet plans, C has more call dura^n so they will be given cheaper call rates

Example of Association

Customer 1	Customer 2	Customer 3
Bread	Bread	If a customer purchases bread he is likely to purchase milk too.
Milk	Milk	
fruits	Rice	
wheat	Butter	

Applica^n of UL

- Recommendation System
- Network Analysis → identifying plagiarism & copyright
- Semantic Clustering → groups similar queries into clusters

3). Semi-Supervised Learning

→ lies between Supervised and Unsupervised machine learning

Problem with supervised & unsupervised

SL needs labeled data. We have enormous amount of data available in the world in the form of text, images, audio, videos but only a small fraction of such data is actually labeled. This process of labeling data whether algorithmically or by hand is a costly process. UL disadvantage is that its applicⁿ spectrum is limited

Semi-supervised = SL + UL

- Pick up the large unlabeled data set
- Label a small portion of the dataset
- Put the unlabeled dataset into clusters using UL algo.
- Build your model to use the labeled data to label & classify the rest of the unlabeled data.

Applications

Internet content classification

There are millions of webpages. It is practically impossible to label all the webpages SSL helps to classify webpages

Audio / Video analysis

Large amount of audio & video files, labeling them is a massive task. SSL solves this problem

4) Reinforcement Learning

→ there is an agent which is trying to accomplish a task There is an environment in which an agent is working The environment could be a maze in which an agent is finding the way out, an agent trying to see what sort of request to be entertain in cab service scenario

- solving a maze
- Managing Investment Portfolio
- Deciding among the pickup requests in cab service scenario
- Process Control System

- The agent is any robot that is trying to learn the task. A agent is an abstraction, it could be person, robot, a piece of software, a controller, h/w. trying to behave optimally in a given environment.
- Environment is whatever is surrounding an agent Environment is the world around it that gives it the feedback.
The agent doesn't have a direct control over the environment.

Reinforcement learning is similar to human learning. Remember the first time you were trying to learn to ride a bicycle. Learning how to balance comes with experience. Maybe, when you had a fall (negative experience), you learnt that the action which led to the fall was wrong and you should not do that again. Similarly when you had a positive experience you learnt what actions led to a happy ride.

- It is about taking suitable action to maximize reward in a particular position. It is employed by various softwares & machines to find the best possible behavior or path it should take in a specific situation.

Input - initial state from which the model will start
Output - There are many possible outputs as there are a variety of solutions to a particular problem
Training - is based on i/p. The model will return a state & the user will decide to reward or punish the model based on its o/p.

The model continues to learn.
The best solution is decided on the max reward.

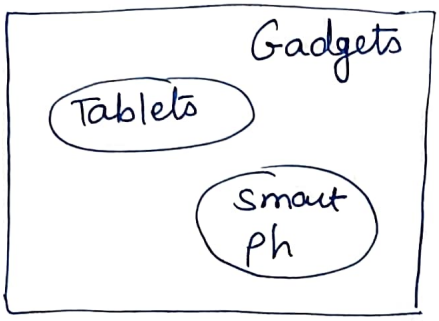
Difference b/w Supervised and Reinforcement Learning -

In supervised learning the training data has the answer key with it, so the model is trained with. correct answer itself but in reinforcement learning there is no answer but the agent decides what to do to perform the given task. In the absence of a training dataset, it the bound to learn from its experience.

Applications

1) used in robotics for industrial automation
2) Robot navigation
3) Game playing such as tic-tac-toe, chess, etc
4) Finance sector for evaluating trading strategies.

Concept Learning - helps to ~~learn~~ find all consistent hypothesis or concepts.



Each and every gadget will have some features Those features are defined as Binary valued attributes.

Size : large, small — x_1
Color : blue, black — x_2
Screentype: flat, folded — x_3
Shape: square, rectangle — x_4

Concept = $< x_1, x_2, x_3, x_4 >$

Tablet = < large, blue, flat, square>

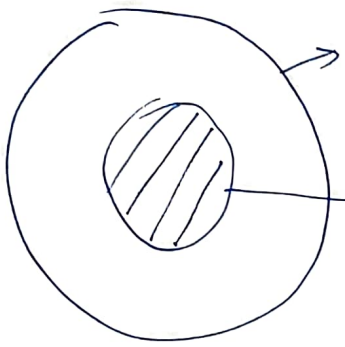
Smart phone = < small, black, folded, rectangle>

No. of possible instances (X) = 2^d (d = no. of features)
$= 2^4 = 16$

Total possible concepts = $2^{2^d} = 2^{16}$

We are not going to train the machine with these 2^{16} concepts.

Only the concepts which are consistent will be Choosen.



* Based on the features of say tablet $< C, B, F, S >$ we will either reject or accept the hypothesis

There are 2 types of hypothesis

1) Most specific hypothesis

$< \phi, \phi, \phi, \phi >$ Reject all

(Rejecting all types of features)

2) Most general hypothesis

$< ?, ?, ?, ? >$ → accept all

Concept Learning

→ boolean-valued function defined over a large set of training data.

ex Find a day when XYZ enjoys his favorite sport.
Attributes like Sky, AirTemp, Humidity and based on this we have target concept EnjoySport.

TPE (Task, Performance, Experience)

Problem: Learning the day when XYZ enjoys the sport

Task T: Learn to predict the value of EnjoySport for an arbitrary day based

Performance measure P: Total percent of days (EnjoySport) correctly predicted

Training Experience E - A set of days with given labels (EnjoySp

Concept Learning

Target concept / Ground Truth

Example	Sky	Temp	Wind	Play
1	Sunny	Warm	Strong	Yes
2	Sunny	Freezing	Strong	Yes
3	Rainy	Cold	Strong	No
4	Sunny	Warm	Strong	Yes

Attributes

Each attribute can have different values and depending on these values we can have two answers for play Yes or No. So it is a binary-valued function

Learn what? Learn if XYZ will play on an arbitary day or not so that we can predict this in future

Learn from what? From training examples (1 to 4)

Can the machine generalise what it has learnt from training examples

Concept - spirit / group / category of something
eg bird is a concept that includes all the birds that are existing in the world
→ concept is a boolean func^{n}
In bird example if given Tiger the answer would be No as tiger doesn't fall under bird category

Concept learning is a Search problem. Search happens in hypothesis space

hypothesis is a set of constraints on attributes
like <Sunny, Warm, Strong>

The m/c tries to search all possible hypothesis to find the best hypothesis. The best hypothesis is one that fits the training set the best → attributes on a particular day
$h(X)$ = Yes/No.

The h(x) has to be strong enough that if we give first row <Sunny, Warm, Strong> it is able to produce Yes.

We want to successfully find the hypothesis which is optimal which fits our training set successfully

General hypothesis → ? (accept) → any value is acceptable <Sunny, ?, Strong>

Specific → ϕ (reject) <ϕ, Warm, Strong>

↓ no value acceptable

↘ null

Find S algorithm

→ most specific hypothesis that fits all training examples.

Algorithm (considers only positive examples)

Yes - positive, No - negative

Step 1 Initialise with most specific hypothesis

$h_0 = \langle \phi, \phi, \phi, \phi, \phi \rangle$ ⇒ 5 attributes

Step 2 For each +ve sample → leave -ve

for each attribute

if (value = hypothesis value) ⇒ Ignore

↓ (removal) ↓ (which is ϕ) (write it as it is)

else

Replace with the next most general hypothesis (?)

* moves from most specific to most general hypothesis

1. Initialize h to most specific hypothesis (2)

$h_0 = \langle \phi, \phi, \phi, \phi, \phi, \phi \rangle$

Example	Sky	Air Temp	Humidity	Wind	Water	Forecast	EnjoySport
1	Sunny	Warm	Normal	Strong	Warm	Same	Yes
2	Sunny	Warm	High	Strong	Warm	Same	Yes
3	Rainy	Cold	High	Strong	Warm	Change	No
4	Sunny	Warm	High	Strong	Cool	Change	Yes

2. $a_1\ a_2\ a_3\ a_4\ a_5\ a_6$

$h_0 = \langle \phi, \phi, \phi, \phi, \phi, \phi \rangle$

$x_1 = \langle$ Sunny, Warm, Normal, Strong, Warm, Same $\rangle$

For each positive training instance x
  For each attribute constraint a_i in h
    If the constraint a_i is satisfied by x
    Then do nothing
    else replace a_i in h by the next more general constraint that is satisfied by x

There is no match so → a_i

$\langle \phi, \phi, \phi, \phi, \phi, \phi \rangle$

This null is replaced by

$h_1 = \langle$ Sunny, Warm, Normal, Strong, Warm, Same $\rangle$

$x_2 = \langle$ Sunny, Warm, High, Strong, Warm, Same $\rangle$

$h_2 = \langle$ Sunny, Warm, ?, Warm, Same $\rangle$

$x_3 \rightarrow$ -ve

So $h_3 = h_2$ (kept as previous one)

$h_3 = \langle$ Sunny, Warm, ?, Strong, Warm, Same $\rangle$

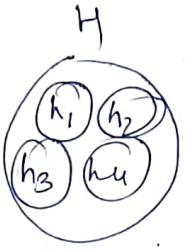
$x_4 = \langle$ Sunny, Warm, High, Strong, Cool, Change $\rangle$

Output $h_4 = \langle$ Sunny, Warm, ?, Strong, ?, ? $\rangle$ → Maximally Specific hypothesis

Version space

Subset of hypothesis H consistent with the training examples D

H ← hypothesis space.



$$VS_{H,D} = \{ h \in H \mid \text{consistent}(h, D) \}$$

h is consistent with D

$h(x) = c(x)$ ↑ target func^n

subset of H

Candidate Elimination Algorithm –

→ uses the concept of version space.
→ Considers both +ve and -ve values (Yes & No)
→ both specific and general hypothesis.

For positive samples, move from specific to general
For negative samples, move from general to specific

$S = \{ \phi \; \phi \; \phi \; \phi \}$ + ↓

$G = \{ ? \; ? \; ? \; ? \}$ − ↑

Dataset (Enjoysport)

	Sky	Temp	Humidity	Wind	water	Forecast	Enjoy
1.	Sunny	Warm	Normal	Strong	Warm	Same	Yes
2.	Sunny	Warm	High	Strong	Warm	Same	Yes
3.	Rainy	Cold	High	Strong	Warm	Change	No
4.	Sunny	Warm	High	Strong	Cool	Change	Yes

Algorithm

1. Initialise both general and specific hypothesis
   $S_0 = \{\phi\ \phi\ \phi\ \phi\ \phi\ \phi\}$ $G_0 = \{\phi\ ?\ ?\ ?\ ?\ ?\ ?\}$
2. For each example,
   if example is positive
   make specific to general
   else example is negative
   make general to specific

1) +ve (specific to general)

$S_1 = \{sunny, warm, normal, strong, warm, same\}$

$G_1 = \{?\ ?\ ?\ ?\ ?\ ?\}$

Compare S1 with example 2 as there is no initial hypothesis

$S_1 = \{S, W, N, S, W, S\}$

example2 $\{S, W, H, S, W, S\}$ → this is +ve so we will move from sp to ge.

$S_2 = \{S, W, ?, S, W, S\}$

$G_2 = \{?\ ?\ ?\ ?\ ?\ ?\}$

2) $S_2 = \{S, W, ?, S, W, S\}$

example 3 $\{R, C, H, S, W, C\}$ -ve so move from g to s

Keep specific hypothesis as it is so

$S_3 = S_2 = \{S, W, ?, S, W, S\}$ since here is ? which means it can take any value so we will not include

$G_3 = \{$

example3 $= \{R, C, H, S, W, C\}$

$G_3 = \{\langle sunny\ ?\ ?\ ?\ ?\ ?\rangle, \langle ?\ warm\ ?\ ?\ ?\ ?\rangle, \langle ?\ ?\ ?\ ?\ ?\ same\rangle\}$

Specialize G to exclude -ve example

~~$S_3 = \{S, W, H, S, C, C$~~

$S_3 = \{S, W, ?, S, W, S\}$

Example 4 $= \{S, W, H, S, C, C\}$ +ve

$S_4 = \{S, W, ?, S, ?, ?\}$.

$G_4 = \{\langle S, ?, ?, ?, ?, ?\rangle \langle ?, W, ?, ?, ?, ?\rangle\}$

In $G_3 = \{$ ——— $\langle ?, ?, ?, ?, ?, \text{same}\rangle$

and $S_4 = \{$? $\}$ There is a contradiction

Hence final ans is

$S_4 = \{S, W, ?, S, ?, ?\}$

$G_4 = \{\langle S, ?, ?, ?, ?, ?\rangle \langle ?, W, ?, ?, ?, ?\rangle\}$.

~~Linear Discriminants~~

Example	Citations	Size	Inlibrary	Price	Editions	Buy
1	Some	Small	No	Affordable	One	No
2	Many	Big	No	Expensive	Many	Yes
3	Many	Medium	No	Expensive	Few	Yes
4	Many	Small	No	Affordable	Many	Yes

Module 1

Origin	Manufacturer	Color	Decade	Type	ExampleType
Japan	Honda	Blue	1980	Economy	+ve
Japan	Toyota	Green	1970	Sports	-ve
Japan	Toyota	Blue	1990	E	+ve
USA	Chrysler	Red	1980	E	-ve
Japan	Honda	White	1980	E	+ve

1) +ve $G_1 = \{?\ ?\ ?\ ?\ ?\}$

$S_1 = \{J, H, B, 1980, E\}$

2) -ve example (J, T, G, 1970, S) compare with S_1

$S_2 = \{J, H, B, 1980, E\}$ specialize G to exclude negative example

$G_2 = \{(?\ H, ?\ ?\ ?), (?\ ?\ B\ ?\ ?), (?\ ?\ ?\ 1980\ ?), (?\ ?\ ?\ ?\ E)\}$

3) +ve example (J, T, B, 1990, E) compare with S_2

$S_3 = \{J, ?\ B, ?, E\}$

$G_3 = \{(?\ ?\ B\ ?\ ?), (?\ ?\ ?\ ?\ E)\}$

4) -ve example (U, C, R, 1980, E)

$G_4 = \{(?\ ?\ B\ ?\ ?), (?\ ?\ ?\ ?\ E)\}$

$S_4 = \{J, ?\ B\ ?\ E\}$