UNIT- I

INTRODUCTION TO MACHINE LEARNING.

MACHINE LEARNING:-

Machine learning is said as a subset of AI that is Mainly Concerned with the development of algo- which allow a computer to learn from the past data experience on their own. The term Machine learning was first introduced by Arthur samuel in 1959.

DEFINITION:-

ML enables a machine to automatically learn from data, improve performance from experiences and predict things without being explicity Programmed.

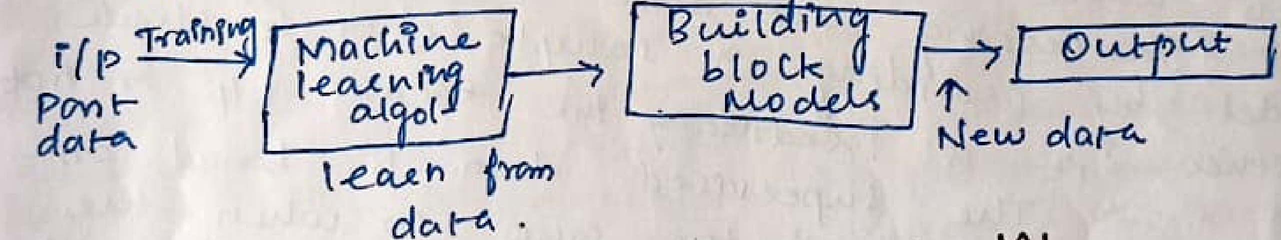


Fig: Working of ML.

A Machine s/m learns from historical data, builds the prediction models and whenever it receives new data and predicts the o/p for it.

Features of Machine learning:-

Machine learning uses data to detect various patterns in a given set.

It can learn from past data and improve automatically.

It is a data driven technology.

Machine learning is much similar to data mining as it also deals with huge amount of the data.

Classification of Machine learning:-

It is classified into three types. They are

1) Supervised learning.
2) Unsupervised learning.
3) Reinforcement learning.

SUPERVISED LEARNING:-

It is a type of ML Method in which we provide sample labelled data to the ML S/w in order to train it and on that basis it predict the o/p.

The S/w creates a model using labelled data to understand the datasets and learn about each data once the training and processing are done then we test the Model by providing a sample data to check whether it is predicting the exact o/p or not.

The supervised data is based on supervision and it is the same as when the student learns things in the supervision of the teacher. The example of supervised learning is spam filtering.

Supervised learning can be grouped further into two categories of algol-

1) Classification
2) Regression.

UNSUPERVISED LEARNING:-

It is a learning Method in which a Machine learns without any supervision.

The training is Provided to the Machine with the set of data that has not been labeled classified or categorized and the algo- needs to act on the data without any supervision. The goal of unsupervised learning is the to restructure the i/p data into few features or a group of objects with similar patterns.

In supervised learning we don't have a Predetermined result. The Machine tries to find useful insights from the huge amount of data. It can be further classified into two categories of algo.

1. Clustering.
2. Association.

Reinforcement learning:-

It is a feedback based learning method, in which learning agents gets a reward for each right action and gets a Penalty for each wrong action. The agent learns automatically with the feedbacks and improve its performance. In reinforcement learning, the agent interacts with the environment and explores it. The goal of an agent is to get the most reward points and hence it improve its performance.

LINEAR ALGEBRA :-

Linear algebra is an essential field of Mathematics, which defines the study of vectors, Matrices, planes, Mapping and lines required for linear transformation.

It plays a vital role and key foundation in Machine learning and it enables Machine learning algo- to run a huge number of datasets. The concepts of linear algebra are widely used in developing algol- in Machine learning. It can also Perform the following task:

1) Optimization of data.

2) Applicable in loss functions, regularisation, Covariance Matrices, singular value decomposition (SVD) matrix operations and support vector Machine classifications.

Implementation of linear regression in ML.

The linear algebra is also used in neural n/ws and data science field.

The general linear equation is represented as

$$\boxed{a_1 x_1 + a_2 x_2 \ldots\ldots + a_n x_n = b}$$

Where,

a = Represents the co-efficients

x = Represents the unknowns

b = Represents the constant.

Transposes and Inner Products :-

A collection of variables may be treated as a single entity by writing them as a vector.

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$x^T = [x_1 \quad x_2 \quad x_3].$$

The transpose operator also turns row vectors into column vectors. By defining the inner product of two vectors

$$x^T y = [x_1 \quad x_2 \quad x_3] \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

$$= x_1 y_1 + x_2 y_2 + x_3 y_3$$

$$= \sum_{i=1}^{3} x_i y_i .$$

The outer Product of two vectors produces a Matrix

$$xy^T = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \begin{bmatrix} y_1 & y_2 & y_3 \end{bmatrix}$$

$$= \begin{bmatrix} x_1 y_1 & x_1 y_2 & x_1 y_3 \\ x_2 y_1 & x_2 y_2 & x_2 y_3 \\ x_3 y_1 & x_3 y_2 & x_3 y_3 \end{bmatrix}$$

When applying the transpose matrix the i^{th} row becomes the i^{th} column. That is, if

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}.$$

Then

$$A^T = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & a_{33} \end{bmatrix}$$

Outer Product:-

The outer product of two Co-ordinate vectors is a matrix. If the two vectors have dimensions n and m, then their outer product is an $n \times m$ matrix.

$$c = a_{*1} B_{1*} + a_{*2} B_{2*} + a_{*3} B_{3*} + \cdots + a_{*n} B_{n*}.$$

Inverse:-

A matrix X its inverse X^{-1} is defined by the properties

$$X^{-1} X = I$$
$$X X^{-1} = I.$$

Eigen values and Eigen vectors:-

The eigen vectors x and eigen values λ of a matrix A satisfy → $Ax = \lambda x$.

Singular values and Singular vectors :-

A Singular value and pair of singular vectors of a square or rectangular matrix A are a nonnegative scalar σ and two non-zero vectors u and v so that

$$A_v = \sigma u,$$

$$A^H u = \sigma v .$$

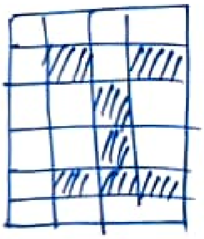
Examples of linear algebra in Machine learning:-

Datasets and Datafiles :-

Each ML Project works on the dataset and we fit the Machine learning model using this dataset.

$$\text{dog} = \begin{bmatrix} 0 \\ 7 \\ 0 \\ 0 \\ 5 \\ 1 \\ \vdots \\ 0 \end{bmatrix}$$

Images :-



$$\rightarrow \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 1 \end{bmatrix} .$$

YES/NO or Ratings :-

Given users and items (eg. movies), vectors can indicate if a user has interacted with the item (1 = yes, 0 = no) or the users ratings. say a number b/w 0 and 5.

$$\text{User 1} = \begin{array}{c|l} \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ \vdots \\ 1 \\ 0 \end{bmatrix} & \begin{matrix} \text{NO} \\ \text{YES} \\ \text{NO} \\ \text{NO} \\ \vdots \\ \text{YES} \\ \text{NO} \end{matrix} \end{array} \quad \text{or} \quad \text{USER 4} = \begin{array}{c|l} \begin{bmatrix} 0 \\ 5 \\ 0 \\ 3 \\ \vdots \\ 0 \\ 2 \end{bmatrix} & \begin{matrix} ? \\ \text{Love} \\ ? \\ \text{Like} \\ \vdots \\ ? \\ \text{Dislike.} \end{matrix} \end{array}$$

One Hot Encoding :-

How to represent non-numerical data such as language. One way is to apply one hot encoding.

Assign to each word a vector with one 1 and 0s elsewhere

$$apple = \begin{bmatrix}1\\0\\0\\0\end{bmatrix} \quad cat = \begin{bmatrix}0\\1\\0\\0\end{bmatrix} \quad home = \begin{bmatrix}0\\0\\1\\0\end{bmatrix}$$

$$tiger = \begin{bmatrix}0\\0\\0\\1\end{bmatrix}$$

How Machines learn ?

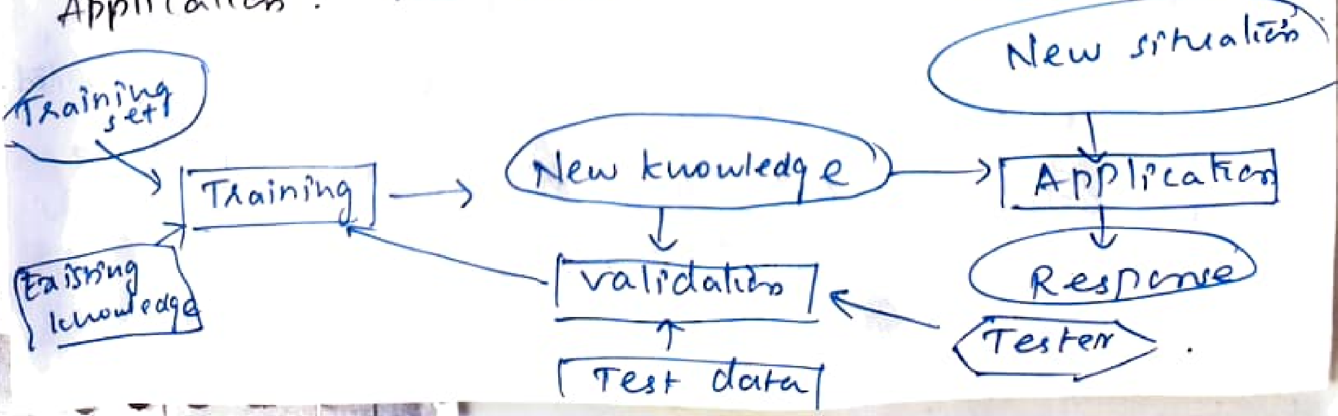
Machine learning typically follows three phases :

1. Training
2. validation
3. Application.

Training: A training set of examples of correct behavior is analyzed and some representation of the newly learnt knowledge is stored.

Validation: The rules are checked and if necessary, additional training is given.

Application: The rules are used in responding to some new situation.



Why Machine learning is important?

ML algol- can figure out how to perform important tasks by generalizing from examples.

Machine learning provides business insight and intelligence. Decision makers are provided with greater insights into their organizations. This adaptive technology is being used by global enterprises to gain a competitive edge.

Ingredients of Machine Learning:-

The ingredients of ML are as follows:

1 Tasks: The problems that can be solved with Machine learning. A task is an abstract representation of a problem. The standard methodology in Machine learning is to learn one task at a time. Large problems are broken into small, reasonably independent sub-problems that are learned separately and then recombined.

2. Models: The o/p of Machine learning. Different models are geometric models, probailistic models, logical models, grouping and grading.

3. Features: The workhorses of Machine learning. A good feature representation is central to achieving high performance in any Machine learning task.

Examples of Machine learning Applications:-

Optical character recognition: categorize images of handwritten characters by the letters represented.

Face detection: Find faces in images (or indicate if a face is present).

Spam filtering :

Identify email messages as spam or non - spam topic spotting: Categorize news articles (say) as to whether they are about politics, sports, entertainment etc..,

Spoken language understanding: within the Content of a limited domain. determine the Meaning of something uttered by a speaker to the extent that it can be classified into One of a fixed set of categories.

Face Recognition and Medical Diagnosis:-

Face Recognition :-

It is effortlessly and every day we recognize our friends, relative and family Members. We also recognition by looking at the Photographs. In photographs, they are in different Pose, hair styles, background light, Makeup and without Makeup.

Medical diagnosis :-

The i/p are the Relavant information about the patient and the Classes are the illness. The i/p contains the age of Patient's, gender, Past medical history and current symptoms.

Google Home and Amazon Alexa :-

Amazon Alexa / Siri :-

Everytime Alexa or Siri Make a Mistake when responding to our request, it uses the data it receives based on how it responded to the original query to improve the next time.

Google Home:-

Google services such as its image search and translation tools are sophisticated machine learning which allow computers to see, listen and speak in much the same way as human do.

Unmanned Vehicles:-

An unmanned Aerial vehicle (UAV), sometimes known as a drone, is an aircraft or airborne s/m that is controlled remotely by an onboard computer or a human operator. The grand control station, aircraft components, and various types of sensors make up the UAV s/m.

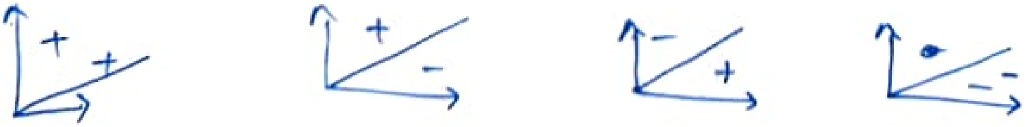
VAPNIK - CHERVONENKIS (VC) DIMENSION:-

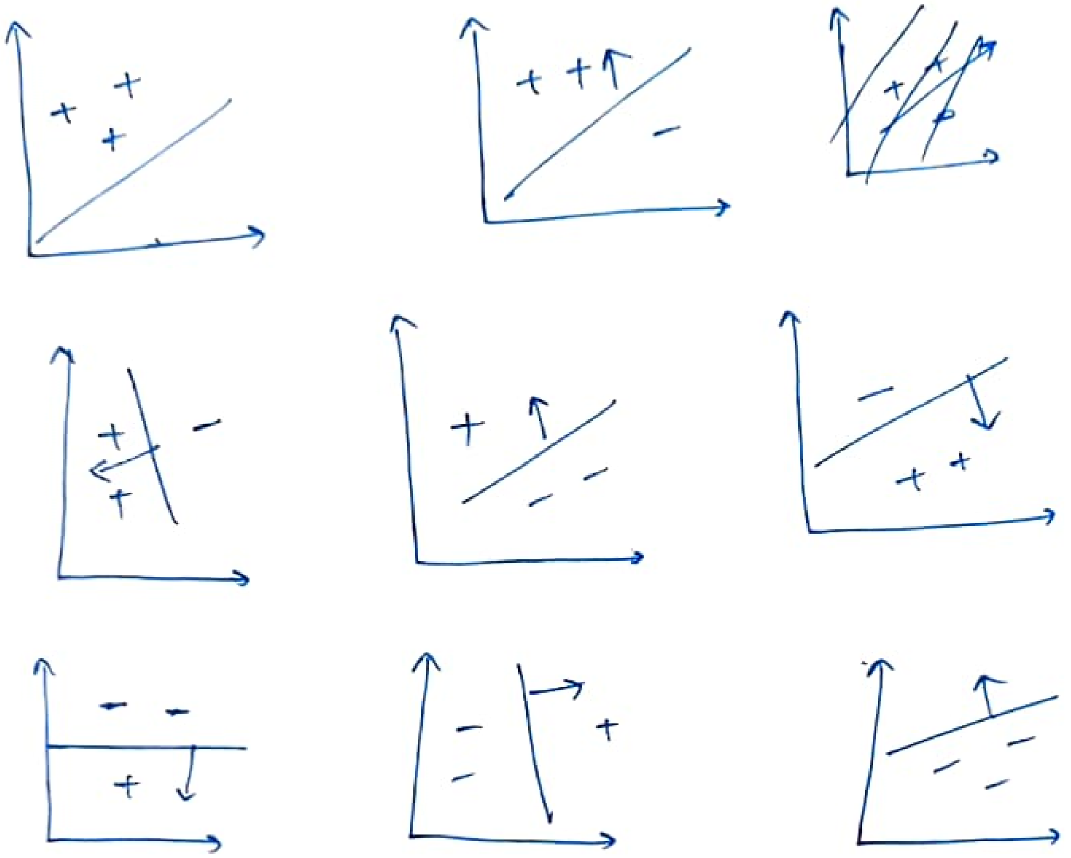
The Vapnik - Chervonenkis dimension, more commonly known as the VC dimension, is a model capacity measurement used in statistics and machine learning.

Shattering:-

It is the ability of a model to classify a set of points perfectly. The model can create a function that can divide the points into two distinct classes without overlapping. It is different from simple classification because it considers all possible combinations of labels upon those points. (2^N Possible ways).

Find VC Dimension:-



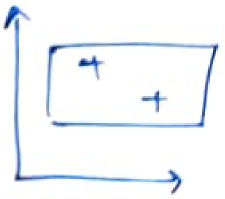


2^4 → is not Possible

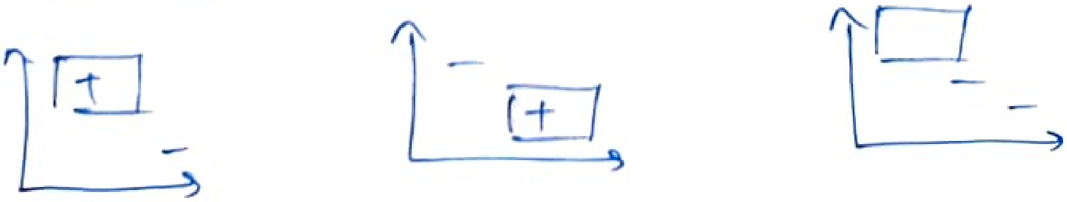
2^3 is Possible

In VC dimension Maximum no. of. point in R^2 that be shattered by straight line is 3.

VCD (Axis aligned Rectangle)



→ outside the rectangle negative
Inside the Rectangle Positive



until 4 Point. Possible
5 → not Possible $\boxed{VCD = 4 \text{ (axis aligned)}}$.

Probably Approximately Correct Learning :-

PAC learning is a framework for the mathematical analysis of machine learning.

<u>Goal of PAC</u> :- With high probability ("Probably"), the selected hypothesis will have lower error ("Approximately Correct")

In the PAC model, we specify two small parameters, ε and δ, and require that with probability atleast (1- δ) a system learn a concept with error at most ε.

ε and δ Parameters :-

ε gives an upper bound on the error in accuracy with which h approximated (accuracy: 1- ε).

δ gives the probability of failure in achieving this accuracy (confidence : 1- δ).

X instance
C Target
H hypothesis
D Training data

Example :-

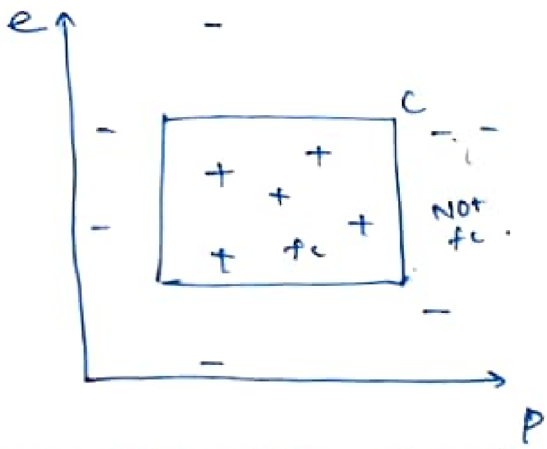
N number of car having Price and Engine Power, as training set (p, e), find the car is family car or not.

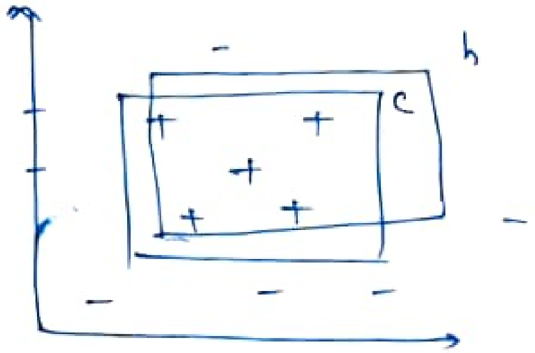
An algo- gives answer whether the car is family car or not.

C - Target function.

Instances within rectangle represents family cars and outside are not family cars.

<u>Hypothesis h</u> - closely approximate c, and there may be error region.

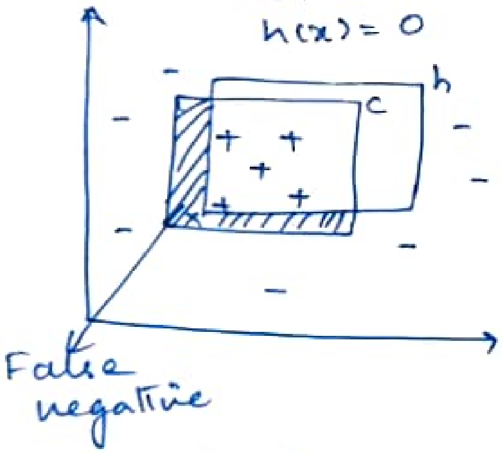




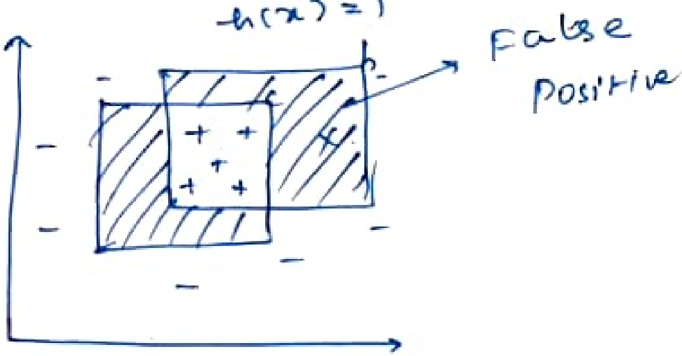
False Negative & False +ve :-

Instances lies on shaded region are +ve/-ve according to our actual function 'c', but those are -ve/+ve based on the hypothesis h. Hence it is called as false -ve or false +ve.

$c(x) = 1$
$h(x) = 0$



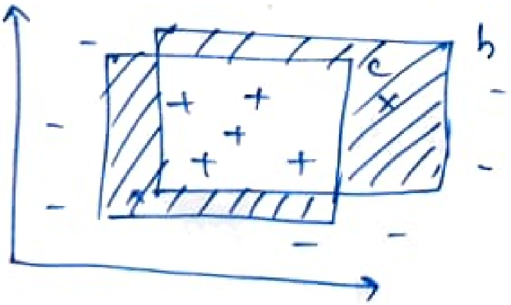
$c(x) = 0$
$h(x) = 1$



Error Region :-

The Probability of error region to be Small. The error region : $P(C \text{ XOR } h) <= \varepsilon$.

$P(C \oplus h)$.



Error region : c xor h

Approximately correct :-

the hypothesis h, that approximately correct, and error is less than or equal to ε where $0 <= \varepsilon <= 1/2$. (i.e) $P(C \text{ XOR } h) <= \varepsilon$

Probably Approximately Correct :-

low generalization error with high Probability

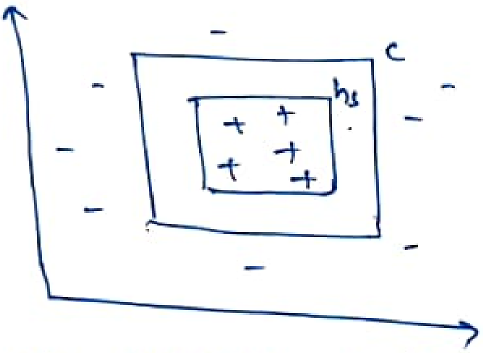
$[P[\text{error}(h) <= \varepsilon)] < = 1-\delta$

$P[P(C \text{ XOR } h) <= \varepsilon] <= 1-\delta$.

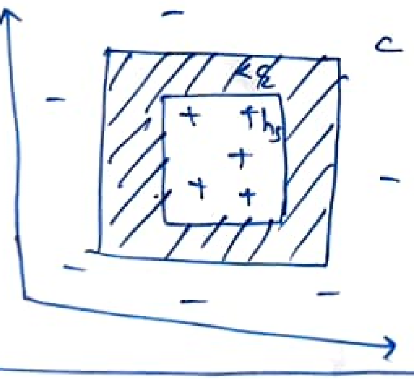
PAC learnability for axis - aligned rectangle :-

Specialization :-

h_s is the tighest possible rectangle around a set of +ve training. h_s is subset of c, Hence error region = c-h.



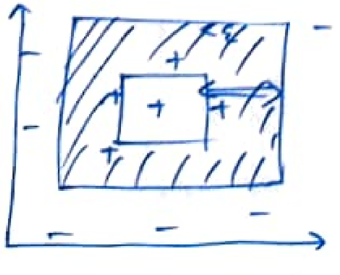
If an hypothesis lies b/w h and c (shaded region) then it is approximately correct.

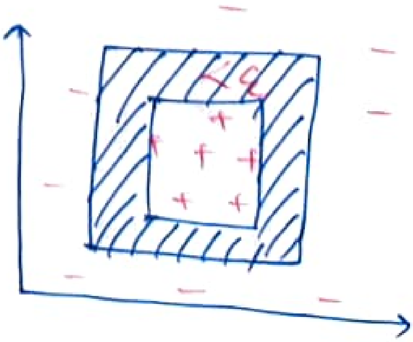


If the generated hypothe does not touch any of these region.

Error region is greater than ε and not approximately correct, because the error region got increased

Atleast one +ve example at each side of the rectangle.





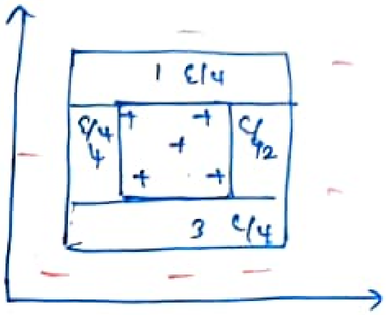
Error Region :-

Error region = Sum of four rectangular strips < ε.

Each strip is at most $\varepsilon/4$.

Probability of +ve example falling in any one of the strip (error region = $\varepsilon/4$)

Probability that a randomly drawn +ve example misses a strip = $1 - \varepsilon/4$. ⓧ.



P(m instance miss a strip) ≠ ~~ε/4~~
$= (1 - \varepsilon/4)^m$.

P(m instance miss any strip) < $4 (1 - \varepsilon/4)^m$.

Finally we get $m > 4/\varepsilon \log 4/\delta$.

Example 1 :-

Hypothesis h1 generated the errors power of with respect to price and engine power of given 10 samples.

Given $\varepsilon = 0.05$ $\delta = 0.20$

$P(h_1) >= 1 - \delta$

$P(h_1) = 8/10 = 0.80$

(3rd and 8th values are greater than ε)

$\therefore 0.80 >= (1 - 0.20)$ i.e. $0.80 = 0.80$ ⓧ

H1 is probably approximately correct.

Sl. No	Error(h1)
1	0.001
2	0.025
3	0.07
4	0.003
5	0.025
6	0.045
7	0.021
8	0.065
9	0.012
10	0.036.

HYPOTHESIS SPACES :-

The hypothesis space is defined on the supposition or proposed explanation based on insufficient evidence or assumptions. It is just as guess based on some known facts but has not yet been proved. A good hypothesis

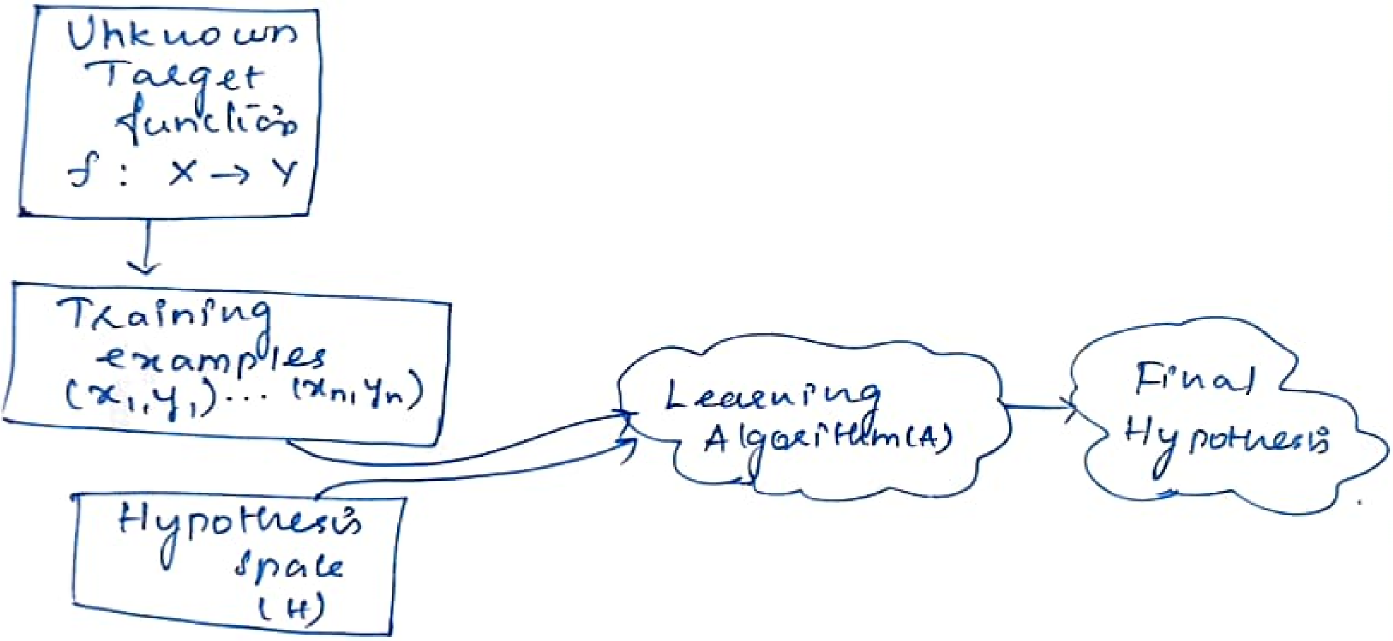
TRADE OFF :-

content is not contained in

- is testable, which results is either true or false.

Hypothesis in Machine learning (ML) :-

The hypothesis is one of the commonly used concepts of statistics in ML. It is specifically used in supervised learning where an ML model learns a function that best maps the i/p to corresponding o/ps with the help of an available dataset.



CANDIDATE ELIMINATION ALGORITHM :-

It is a method of learning concepts from data that is supervised. Given a hypothesis space H and a collection E of instances, the candidate Elimination Procedures develop the version space continuously.

TIME	WEATHER	TEMP	COMPANY.	HUMIDITY	WIND	GOES
MORNING	Sunny	Warm	Yes	Mild	Strong	yes
EVENING	Rainy	Cold	No	Mild	Normal	No
Moxing	Sunny	Moderate	Yes	Normal	Normal	yes
Evening	Sunny	Cold	yes	High	Strong	yes

VARIANCE TRADE OFF :-

G_3 = <Morning ? ? ? ? ?> <? Sunny ? ? ? ? ?> <? ? ? ? yes ? ? ?>

x_3 = < evening, Sunny, Cold, yes, High, Strong>.

+ve → specific

(x_2, x_3) S_4 = < ?, Sunny, ?, yes, ?, ? >

G_4 = < ? Sunny ? ? ? ? > <? ? ? ? yes ? ? >.

Specific hypothesis

< ?, Sunny, ?, yes, ?, ? >

General hypothesis

<? Sunny ? ? ? ? > <? ? ? ? yes ? ? >

Sample Error and True Error :-

The Sample error ($err_s(h)$) of h with respect to target function (f) and data sample (S) is the proportion of examples h misclassifies. The sample test error is the mean error over the test Sample.

$$err_s(h) = \frac{1}{n} \sum_{i=1}^{n} L(f(x_i), h(x_i)).$$

The true error of hypothesis h with respect to target function f and Distribution D is the probability that h will misclassify an instance drawn at random according to D.

$$Err(h) = E(L f(x)), h(x))].$$

INDUCTIVE BIAS :-

The Candidate Elimination Algol- will converge toward the true target concept provided

- it is given accurate training examples ✓
- its initial hypothesis space contains the target concept.

Bias → average squared diff. btw Prediction and true values

Generalize

Solve on what days the Person likes to go on walk using candidate elimination algol-.

Sol

CEA:-

- extended form of Find S algol-
- considers both +ve and -ve instances
- finds all the hypotheses that Math all the given training examples.

Attributes:- Time, weather, Temp, company, Humidity, wind.

Target : Goes for a walk or not.
(+ve = yes ; -ve = No).

S (specific hypothesis) = < ϕ, ϕ, ϕ, ϕ, ϕ, ϕ >

G (General hypothesis) = < ?, ?, ?, ?, ?, ? >

1) x_0 = <Morning, Sunny, warm, yes, mild, strong>

+ve → Specific (S)

(x_0, S_0) S_1 = <Morning, sunny, warm, yes, Mild, strong>

G_1 = < ?, ?, ?, ?, ?, ? >

② x_1 = < Evening, Rainy, cold, No, Mild, Normal >

-ve → General (G)

S_2 = < Morning, sunny, warm, yes, Mild, strong>

(x_1, S_1) G_2 = < Morning, ? ? ? ? ? > <Sunny. ? ? ? ? >
$(g^{nd}_{f}, 1^{st})$ < ? ? warm ? ? ? > < ? ? ? yes ? ? >
$(x_1, +d)$ < ? ? ? ? ? strong >.

③ x_2 = < Morning, Sunny, Moderate, yes, Normal, ~~Normal~~

+ve → specific (S)

(x_2, S_2) S_3 = < Morning, sunny, ? , yes, ? , ? >.

BIAS ~~AND~~ VARIANCE TRADE OFF :-

— ∴ experimental Practice we observe the bias

- What if the target concept is not contained in the hypothesis space ?
- Can we avoid this difficulty by using a hypothesis space that includes every possible hypothesis ?

A Biased Hypothesis space :-

Support we wish to assure that the hypothesis Space Contains the unknown target concept. The Obvious solution is to enrich the hypothesis space to include every possible hypothesis. To illustrate, consider the Enjoysport example in which we restricted the hypothesis space to include only conjunctions of attribute values.

< Sunny, High, Normal, Strong, cool, same >.

Example	Sky	Airtemp	Humidity	Wind	water	Forecast	Enjoysport
1	Sunny	Warm	Normal	Strong	Cool	change	Yes
2	Cloudy	Warm	Normal	Strong	Cool	Change	Yes
3	Rainy	Warm	Normal	Strong	cool	change	No.

S0 : < ϕ, ϕ, ϕ, ϕ, ϕ, ϕ >

S1 : < sunny, warm, Normal, Strong, Cool, change >

S2 :< ?, warm, Normal, Strong, cool, change >

S3 : < ?, warm, Normal, Strong, Cool, Change>

G2 : < ?, ?, ?, ?, ?, ? >

G1 : < ?, ?, ?, ?, ?, ?, ? >

G0 : < ? ? ? ? ? ? >

Because of the restriction, the hypothesis space is unable to represent even simple disjunctive target concepts such as "sky = Sunny or sky = cloudy".

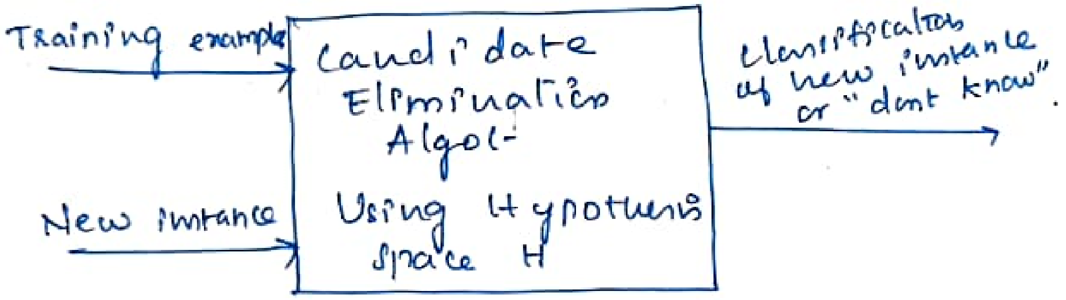
<"sky = Sunny or sky = cloudy", warm, Normal, Strong, Cool, change>.

Unbiased Learner :-

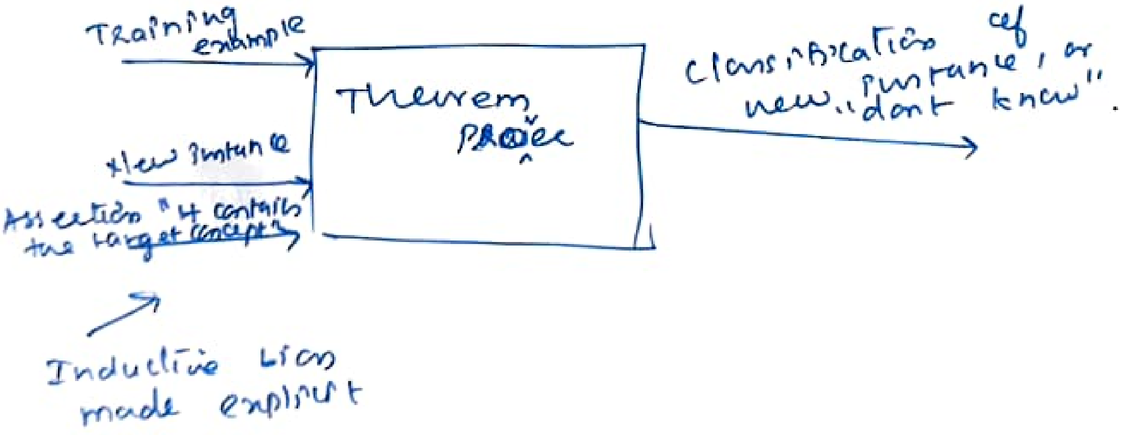
The obvious solution to the problem of assuring that the target concept is in the hypothesis space H is to provide a hypothesis space capable of representing every teachable concept; that is, it is capable of representing every possible subset of the instances X.

In general, the set of all subsets of a set X is called the powerset of X.

Inductive S/m :-


Training example
New instance
Candidate Elimination Algo-
Using Hypothesis space H
Classification of new instance or "dont know".

Equivalent Deductive S/m :-


Training example
New instance
Assertion "H contains the target concept"
Theorem prover
Classification of new instance, or "dont know".
Inductive bias made explicit

Generalization :

Generalization is a definition to demonstrate how well is a trained model to classify or forecast unseen data. Training a generalized machine learning model means, in general, it works for all subset of unseen data.

BIAS AND VARIANCE TRADE OFF :-

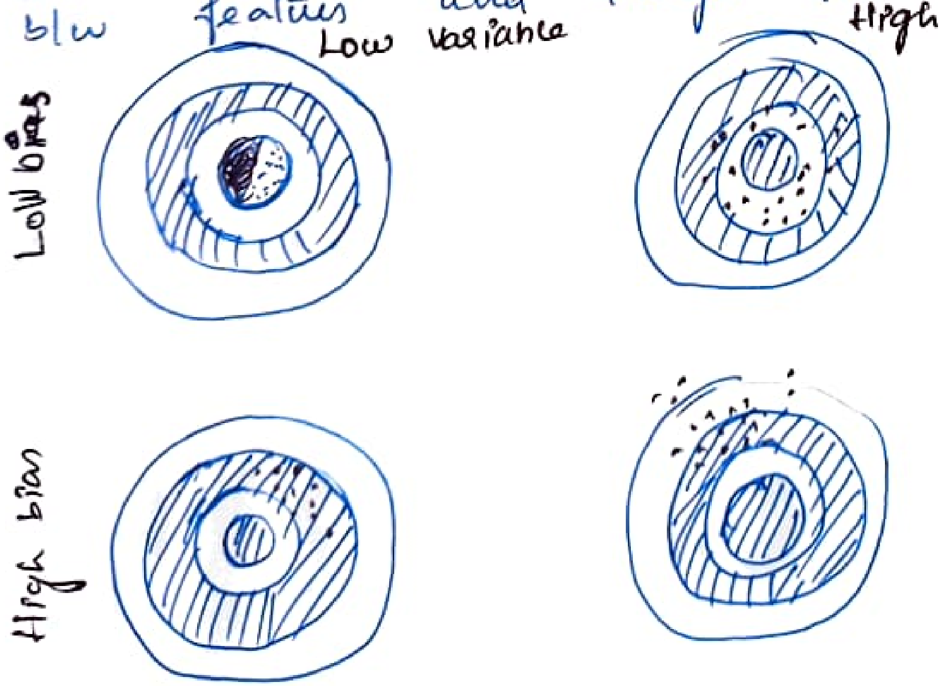
In the experimental Practice we observe an important phenomenon called the bias variance dilemma.

In Supervised learning, the class value assigned by the learning model build based on the training data may differ from the actual class value. This error in learning can be of two types, errors due to 'bias' and error due to 'variance'.

The bias - variance dilemma is the Problem of simultaneously minimizing two source of error that Prevent Supervised learning algol- from generalizing beyond their training set.

1. The bias is error from erroneous assumptions in the learning algol-. High bias can cause an algol- to miss the relevant relations b/w features and target O/Ps.

2. The variance is error from sensitivity small fluctuations in the training set. High varie can cause an algol- to miss the relevant relation.. b/w features and target O/Ps.

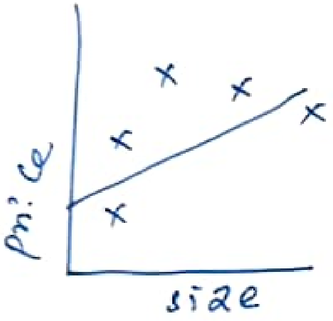


Bias - assumption
variation - deviate (how much)

Underfitting (High bias and low variance) :-

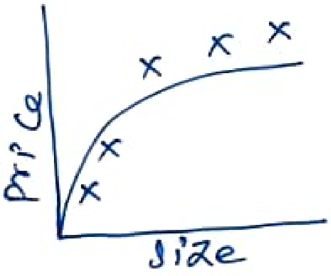
A statistical model or a Machine learning algo- is said to have underfitting when it Cannot capture the underlying trend of the data.

It can be avoided by using more data and also reducing the features by feature selection.



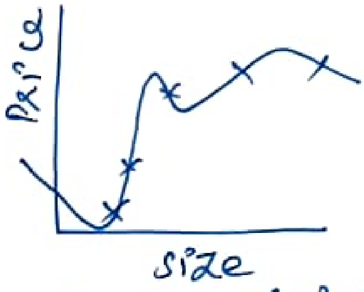
$\theta_0 + \theta_1 x$

High bias (underfit)



$\theta_0 + \theta_1 x + \theta_2 x^2$

High bias (underfit)



$\theta_0 + \theta_1 x + \theta_2 x^2 + \theta_2 x^2 + \theta_2 x^2$

High variance (overfit).

Overfitting (High variance and low bias)

A statistical model is said to be overfitted, when we train it with a lot of data.

When a model gets trained with so much of data, it starts learning from the noise and inaccurate data entries in our data set.

Then the model does not categorize the data correctly, because of too many details & noise.

A solution to avoid overfitting is using a linear algo- if we have linear data or using the parameters like a maximal depth if we are using decision trees.