UNIT - 1

Introduction to AI:

The AI (artificial intelligence) branch of computer science or science by which we can create intelligent machine, which can behave like a human and think like human and able to make decisions.

Artificial intelligence → Thinking power
↓
man made

* AI can create a machine which programmed algorithm which can work with own intelligence.
* AI can not need to use pre program machine to do same work

(or)

* AI is a branch of computer science which deals with helping machines finding solution to complex problems like human

Humanbeings → convert
↓
Algorithms → implemented
↓
machine

* AI is not related to computer science it is also related to maths, biology, [psychology] psylogy etc. etc.

ex:- Automated driving systems (self driving cars)

To calculate 100 students marks

* human take more time but computer can perform very fast.

Need of AI :

* AI can create software or devices which can solve real world problems very easily and accuracy.

ex: Health issues, supermarkets, traffic issues.

* AI can create your personal virtual assistance such as

Gemini, Google assistance, siri etc.

* AI can create [your] robots it works just like a human.

ex: sofia, alexa

Applications of artificial intelligence (AI) :-

Every branch of science, engineering and technology shares the tools and techniques available in the domain of AI.

1. Game playing :

Game playing is one of the leading domains where AI has been applied with great successful.

2. Expert Systems :

* An expert system is a software that manipulated encoded knowledge to problems in a specialization domain that normally requires human experts.
* An expert system is in AI program in which the system knowledge is obtained from an expert sources such that intelligence advice or intelligence decision in solving problems.

3. Natural language processing :

NLP is a technique that builds ability in machines to read and understand the languages that human speaks.

4. Image understanding :

* many of AI programs are engineered to solve problem without humans.
* 2D array contains grey levels can be used to received digital images that recognized by video camera.

5. Robotics :-

* AI is applied in robotics in order to see, hear and react to other sensory stimulation.

ex: Erica and Sophia.

6. Finance:

* The banks we use AI to perform different operations
* Organize operations (credit, debit)
* Invest in stocks
* Financial institutions have used AI to detect changes or claims etc.

7. Music:

Composition of song, performance sounds and research in music

8. Transportation:

fuzzy logic controllers have been developed for automatic gearbox in automobiles.

ex: Audi TT
VW TORG
VW Corvel
} automatic gearbox

9. Hospitals:

A medical clinic use AI system to organise bed schedule make a rotation, heart sound analysis identify tumors

10. Computer vision:

* It is able to extract information from it's vision computer vision plays it's role here to recognise the object as an image and identify the taste.
* The image data only can be in the form of pictures, videos, multidimensional data from a medical scanner or multiple cameras.

-Advantages of AI:

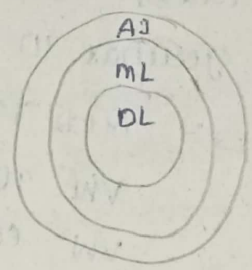
* High accuracy with less error
* High speed
* High reliability
* Useful for risk data
* Digital assistant
* Useful as a public utility

Disadvantages of AI:

* High cost
* No feelings, emotions
* No original creations
* can't think out of box.

⇒ machine Learning :-

* machine learning is the branch of artificial intelligence concerned with the design and development of algorithms.
* It allows the computer to behave in a way based on the empirical data, such as from electronic sensors or database.
* machine learning is a field of studying that gives computers a capability to learn without being explicitly programs.



ex:- Online shopping.

(or)

Machine learning is an subfield of artificial intelligence that provides system that ability to automatically and improve from experience without being explicitly programmed.

- Features of machine learning :-

* machine learning is used to detect various patterns in a given dataset.
* It can learn from past data and improve performance perfectly.
* It is a data driven technology

ML is similar to Datamining and it can work large amount of data.

- Types of ML :

There are three types of machine learning :

1. supervised learning
2. unsupervised learning
3. Reinforcement learning

1. supervised learning:-

* In supervised learning we teach the machine using label data to predict outcome for new unseen data
* An external super vision is there.
* The labelled data means some input data is already taged with current output
* If any wrong data is given then send feedback to machine.
* There are two types of supervised learning algorithm.
  1. Regression
  2. classification.

a. Regression :- It is used for prediction of continuous variables

ex: price, Gender, age and salary

b. classification : It is used for prediction of output variables

ex: True or false, positive or negative, high or loco.

2) Unsupervised learning :-

* unsupervised learning is a self learning algorithm with any super vision.
* The training is provided to the machine with the set of data that not been classified and the algorithms needs to act on that data without any supervision.
* The input data is unlabelled data.
* No supervision and no feedback.
* There are two types of unsupervised learning algorithms:
  1. clustering
  2. Association

a. clustering : Grouping of objects into clusters.

b. Association : to find out relationship between two objects.

3. Reinforcement learning :-

* It is a feedback based learning approach. here an agent learn the environment and perform the action to get result of actions.
* The reinforcement learning is an example or semi supervised learning because it works based on supervised data and unsupervised learning.
* There are two types of reinforcement learning :-
  1. postive reinforcement learning
  2. Negative reinforcement learning

Applications of machine learning :-

* machine learning is a buzzword for today technology, and it is growing very rapidly day by day.
* We are using machine learning in our daily life even without knowing it such as Google maps, Google assistant, Alexa etc.
* Below are some most trending real-word applications of machine learning

1. Image recognition :
* Image recognition is the one of the most common application of machine learning
* It is used to identify objects, persons, places, digital images etc

2. Speech recognition :
* while using Google, we get an option of "search by voice", it comes under speech recognition, and it's a popular application of machine learning
* examples : Google assistant, Siri, cortana and Alexa

3. Traffic prediction :
* If we want to visit a new place, we take help of Google maps, which shows us the correct path with th

shortest route and predicts the traffic conditions
* It predicts the traffic conditions such as whether traffic is cleared, slow-moving, or heavily congested with the help of two ways:
→ Real time location of vehicle form google map app and sensors
→ Average time has taken on past days at the same time.

4. Product recommendations:
• machine learning is widely used by various e-commerce and entertainment companies such as Amazon, Netflix etc for product recommendation to the user.

5. Self driving cars:
* One of the most exciting applications of machine learning is self-driving cars.
* machine learning plays a significant in self-driving cars.
* Tesla, for the most popular car manufacturing company is working on self-driving car.

6. Virtual personal assistant:
* We have various virtual personal assistants such as Google assistant, Alexa, Cortana, Siri.

7. Medical diagnosis:
* In medical science, machine learning is used for disease diagnosis
* with this, medical technology is growing very fast and able to build 3D models that can predict the exact position of lesions in the brain.

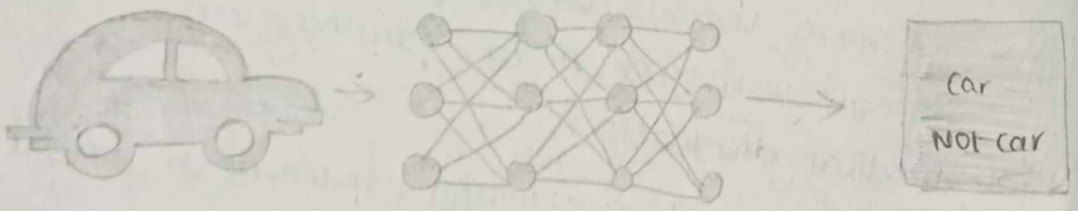
• Advantages of machine learning:-
* easily identifies trends and patterns.
* Continuous improvements.
* No human interaction needed.
* Handling multi-dimensional data.
* wide range of applicability

Disadvantages of machine learning :-

* changing nature of jobs.
* Highly expensive
* High error chances
* results interpretations
* Time & resources
* Data acquisition.

⇒ Deep Learning :-

* Deep learning is a synonym for deep[learning] structure hierarchal learning.
* It is a subfield of machine learning where supervised, unsupervised & semi supervised learning methods are adopted to learn from data representation.
* It contains set of algorithms which have been inspired from the structure and function of human brain.
* Deeplearning is used to feature extraction



input → feature extraction + classification → Output

* Deep learning will do imitate the human brain.
* Deep learning is a subfield of ML for learning feature hierarchies that are actually on artificial neural networks.
* Deep learning is implemented by the help of deep networks, which is nothing but neural networks with multiple hidden layers.

• Neural Networks : Neural networt is also know as artificial neural networks (ANN) simulated

neural network.

* most popular techniques used for implementing deep architectures are:

1. Artificial neural networks (or) multilayer perception (or) feed forward network.
2. convolutional neural network or deep forward N/w.
3. recurrent Neural network.

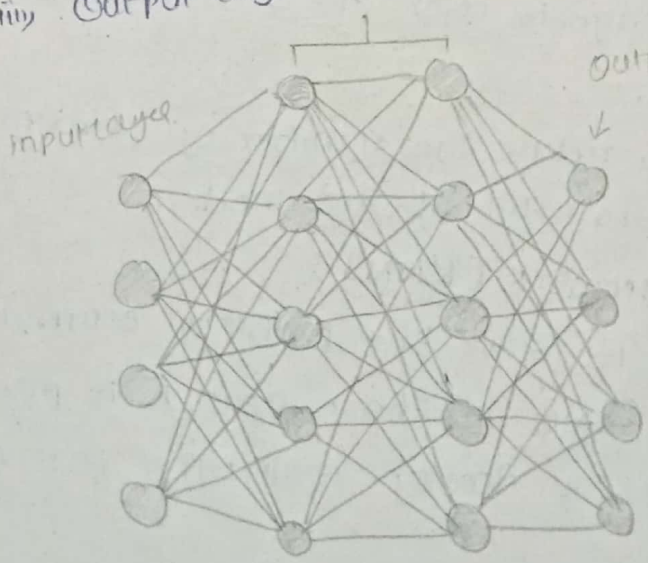
1. Artificial neural networks:

* Artifical neural network is inspired by the working of human brain.
* The human brain has neurons interconnected to one another ANN also have neurons that are interconnected to one another in various layers of the networks
* These neurons are known as nodes or units.
* Neural network are a set of algorithm that tries to recognize the patterns, relationships and information from the data.
* Artificial neural network is also called as multilayer perception or feed forward network.

Components of ANN:

A simple neural network consists of three (3) components

i) input layer
ii) hidden layer (middle layer)
iii) Output layer.



Input Layer : input nodes receives inputs / information fro
the Outside world.

Hidden Layer :

* Hidden Layer is a set of neurons where all the comp
-tations are performed on the input data.
* Any number of hidden layers are used in neural networ
But the simple network consists of only one hidden
layer.

Output Layer :

* The output layer is the output / conclusion derived from
all the computations performed.
* There can be single or multiple nodes in the o/p layer
* If we give binary classification problem the output
node is 1.
* If we are give input as multi class classification, the
output nodes are more than one.

2) Convolutional neural Network :

* Convolutional neural network (CNN) is similar to a
multilayer perceptron network.
* The difference is network learns structures and purpos
they are mostly used.
* CNN can be applied in domain of computer vision prob
/ image processing etc.
* If the given image is color then we use three (3) as size o
gray image use 1.
* The CNN used to reduce the number of parameters &
speed up the training of the model.

3) Recurrent Neural Network (RNN) :

* The RNN is used to solve the problem occurs in CNN.
* problems with CNN. we have input & it produce o
output. In CNN it doesn't maintain internal

* Every input is independent from other input
* CNN cannot solve problems like sentences, stock prices and time series.
* In RNN, each neuron or unit of RNN uses its internal memory to maintain information about the previous o/p when it required to predict the txt of word of sentence The previous word are required.
* The RNN is used to remember the previous output
* The most and main important features of RNN is hidden layers.
* The hidden layers have a memory which remembers all the information what has be calculated.
* It can also reduce the complexity of parameters.

Applications of Deep learning :-

* Automatic colouring of black and white images.
* automatically adding sounds to silent movies.
* automatic machine translation.
* object classification & detection in photographs.
* automatic text generations
* recommendation engine.
* chatbots and speech recognization.
* Image recognizations, IOT, Computer Vision.

⇒ main challenges of machine learning :-

* There are a lot of challenges that machine learning profess-ionals face to inculate ML skills and create an application from scratch.

1. poor quality of Data :

* Data plays a significant role in the machine learning process.
* one of the significant issues that machine learning professionals face is the absence of good quality data. unclean and noisy data can make the whole process extremely exhausting
* we don't want our algorithm to make inaccurate or faulty predictions, hence the quality of data is

essential to enhance the output.

2. Underfitting of training data :-

* The process occurs when data is unable to establish an accurate relationship between input and output variables.
* To overcome this issue:
  i. maximizing the training data.
  ii, Enhance the complexity of the model.
  iii, Add more features to the data.
  iv, reduce regular parameters
  v. increasing the training time of model.

3. Overfitting of training data:

* An overfitting is occurs when the low bias and high variance. that means the model training is very well but testing is very low.
* Training is properly done and but testing is not properly done.

How to avoid the Overfitting :-

a) cross validation
b) Training with more data
c. removing features
d. early stopping of training

* Both overfitting and underfitting cause the degraded performance of the machine learning model.

4) Lack of training Data:

* The most important task you need to do in the machine learning process is to train the data to achieve an accurate output.
* less amount training data will produce inaccurate results.

5. slow implementation:

* This is one of the common issues faced by machine learning professionals.

* the machine learning models are highly efficient in providing accurate results
* But it takes more time to provide accurate results.

Difference between artificial intelligence vs machine learnin vs Deep Learning.

artificial intelligence	machine learning	Deep Learning.
AI stands for artificial intelligence, and is basically the study/process which enables machines to mimic human behaviour through particular algorithm.	ML stands for machine learning, and is the study that uses statistical methods enabling machines to improve with experience	DL stands for deep learning is the study that makes use of neural networks (similar to neurons present in human brain) to imitate functionality just like a human brain
AI is the border family consisting of ML and DL as it's components	ML is the subset of artificial intelligence	DL is the subset of machine learning
AI is a computer algorithm which exhibits intelligence through decision making	ML is an AI algorithm which allows system to learn from data	DL is a ML algorithm that uses deep (more than one layer) neural networks to analyze data and provide output accordingly.
Search trees and much complex (which exhibits) math is involved in AL	if you have a clear idea about the logic involved in behind and you can visualize the complex functionalities like k-means, support vector machines, etc. then it defines the ML aspect	if you are clear about math involved in it but don't have idea about the features. So you break the complex functionalities into linear lower dimensional features by adding more layers, then it defines DL aspect

The aim is to basically increase chances of success and not accuracy	The aim is to increase accuracy not caring much about the success ratio	It attains the highest rank in terms of accuracy when it is trained with large amount of data
Three broad categories /types of AI are: artificial narrow intelligence (ANI), artificial General intelligence (AGI) and artificial super intelligence (ASI)	Three broad categories /types of ML are: Supervise learning Unsupervised learning reinforcement learning	DL can be considered as neural networks with large number of parameters layers lying in one of the four fundamental network architecture: Unsupervised pretrain-ed networks, convolutional neural networks, recurrent neural networks & recursive NN.
The efficiency of AI is basically the efficie-ncy provided by ML and DL respectively	Less efficient than DL as it can't work for longer dimensions or higher amount of data.	more powerful than ML as it can easily work for larger sets of data.
Examples of AI: Google's AI powered predictions, ridersh-aring Apps like Uber and lyft Commercial flights use an AI autopilot etc.	Examples of ML: Virtual personal assistant: siri, Alexa, Google etc, Email spam and malware filtering	examples of DL: sentiment Based news aggregation, Image analysis and caption generation etc.
AI systems can be rule based, knowledge-based or data driven	In reinforcement learning, the algorithm learns by trial and error, receiving	DL also consists of multiple layers of interconnected neurons that proce

	feedback in the form of rewards or punishments	data in hierarchical manner, allowing them to learn increasingly complex representation of data

Q difference between supervised and unsupervised learning:

parameters	supervised learning	Unsupervised learning
Definition	learning from labeled data to predict outputs	learning from unlabeled data to find patterns.
data types	requires labeled data (input - output pairs)	uses unlabeled data (only input features)
Objective	predict outcomes for new unseen data	discover hidden patterns or structure in the data.
Examples	email spam detection, house price prediction	customer segmentation, anatomy detection
Techniques used	regression, classification algorithms (eg: decision trees, SVM)	clustering, dimensionality reduction (eg: K-means PCA)
Output	predicts specific outputs (labels for values)	identifies groups, clusters or patterns.
Dependency on labels	Highly dependent on labeled data	No labels required for training
complexity	Generally simpler as labels guide the learning process	more complex as it involves finding structure on its own.
Real-world use cases	fraud detection, medical diagnosis	Social network analysis, segmentation

⇒ Sampling distribution of an Estimator [population]:

- **population :**

population is the totality of statistical data forming a subject of investigation

- **size :**

* The number of observations in the populations is defined to the size of population. It may be finite or infinite.

* size of the population denoted by N
* size of the sample denoted by n

- **Sample :**

* most of the times study of entire population may not be possible to carry out and hence a part alone is selected from the given population.

Sampling :

The process of selecting the sample is called sampling

Sampling distribution :

* A sample distribution is a probability of a statistics obtained from a larger number of samples drawn from a specific population.
* Sample statistic only estimate population parameters, mean, standard derivation.
* The sample mean will be different to the population mean
* A researcher will never know the exact amount of sampling. But using sampling distribution they can estimate the sampling error

1. Number of samples with replacement = N^n
2. Number of samples without replacement = NC_n

* Samples are classified in two ways :

1. large sample : $n > 30$
2. Small sample : $n \leq 30$

Parameters :

population related mean(μ), standard deviation(σ) Variance (σ^2), median etc are called parameters.

Statistics: Sample related mean($\bar{x}$), standard deviation (s), variance (s^2), median etc are called statistics.

Sample mean:

If $x_1, x_2, \cdots x_n$ represent a random sample of size 'n' then the sample mean is defined by $\bar{x}$

$$\therefore \bar{x} = \sum_{i=1}^{n} \frac{x_i}{n}$$

Sample variance:

If $x_1, x_2, \cdots, x_n$ represents a random sample of size 'n' then the sample variance is defined by s^2.

$$\therefore s^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n}$$

Sample error:

* The difference between sample mean statistic and the population parameter is called sample error.

Estimator:

* A tool in ML that uses data to make predictions or classify new instances.

⇒ Training and testing loss :-

Tip amount prediction

Bill amount	Tip amount
500	45
550	50
600	55
500	45
560	50
650	65
600	60
550	50
500	48
600	50

input → model → function → predict

test data

Linear regression:

i/p → model linear → function eg of line.

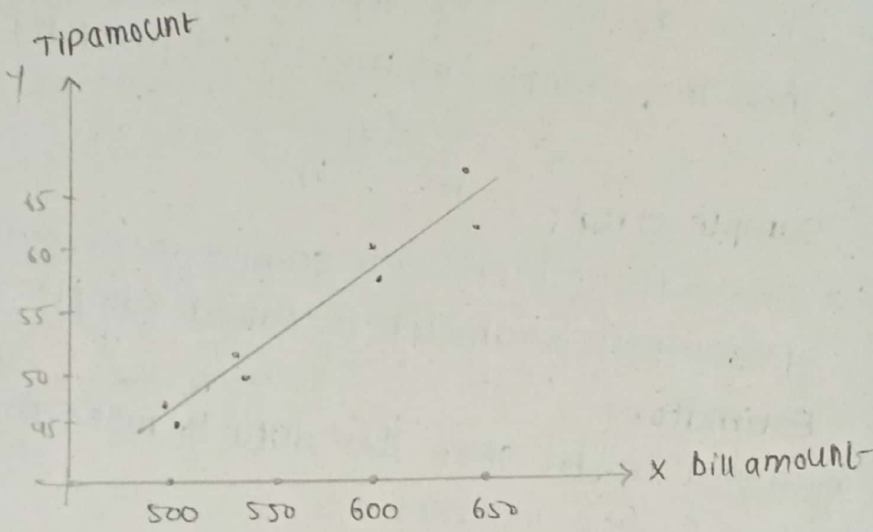
$\hat{y} = 0.1\,x + 2$

$\text{loss} = \frac{\sum_{i=1}^{n} (y - \hat{y})^2}{n}$

for testing : $\text{Test} = \frac{(7)^2 + (4)^2 + (12)^2}{3}$

$\text{Train} = \frac{(7)^2 + (7)^2 + \cdots}{7}$

Graph :-



* Training loss and test loss are used to measure how well a model performs.

Training loss : A measure of how well a model performs on the training data. It is used to guide the model's learning process.

Test loss : A measure of how well a model performs on new unseen data. It's an independent measure of the model's generalisation performance.

⇒ Empirical risk [management] (ERM) minimization :

Empirical risk minimization is a fundamental principle in statistical learning theory used to find a predictive model by minimizing the error or loss on a given training dataset. The objective is to choose a hypothesis from a hypothesis set that minimizes the empirical risk, which is the average loss on the training data.

- key concepts :

Hypothesis Set : A set of possible functions (models) from

which we aim to choose the best one.

Loss function: measures the error between the predicted and actual values.

* Common example include mean squared error for regressio and cross-entropy for classification.

Empirical risk: The average loss over the training sample if the training set consists of N samples and loss function is L, the empirical risk is given by:

$$R_{emp}(h) = (1/N) \Sigma L(h(x_i), y_i)$$

* where $h(x_i)$ is the prediction of the model 'h' for the input (x_i) and y_i is the actual output

Example:

Let's consider a simple example with linear regression.

1. Data: suppose we have a training set with three data points:

$$\{(1,2), (2, 2.5), (3, 3.5)\}$$

2. Hypothesis set: We consider linear function of the form

$$h(x) = wx + b.$$

3. Loss function: mean squared error (MSE):

$$L(h(x_i), y_i) = (h(x_i) - y_i)^2$$

4. Empirical risk calculation:

Suppose our model is $h(x) = 1.2x + 0.5$

- Calculate predictions:

$$h(1) = 1.2(1) + 0.5 = 1.7$$

$$h(2) = 1.2(2) + 0.5 = 2.9$$

$$h(3) = 1.2(3) + 0.5 = 4$$

- Calculate empirical risk:

$$R_{emp}[h] = \frac{1}{3}[(1.7-2)^2 + (2.9-2.5)^2 + (4.1-3.5)^2]$$

$$R_{emp}(h) = \frac{1}{3}(0.09 + 0.16 + 0.36)$$

$$= \frac{1}{3}(0.61) = 0.203$$

$$\therefore \boxed{R_{emp}(h) = 0.203}$$

Why ERM:

practical: provides a concrete criterion to select the best model based on observed data.

Theoretical foundation: Forms the basis for many machine learning algorithms, ensuring they generalize well to unseen data.

Flexibility: can be applied to various types of loss functions and hypothesis tests

* ERM is a core concept in machine learning and helps in building models that perform well on training data, hoping they will also perform well on new, unseen data.

Advantages of Empirical risk minimization:

* Foundation of supervised learning: ERM is a core principle in supervised learning, providing a systematic approach to model training and evaluation.
* Flexibility: ERM can be applied to various types of models and loss functions, making it versatile across different problem domains.
* Optimization framework: provides a clear objective for optimization minimization the empirical risk (avg loss on training data).
* Generalization: When combined with techniques like cross-validation and regularization, ERM help models generalize well on unseen data.
* Adaptability: can be applied to different learning algorithms from linear regression to deep neural networks, by choosing appropriate loss functions and hypothesis sets

- Disadvantages Empirical risk minimization:

1. Overfitting: Focusing solely on minimizing empirical risk may lead to overfitting, where the model performs well on training data set but poorly on new data.

2. Dependence on training data quality : ERM relies heavily on the quality and representativeness of the training data. If the data is biased or noisy, the model's performance will suffer.

3. Computational complexity : For large datasets or complex models, minimizing empirical risk can be computationally intensive, requiring significant resources.

4. Choice of loss function : The effectiveness of ERM depends on the choice of loss function, which can be challenging to select and may not always capture the true cost of prediction errors.

5. Local minima : In non-convex optimization problems, ERM can get stuck in local minima, leading to suboptimal models.

⇒ Estimating Risk statistics :

* risk management plays a crucial role in ML of building effective and reliable.
* At a high level the statistical risk measure the quality of the learning algorithm.
* In ML risk occurs when training data is more and testing data is less otherwise training data is less and testing data is more.

Techniques of risk statistics :

* Empirical risk minimization (ERM)
* [regulation] regularization risk minimization (RRM)
* Cross validation
* loss function etc.

Empirical risk minimization [ERM]

* As we don't know parameter 'p' we can't compute the true risk but we can compute the empirical risk based on x_i, y_i where $i = 1, 2, 3, 4 \cdots$
* It is a statistical learning algorithm used to find the optimal solution out of a set of possible solutions based on sample data.

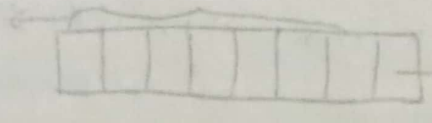
$R_{emp}(h) = \frac{1}{n} \sum_{i=1}^{n} L(y_i, h(x_i))$

* It is flexible and easy to implement
* risk occurs when there is a overfitting and underfitting in data models.
  - Overfitting: It occurs when there is a more training data.
  - Underfitting: It occurs when there is a more testing data than training data.

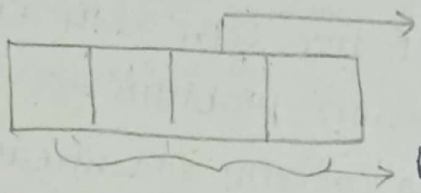
2. [Regulation risk] Regularization risk minimization:
* It is a machine learning technique that uses regularization to reduce overfitting in data models.
* By using set of methods (maths, statistics, computer etc) that trade a small decrease in training accuracy for better generalizability.
* There are, three primary regularization methods they are:
  - L1 regularization
  - L2 regularization
  - Elastic net

- L1 regularization: the process of adding sum of absolute values to the loss function
* it is also called "Lasso".
- L2 regularization: The process of adding sum of square values to the loss function.
* It is also called "ridge".
- Elastic net: The combination of L1 regularization (lasso) and L2 regularization (ridge) is called elastic net.

3) Cross validation:
In machine language, it is not suitable to use same data for both training and testing data.



cross validation is the mechanism to identify how much data is used for training and how much data is used for testing to get an accurate results.



Formula:

$$CV_k = \frac{1}{n} \sum_{i=1}^{n} loss\ (gT_{-k(i)}\ (x_i', y_i'))$$

* consider a dataset T of size n is divided into k folds
* loss function of training set $l(gT)$

4. loss function:

* loss function quantifies the difference between the predicted values and actual values.
* it is also known, as error function (or) cost function

categories of loss functions.

- classification model
- regression model

Regression model: It is a powerful tool for understanding & predicting relationships between the variables.

Classification model:

classification loss quantifies the error between prediction class labels and true class labels.

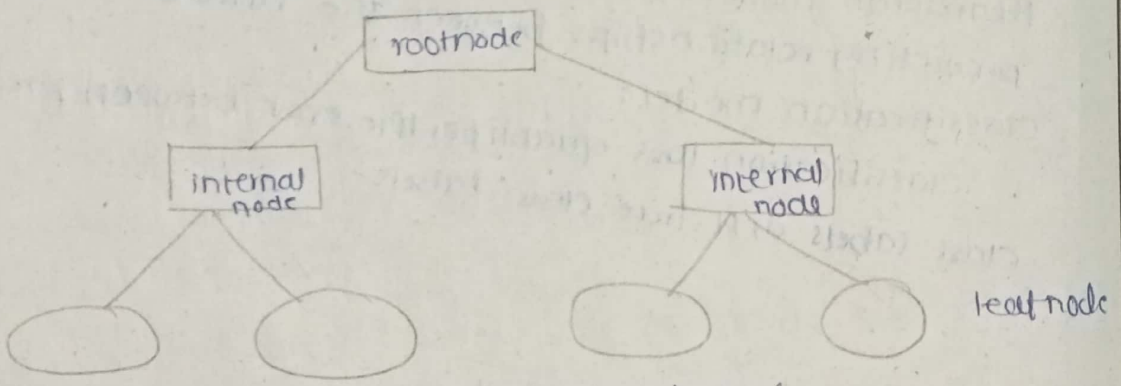
Unit-2

⇒ Decision tree Induction (or) Decision tree:

* Decision tree is supervised learning technique and by using this decision tree solve both classification problems as well as regression problems.
* Decision tree will represents classification model and regression model in the form of tree structure.
* Decision tree contains 3 types of nodes:
  1. root node
  2. internal node
  3. leaf node.

1. Root node: It has no incoming edge and zero or more outgoing edges.
2. Internal node: It has exactly one incoming edge and two or more outgoing edges.
3. Leaf node: It has exactly one incoming edge and no outgoing edges.

* each leaf node represents a class.



* Basically a decision tree are two types:
  1. regression tree
  2. classification tree

1. regression tree:

* for continues quantitative target variables
  ex: predictive, revenue, predicating rainfalls, predicating marks etc.

2. classification tree:

* for discrete categorical target variable
  ex: predicating positive or negative, predicating yes or no

predicating coin or loss etc.

* There are two types of algorithm in classification tree
  1. ID tree (iterative discodemizers)
  2. CART (classification regression tree)

* In ID tree algorithm by using entropy we built classification tree.

* In CART algorithm by using GINI index and impurity use

• Entropy : Entropy is measure of randomness or disorder (or) -ness

* Entropy measures the impurity of a collection of examples or points. It depends from the distribution of the random variables (P).

Formula :

$$\text{Entropy}(S) = -P_+ \log_2^{(P_+)} - P_- \log_2(P_-)$$

here S = collection of training examples / records

P_+ = it is a proportion of positive examples in 'S'.

P_- = it is a proportion of negative examples in 'S'.

• Information gain :

* Given entropy as a measure of the impurity in a collection of training examples, the information gain, is simply the expected reduction in entropy caused by partitioning the examples according to an attributes.

* The information gain, Gain (S, A) of an attribute (A). relative to a collection of examples (S) is defined as

$$\text{Gain}(S, A) = \text{entropy}(S) - \sum_{v \in \text{values}} \frac{|S_v|}{|S|} * \text{entropy}(S_v)$$

here. $S_v \rightarrow$ subset of S

value (A) $\rightarrow$ set of all possible values for the attribute A.

Decision tree algorithm - ID3 solved examples.

Instance	classification	a1	a2
1	+	T	T
2	+	T	T
3	-	T	F
4	+	F	F
5	-	F	T
6	-	F	T

Attribute : a_1

We have 6 examples or points or records There are class labels + (+), -.

$S[3+, 3-]$

$\text{entropy}(S) = -P_+ \log_2(P_+) - P_- \log_2(P_-)$

$= 1.0$

$\therefore$ entropy (S) = 1

1st attribute : True values {+, -}

$S[2+, 1-]$

$\therefore \text{entropy}(S_T) = -P_+ \log_2(P_+) - P_- \log_2(P_-)$

$= -\frac{2}{3}\log_2(2/3) - (1/3)\log_2(1/3) = 0.9183$

2nd attribute : False, values {+, -, -}

$\therefore S[1+, 2-]$

$\therefore \text{entropy}(S_F) = -P_+ \log_2(P_+) - P_- \log_2(P_-)$

$= -\frac{1}{3}\log_2(1/3) - \frac{2}{3}\log_2(2/3) = 0.9183$

$\therefore$ information Gain $(S, a_1) = \text{entropy}(S) - \sum \frac{|S_v|}{|S|} * \text{entropy } P(S)$

$= 1 - \frac{3}{6}\text{entropy}(S_T) - \frac{3}{6}\text{entropy}(S_F)$

$= 1 - \frac{3}{6}(0.9183) - \frac{3}{6}(0.9183)$

$= 1 - 0.45915 - 0.45915 = 1 - 0.9183$

$= 0.0817$

Attribute : a_2

1st attribute : True values = {T, T, F, F}

$\therefore S[3+, 3-] = 1$

$S_T = [2+, 2-]$

$\therefore$ entropy $(S_T) = 1$

2nd attribute : $S_F = [1+, 1-]$

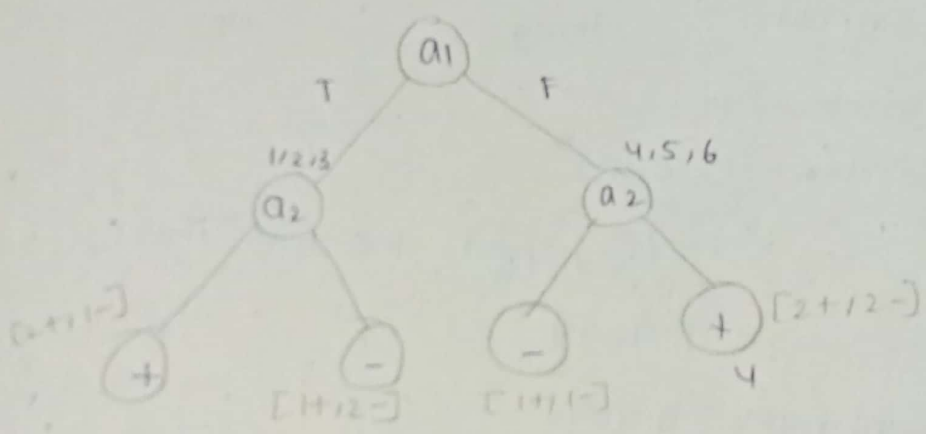
entropy $(S_F) = 1$

information Gain (S, a_2) = entropy $(S) - \sum \frac{|S_v|}{|S|} *$ entropy (S)

$= 1 - \frac{4}{6} * 1 - \frac{2}{6}(1)$

$= 1 - 0.6666 - 0.3333$

$= 0.0001$



⇒ GINI Index :

- The GINI index is used in CART algorithm
- Decision tree algorithms are famous decision for classification and regression tree.
- GINI index is also called as GINI coefficient or GINI impurity.
- GINI coefficient is a significant impurity measure utilized in decision tree algorithm.
- The GINI index is a proportion of impurity or inequality or in statistical settings
- They will investigate the idea of GINI index exclusively it's a numerical formula and it's applications in machine learning

Formula : GINI index $= 1 - \sum_{i=1} (P_i)^2$

here P_i = ith probability

ex> Identify the attribute that will act as the root node of decision tree to predict golf play for following data -base with index. Indicate an intermediate steps.

Outlook	wind	playgolf
Rain	strong	No
sunny	weak	yes
Overcast	weak	yes
Rain	weak	yes
sunny	strong	yes
rain	strong	No
Overcast	strong	No

Gini index = [4+, 3-]

Gini index = $1 - \sum_{i=1}^{v} (P_i)^2$

$= 1 - \left(\frac{4}{7}\right)^2 - \left(\frac{3}{7}\right)^2 = 1 - 0.3265 - 0.1836$

$= 0.4899$

∴ Gini index = 0.4899

Attribute 1: Outlook.

values = {rain, sonny, Overcast}

Srain = [1+, 2-]

Gini(rain) = $1 - \left(\frac{1}{3}\right)^2 - \left(\frac{2}{3}\right)^2 = 0.4444$

∴ Gini(sunny) = [2+, 0-]

∴ Gini (sunny) = 0

→ Gini (Overcast) = [1+, 1-] = $1 - \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2 = 0.5$

∴ Gini index (Outlook) = $\frac{3}{7}$ (Gini(rain)) + $\frac{2}{7}$ * Gini (sunny) + $\frac{2}{7}$ (Gini (Overcast))

$= \frac{3}{7}(0.4444) + \frac{2}{7}(0) + \frac{2}{7}(0.5)$

$= 0.3333$

Attribute 2: wind

values = {strong, weak}

∴ Gini [index] (strong) = [1+, 3-]

$= 1 - \left(\frac{1}{4}\right)^2 - \left(\frac{3}{4}\right)^2 = 1 - 0.0625 - 0.$

$= 0.375$

: Gini(weak) = [3+, 0-] = 0

∴ Gini (wind) = $\frac{4}{7}$ * gini(strong) + $\frac{3}{7}$ (Gini (weak))

$= \frac{4}{7}(0.375) + \frac{3}{7}(0)$

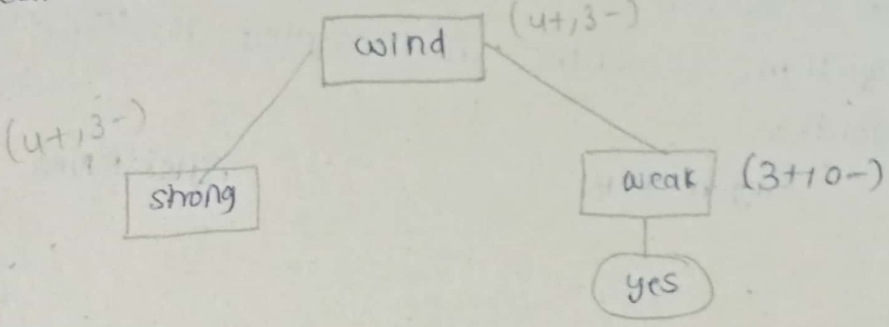
$= 0.2142$

from two attributes;

Gini (Outlook) = 0.3333

Gini (wind) = 0.2142

∴ here Gini(wind) is smaller. hence we consider the wind as root node.



Outlook	wind	play golf
Rain	strong	no
sonny	strong	yes
Rain	strong	no
Overcast	strong	no

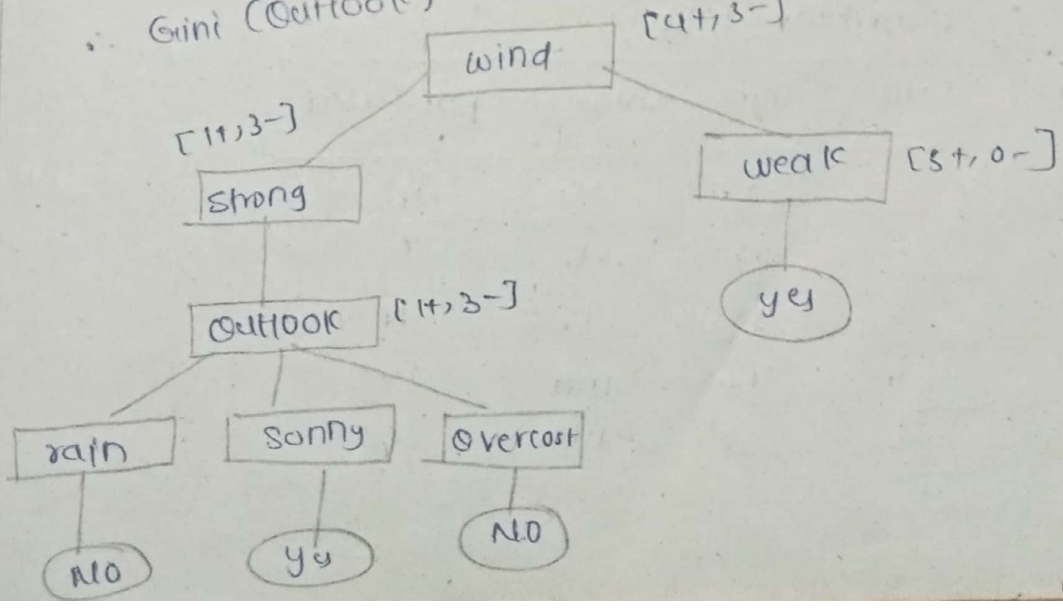
Attribute 1: Outlook, Value = {rain, sunny, Overcast}

Gini (rain) = [0+, 2-] = 0

Gini (sunny) = [1+, 0-] = 0

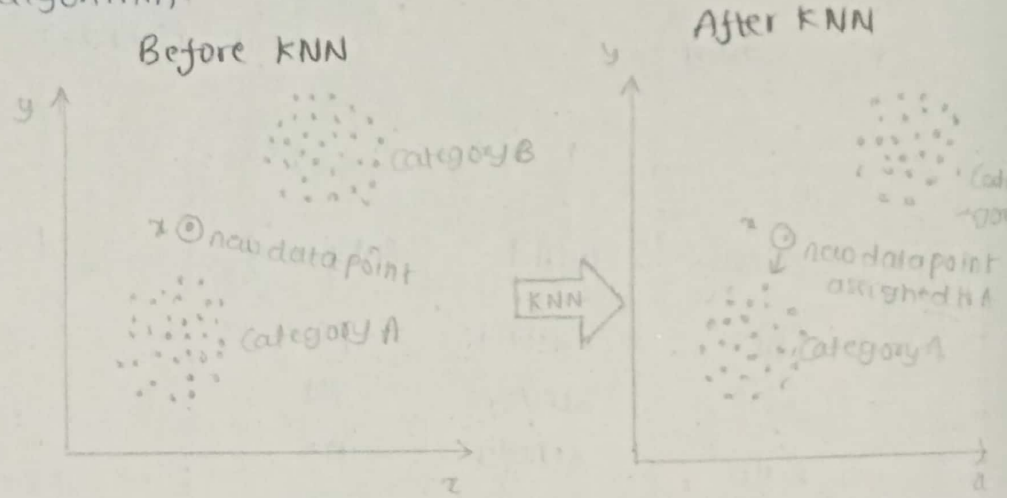
Gini (Overcast) = [0+, 1-] = 0

∴ Gini (Outlook) = 0



⇒ K- Nearest Neighbours

* K- Nearest Neighbour is one of the simplest machi learning algorithm based on supervised learning technique.
* It can be used in many applications.
* It stores all of the training data and classify the new data "x" by finding the training data $(x_i, y_i$
* The classifies the data into a category that is much similar to the new data.
* KNN algorithm is also called as non-parametric algorithm, sometimes it is called as lazy-learner algorithm.



* There are two points category A and category B and we have a new datapoint. So, this data point will lies in which of these category.
* TO solve this type of problem we can use KNN algorithm
* with the help of KNN, we can easily identify the class of particular dataset.

(Y)

SNO	age	salary	purchased
1	25	20	N
2	63	120	Y
3	33	75	N
4	42	100	Y
5	50	140	?

sol: First calculate the distance between test instance and training instance using euclidian distance, K = 3

SNO	age (x1)	salary (y1)	purchased	Distance
1	25	20	N	122.57
2	63	120	Y	23.85
3	33	75	N	67.186
4	42	100	Y	40.79
5	50	140	? (N)	

1. $(x_1, y_1) = (25, 20)$ and $(x_2, y_2) = (50, 140)$

By using euclidian distance = $\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$

$= \sqrt{(50-25)^2 + (140-20)^2} = \sqrt{625 + 14400} = 45.4422$
$= 122.5765$

2. $(x_1, y_1) = (63, 120)$ and $(x_2, y_2) = (50, 140)$

Euclidian distance = $\sqrt{(50-63)^2 + (140-120)^2}$

$= \sqrt{169 + 400} = 23.8537$

3) $(x_1, y_1) = (33, 75)$ and $(x_2, y_2) = (50, 140)$

Euclidian distance = $\sqrt{(50-33)^2 + (140-75)^2} = \sqrt{289 + 4225}$

$= 67.1863$

4) $(x_1, y_1) = (42, 100)$ and $(x_2, y_2) = (50, 140)$

∴ Euclidian distance = $\sqrt{(50-42)^2 + (140-100)^2}$

$= \sqrt{64 + 1600} = 40.7921$

⇒ Without giving 'k' value:-

Height	weight (w)	Class	Distance
167	51	Underweight	6.70
182	62	Normal	13
176	69	Normal	13.41
173	64	Normal	7.61
172	65	Normal	8.24
174	51	Underweigth	4.12
169	58	Normal	1.41
173	57	Normal	3
170	55	Normal	2
170	58	?	0

1) $(x_1, y_1) = (167, 51)$ & $(x_2, y_2) = (170, 57)$

∴ Euclidian distance $= \sqrt{(170-167)^2 + (57-51)^2} = \sqrt{9+36}$

$= \sqrt{45} = 6.708$

2) $(x_1, y_1) = (182, 62)$ and $(x_2, y_2) = (170, 57)$

∴ Euclidian distance $= \sqrt{(170-182)^2 + (57-62)^2} = 13$

3) $(x_1, y_1) = 176, 69)$ and $(x_2, y_2) = (170, 57)$

∴ Euclidian distance $= \sqrt{(170-176)^2 + (57-69)^2} = 13.41$

4) $(x_1, y_1) = (173, 64)$ and $(x_2, y_2) = (170, 57)$

Euclidian distance $= \sqrt{(170-173)^2 + (57-64)^2} = 7.61$

5) Euclidian distance $= \sqrt{(170-172)^2 + (57-65)^2} = 8.24$

6. $(x_1, y_1) = (174, 56)$

∴ Euclidian distance $= \sqrt{(170-174)^2 + (57-56)^2} = 4.123$

7) $(x_1, y_1) = (169, 58)$

∴ Euclidian distance $= \sqrt{(170-169)^2 + (57-58)^2} = 1.41$

8) $(x_1, y_1) = (173, 57)$

Euclidian distance $= \sqrt{(170-173)^2 + (57-57)^2} = 3$

9) $(x_1, y_1) = (170, 55)$

Euclidian distance $= \sqrt{(170-170)^2 + (57-55)^2} = 2$

10) $(x_1, y_1) = (170, 57)$

Euclidian distance $= \sqrt{(170-170)^2 + (57-57)^2} = 0$

⇒ Naive Bayers Algorithm :-

- Naive Bayes algorithm is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems.
- It is mainly used in text classification that includes a high dimensional training data set.
- Naive Bayes classifier is one of the simple and most effective classification algorithm.

* It is a probabilistic classifier, which means it predicts on the basis of the probability of an object.
* Examples: Spam filters, Sentimental analysis and text classification, article categorization etc.

- Naive: It assumes that the occurence of a certain features is independent of the occurence of other features.
- Bayes: It depends on the principle of bayes theorem.
- Bayes theorem (or) Bayes rule (or) Bayes law:

Equation: $P(A/B) = \frac{P(B/A) * P(A)}{P(B)}$

here $P(A/B)$ = posterior probability

$P(B/A)$ = likelihood probability

$P(A)$ = prior probability

$P(B)$ =

- Advantages:

* It can be used for binary as well as multiclass classification
* It is the most popular choice for text classification problems.
* It is one of the fast and easy machine learning algorithm
* It performs well in multiclass predictions as compared to the other algorithms.

- Disadvantages:

* It assumes that all of features are independent or unrelated
ex: Estimate conditional probability of each attribute { color, legs, height, smile } for species classes: { m, H } using the data given in table.
Using these probability estimate the probability value for new instance
{ color: green, legs: 2, height: tall, smile: NO }

SNO	Color	legs	height	smile	Species
1	White	3	short	yes	M
2	Green	2	tall	NO	M
3	Green	3	short	yes	M
4	White	3	short	yes	M
5	Green	2	short	NO	H
6	white	2	tall	NO	H
7	White	2	tall	NO	H
8	White	2	short	yes	H

Sol:- First we try to [solution] Calculate conditional probabilit as well as prior [propaba] probability to up with probability of new instances.

new instance : {color : green; leg : 2, height : tall; smile : NO}

$P(M) = \frac{4}{8} = 1/2 = 0.5$

$P(H) = \frac{4}{8} = 1/2 = 0.5$

We need to calculate conditional probability of each these attributes

- First attribute : color (white, Green)

color	M	H
White	2/4	3/4
Green	2/4	1/4

→ color (white, m) = 2/4 | color (Green, m) = 2/4

color (white, H) = 3/4 | Color (Green, H) = 1/4

→ 2nd attribute : legs (3, 2)

legs	m	H
3	3/4	0/4
2	1/4	4/4

- 3rd attribute : height (short, tall)

Height	M	H
short	3/4	2/4
tall	1/4	2/4

, 4th attribute: Smile (Yes, NO)

Smile	M	H
Yes	3/4	1/4
no	1/4	3/4

- we need to calculate prior probability:

new instances: {color: green, leg: 2, height: tall, smile: NO}

$\therefore P(M/\text{newinstances}) = P(m) * P(\text{color} = \text{green}) * P(\text{leg} = 2) * P(\text{height} = \text{tall}) * P(\text{smile} = NO)$

$= 0.5 * \frac{2}{4} * 1/4 * \frac{1}{4} * \frac{1}{4}$

$= 0.0039$

$\therefore P(H/\text{newinstance}) = P(H) * P(\text{color} = \text{green}) * P(\text{leg} = 2) * P(\text{height} = \text{tall}) * P(\text{smile} = NO)$

$= 0.5 * \frac{1}{4} * 4/4 * \frac{2}{4} * 3/4$

$= 0.04687$

$\therefore \boxed{P(H/\text{newinstance}) = 0.04687}$

⇒ polynomial Regression:

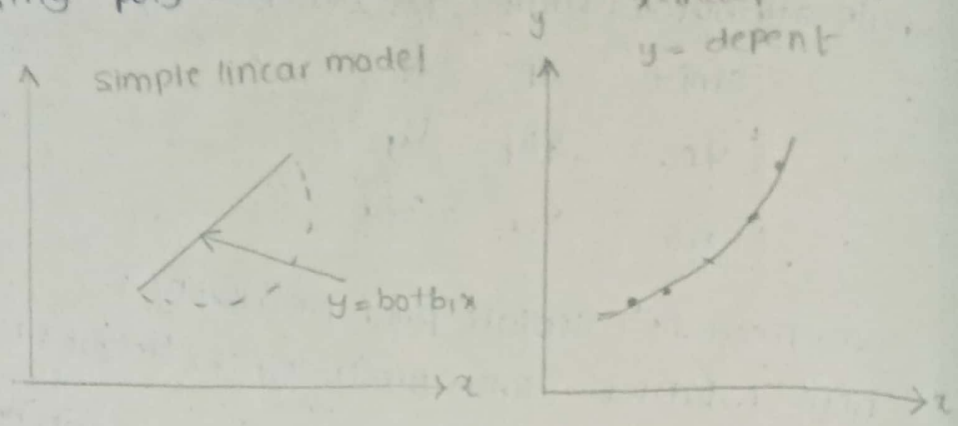
* polynomial regression is a type of regression analysis and used on statistics and machine learning
* the relationship b/w independent variable 'x' and dependent variable 'y' is modelled by the relation

$y = a_0 + a_1 x + a_2 x_2^2 + a_3 x_3^3 - - -$

* the relationship between x and y are linear then linear regression
* If the relationship between x and y are not linear then linear expression can't be used as it will result in large errors
* the problem of non linear regression can be solved by two

Methods:

1) Transformation of non-linear to linear data so that the linear regression can handle the data.
2) Using polynomial regression.



- Transformation:

* This convert non-linear data to linear data that can be handle using linear regression method.
* let us consider an exponential function $y = a e^{bx}$
* to avoid that large errors
* The transformation can be done by applying function on both sides to get:

$$\ln(y) = \ln(a) + \ln(e^{bx})$$
$$\ln(y) = \ln(a) + bx * \ln(e) \quad (\because \log e = 1)$$
$$\ln(y) = \ln(a) + bx \rightarrow \text{linear function.}$$

- polynomial Regression:

* It can handle non-linear relationship among variables by using n^{th} degree of polynomial
* For example, the second-degree polynomial is given as $y = a_0 + a_1 x + a_2 x_1^2$ and third-degree polynomial is called cubic transformation given as

$$y = a_0 + a_1 x + a_2 x_2^2 + a_3 x_3^3$$

* Generally polynomial of maximum degree 4 are used as higher order polynomials take some different (or) strange shapes.
* Consider the polynomial of 2nd degree,
* Then polynomial equation (or) given by

$$y = a_0 + a_1 x + a_2 x_1^2$$

* The coefficient of a_0, a_1 and a_2 are calculating using the formula.

$\therefore a = X^{-1}B$

where

$$X = \begin{bmatrix} n & \Sigma x_i & \Sigma x_i^2 \\ \Sigma x_i & \Sigma x_i^2 & \Sigma x_i^3 \\ \Sigma x_i^2 & \Sigma x_i^3 & \Sigma x_i^4 \end{bmatrix} \qquad B = \begin{bmatrix} \Sigma y_i \\ \Sigma (x_i - y_i) \\ \Sigma [x_i^2 - y_i) \end{bmatrix}$$

Eg:

independent(x)	dependent(y)
1	1
2	4
3	9
4	15

x_i	y_i	$x_i y_i$	x_i^2	$x_i^2 y_i$	x_i^3	x_i^4
1	1	1	1	1	1	1
2	4	8	4	16	8	16
3	9	27	9	81	27	81
4	15	60	16	240	64	256
10	29	96	30	338	100	354

$$\therefore \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 4 & 10 & 30 \\ 10 & 30 & 100 \\ 30 & 100 & 354 \end{bmatrix}^{-1} \qquad X = \begin{bmatrix} 29 \\ 96 \\ 338 \end{bmatrix}$$

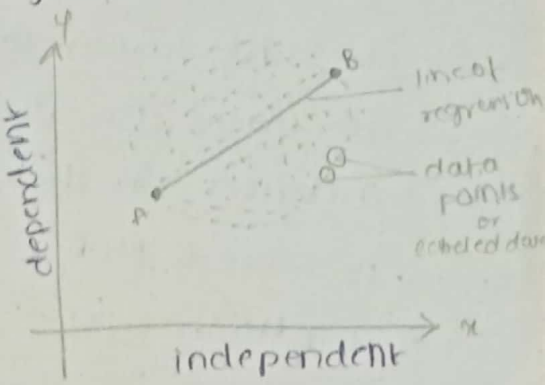
$$= \begin{bmatrix} -0.75 \\ 0.95 \\ 0.75 \end{bmatrix}$$

Final polynomial regression

Linear Regression :-

* linear regression is the most one of the most easiest and popular machine learning.
* It is a statistical method that is used for predictive analysis.
* Linear regression makes prediction for continuous and numeric variables such as sales, temperature, price etc..
* LR algorithm shows a linear relationship between dependent variables and one or more independent variables; hence it is called a linear regression.
* LR is a mathematical representation :

$$y = a_0 + a_1 x + \epsilon$$



here a_0 = intercepts of the line

x, y = variables / datapoints

a_1 = coefficient of line of regression

ϵ = random error.

* There are two types of LR :
  1. simple LR
  2. multiple linear regression.

1. Simple linear regression : if a single independent variable is used to predict the value of a numeric dependent variable, then such a LR algorithm is called SLR (simple linear regression).
2. multiple linear regression : if more than one independent variable is used to predict the value of a numerical dependent variable, then such a LR algorithm is called MLR.

Linear regression line :

A linear line showing the relationship between the independent and dependent variables is called a regression line.

there are two types of regression lines:

1. positive linear relationship
2. Negative linear relationship

1. positive linear relationship: If the dependent variable is increased on the y-axis and independent variable increases on x-axis, then such as relationship is called PLR

[ii, Negative linear]

* Equation : $y = a_0 + a_1 x$

2. Negative linear relationship: If the dependent variable decreases on the y-axis and independent variable increases on the x-axis, then such a relationship is called NLR.

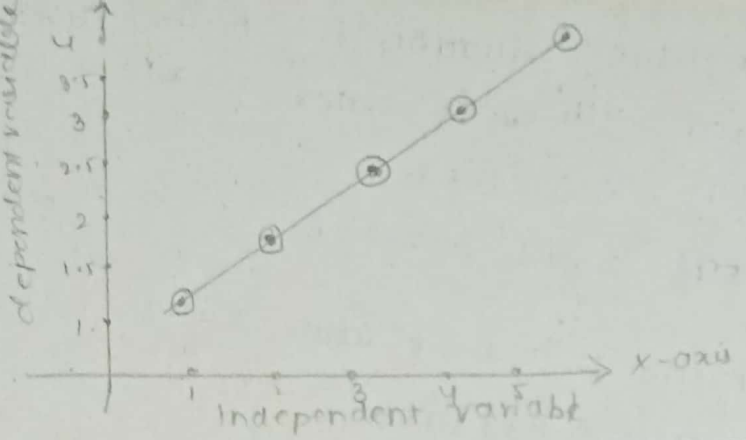
* Equation : $y = -(a_0) + (a_1 x) = -a_0 + a_1 x$

example; let us consider the 5 weeks sales data is given below.

x_i(week)	y_i(sales in thousands)
1	1.2
2	1.8
3	2.6
4	3.2
5	3.8

Apply linear regression technique to predict the seventh & 12th week sales.

sol: In linear regression algorithm,
we plot the independent variable and dependent variable
independent variable is represented as x-axis
dependent variable is represented as y-axis



Linear regression equation ; $y = a_0 + a_1 x + \epsilon$

$a_0 = \bar{y} - a_1 \bar{x}$

$a_1 = \dfrac{\overline{xy} - (\bar{x})(\bar{y})}{(\bar{x})^2 - \bar{x}^2}$

Here there are 5 items : 1, 2, 3, 4, 5

x_i	y_i	x^2	xy
1	1.2	1	1.2
2	1.8	4	3.6
3	2.6	9	7.8
4	3.2	16	12.8
5	3.8	25	19
15	12.6	55	44.4

$\bar{x} = \dfrac{\Sigma x_i}{n} = \dfrac{15}{5} = 3$

$\bar{y} = \dfrac{12.6}{5} = 2.5$

$\bar{x}^2 = \dfrac{\Sigma x^2}{n} = \dfrac{55}{5} = 11$

$[\bar{x}^2 = \dfrac{55}{5} = 11]$

$\overline{xy} = \dfrac{\Sigma xy}{n} = \dfrac{44.4}{5} = 8.88$

$a_0 = \bar{y} - a_1 \bar{x}$

$= 2.5 - a_1(3)$

$a_1 = \dfrac{\overline{xy} - (\bar{x})(\bar{y})}{\bar{x}^2 - \bar{x}^2} = \dfrac{8.88 - (3)(2.5)}{11 - 55}$

$= \dfrac{1.38}{-44} = -0.031$

$\therefore a_0 = 2.5 - (-0.031)(3) = 2.593$

a linear regression equation is $y = a_0 + a_1 x + \varepsilon$

the predicted 7th week sales $x = 7$

$y = 2.593 - 0.031(7)$

$= 2.593 - 0.217 = 2.376$

to predict the 12th week sales $x = 12$

$y = 2.593 - 0.031(12)$

$= 2.593 - 0.372 = 2.221$

∴ 7th week sales is a 2.376 and 12th week sales is a 2.221

2. multi linear Regression:

* A single linear regression model involves one independent variable and one dependent variable.
* In multiple linear regression model involves one or more independent variables and one dependent variable.
* This is the extension of the linear regression equation: $y = a_0 + a_1 x_1 + a_2 x_2 + \varepsilon$

Here x_1, x_2 = independent variables

y = dependent variable

a_0 = inception of line

a_2, a_1 = coefficient of line of regression

ε = random error.

example: Let us consider weekly sales along with sales for products x_1, x_2 are provided.

x_1 (product1 sales)	x_2 product 2 sales	weekly sales (y)
1	4	1
2	5	6
3	8	8
4	2	12

How to find the multiple regression for the values.

x_1, x_2 are two independent variables, y is a dependent variable. The value of y depends on x_1 and x_2 in the particular case.

We use matrix approach for finding multiple regression. The matrix for x and y are given as follows:

$$X = \begin{bmatrix} 1 & 1 & 4 \\ 1 & 2 & 5 \\ 1 & 3 & 8 \\ 1 & 4 & 2 \end{bmatrix}, \quad Y = \begin{bmatrix} 1 \\ 6 \\ 8 \\ 12 \end{bmatrix}$$

The coefficient of the multiple regression equation is given

$$\therefore a = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix}$$

The regression coefficient for multiple regression is calculated

$$\therefore \hat{a} = ((x^T x)^{-1} x^T) Y$$

Here $x^T = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 4 & 5 & 8 & 2 \end{bmatrix}$

$$x^T \cdot x = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 4 & 5 & 8 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 & 4 \\ 1 & 2 & 5 \\ 1 & 3 & 8 \\ 1 & 4 & 2 \end{bmatrix}$$

$$= \begin{bmatrix} 1(1)+1(1)+1(1)+1(1) & 1(1)+1(2)+1(3)+1(4) & 1(4)+1(5)+1(8)+1(2) \\ 1(1)+2(1)+3(1)+4(1) & 1(1)+2(2)+3(3)+4(4) & 1(4)+2(5)+3(8)+4(2) \\ 4(1)+5(1)+8(1)+2(1) & 4(1)+5(2)+8(3)+2(4) & 4(4)+5(5)+8(8)+2(2) \end{bmatrix}$$

$$x^T \cdot x = \begin{bmatrix} 4 & 10 & 19 \\ 10 & 30 & 46 \\ 19 & 46 & 109 \end{bmatrix}$$

$$\therefore (x^T \cdot x)^{-1} = \frac{\text{Adj } A}{\det A}$$

$$\rightarrow \det A = 4(30 \times 109 - 46 \times 46) - 10(10 \times 109 - 46 \times 19) + 19(10 \times 46 - 30 \times 19)$$

$$= 4616 - 2160 - 2090$$

$\therefore \det A = 366$

$\therefore$ Adj A :

$$\det A = \begin{bmatrix} 4 & 10 & 19 \\ 10 & 30 & 46 \\ 19 & 46 & 109 \end{bmatrix}$$

$$A_1 = +4\begin{bmatrix} 30 & 46 \\ 46 & 109 \end{bmatrix} = 4[30\times109 - 46\times46]$$
$$= 4(1,154) = [4616]$$

$$B_1 = -10\begin{bmatrix} 10 & 46 \\ 19 & 109 \end{bmatrix} = -10[10\times109 - 46\times19] = -2160$$
$$= -216$$

$$C_1 = 19\begin{bmatrix} 10 & 30 \\ 19 & 46 \end{bmatrix} = +19[10\times46 - 30\times19] = +110$$
$$= -110$$

$$A_2 = -10\begin{bmatrix} 10 & 19 \\ 46 & 109 \end{bmatrix} = -10[10\times109 - 19\times46] = +216$$

$$B_2 = +30\begin{bmatrix} 4 & 19 \\ 19 & 109 \end{bmatrix} = 30[4\times109 - 19\times19] = 75$$

$$C_2 = -46\begin{bmatrix} 4 & 10 \\ 19 & 46 \end{bmatrix} = -46[4\times46 - 10\times19] = -[-6] = 6$$
$$= (-6)$$
$$= 6$$

$$A_3 = +\begin{bmatrix} 10 & 19 \\ 30 & 46 \end{bmatrix} = +[10\times46 - 19\times30] = +110$$

$$B_3 = -\begin{bmatrix} 4 & 19 \\ 10 & 46 \end{bmatrix} = -[4\times46 - 19\times10] = -[-6] = 6$$

$$C_3 = +\begin{bmatrix} 4 & 10 \\ 10 & 30 \end{bmatrix} = +[4\times30 - 10\times10] = 20$$

$$\text{adj } A = \begin{bmatrix} 1154 & +216 & +110 \\ -216 & 75 & +6 \\ -110 & +6 & 20 \end{bmatrix} \times \frac{1}{366}$$

$$\frac{adjA}{detA} = \begin{bmatrix} 1154 & -216 & -110 \\ -216 & 75 & 46 \\ -110 & 6 & 20 \end{bmatrix} \Big/ 366$$

$$\frac{adjA}{detA} = \begin{bmatrix} 3.1530 & -0.590 & -0.3005 \\ -0.590 & 0.204 & 0.0163 \\ -0.300 & 0.01 & 0.05 \end{bmatrix}$$

$$\therefore \frac{adjA}{detA} = \begin{bmatrix} 3.1530 & -0.590 & -0.3005 \\ -0.590 & 0.204 & 0.0163 \\ -0.300 & 0.01 & 0.05 \end{bmatrix}$$

$$(X^T \cdot X)^{-1} = \begin{bmatrix} 3.15 & -0.59 & -0.30 \\ -0.59 & 0.20 & 0.01 \\ -0.30 & 0.01 & 0.05 \end{bmatrix}$$

$$\therefore [X^T \cdot X]^{-1} X^T = \begin{bmatrix} 3.15 & -0.59 & -0.30 \\ -0.59 & 0.20 & 0.01 \\ -0.30 & 0.01 & 0.05 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 4 & 5 & 8 & 2 \end{bmatrix}$$

$$= \begin{bmatrix} 3.15(1)+(0.59)(1)+(-0.30)(4) & 3.15(1)+(0.59)(2)+(-0.30)(5) & \cdots \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \end{bmatrix}$$

$$= \begin{bmatrix} 1.36 & 0.47 & -1.02 & 0.19 \\ -0.35 & -0.14 & 0.09 & 0.23 \\ -0.09 & -0.03 & 0.13 & -0.16 \end{bmatrix}$$

$$\therefore [[X^T \cdot X]^{-1} X^T] Y = \begin{bmatrix} 1.36 & 0.47 & -1.02 & 0.19 \\ -0.35 & -0.14 & 0.09 & 0.23 \\ -0.09 & -0.03 & 0.13 & -0.16 \end{bmatrix} \begin{bmatrix} 1 \\ 6 \\ 8 \\ 12 \end{bmatrix}$$

$$= \begin{bmatrix} 1.36 + 2.82 - 8.16 + 2.28 \\ -0.35 + (-0.84) + 0.72 + 2.76 \\ -0.09 + (-0.18) + 1.04 - 1.92 \end{bmatrix} = \begin{bmatrix} -1.7 \\ 2.29 \\ -1.15 \end{bmatrix}$$

$$\therefore y = a_0 + a_1 x_1 + a_2 x_2$$

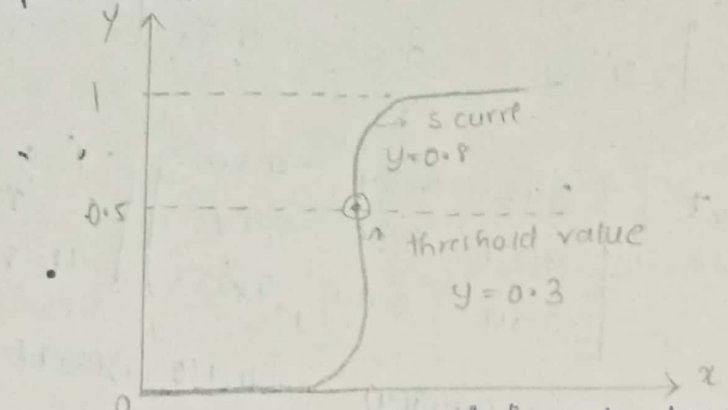
$$= (-1.7) + 2.29(1) + (-1.15)(4)$$

$$= -1.7 + 2.29 - 4.6 = -4.01$$

$$\boxed{y = -4.01}$$

⇒ Logistic Regression:

* Logistic regression is a supervised machine learning algorithm used for classification tasks (true or false, yes or No, + or -)
* where the goal is to predict the probability than an instance belongs to given class or not
* logistic regression is a statistical algorithm which analyse the relationship between two data factors.
* logistic regression is used for binary classification where we use sigmoid function (or) logistic function.
* It takes input as independent variables and produces a probability value between '0' and '1'.



* We have two classes class "0" and class "1".
* If the value of the logistic function for an input is greater than 0.5.
* Then it belongs to class "1" otherwise it belongs to class "0".
* The sigmoid function is a mathematical function used to map the predicted values to probabilities
* It maps any real value into another value within a range of "0" and "1".
* In logistic regression, instead of fitting a regression line, we fit an "S" shaped logistic function, which predicts two maximum value (0 or 1)

sigmoid function formula:

$$S(x) = \frac{1}{1+e^{-x}}$$

simple linear and multiple linear equations

$$Z = a_0 + a_1x_1 + a_2x_2 + \cdots$$

logistic regression equation:

$$h = S(x) = \frac{1}{1+e^{-z}}$$

main aim: logistic regression is used when the dependent variable is categorical value (or) discrete value.

ex: to predict whether an email is spam "1".

- Types of logistic regression:

There are three(3) types of logistic regression:

1. Binomial
2. multinomial
3. Ordinal

1. Binomial: There can be only two(2) possible types of the dependent variables, such as "0" or "1", pass or fail

2. multinomial:

There can be three(3) or more possible unordered types of the dependent variable, such as "cat", "dog" or "sheep",

3. Ordinal:

there can be three(3) or more possible ordered types of dependent variables, such as "low", "medium", or "high".

Q Difference between linear regression and logistic regression.

Feature	Linear regression	logistic regression
purpose	used for predicting continuous values	Used for classification (categorical output)
output type	produces continuous numerical values	produces probabilities mapped to classes (0/1)
mathematical Function	Uses a linear equation $y = mx + c$	Uses a sigmoid function $S(x) = \frac{1}{1+e^{-z}}$

Type of problem	regression problems (eg: predicting sales, temperatures)	classification problems (eg: spam detection, fraud detection.
Algorithm type	Based on least squares method	Based on maximum likelihood estimation.
Linearity	assumes a linear relationship between input and output	Does not assume a linear relationship; applies a non-linear transformation (sigmoid function)
Example Use cases	predicting house prices, stock prices etc	Email spam detection, disease diagnosis (yes/no)