Deep learning

Deep learning is a type of Machine learning and artificial Intelligence (AI) that immitates the way human gain certain type of knowledge while traditional machine learning algorithms are linear, deep learning algorithms are stacked in a hierarchy of Increasing Complexity and abstraction.

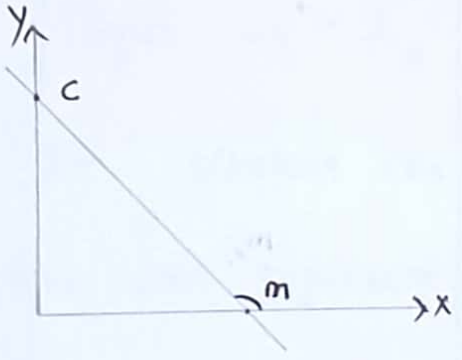
Deep Learning Utilizes both Structured and Unstructured data of training, In practically we are used in virtual assistance, vision for driverless cars, money laundering, face recognization and more.

Pre Requisites :- Maths

As we learned in machine learning algorithms, Straight line eqn.

Every straight line $\Rightarrow$ $y = mx + c$ $\rightarrow$ Constant

$\downarrow$ slope



$$\text{Slope } (m) = \frac{\text{change in Y value}}{\text{change in X value}}$$

$$m = \frac{\Delta y}{\Delta x}$$

Where if $m = 0 \Rightarrow y = 0 \cdot x + c \Rightarrow y = c$

Input, if $x = 0 \Rightarrow y = m \times 0 + c \Rightarrow y = c$

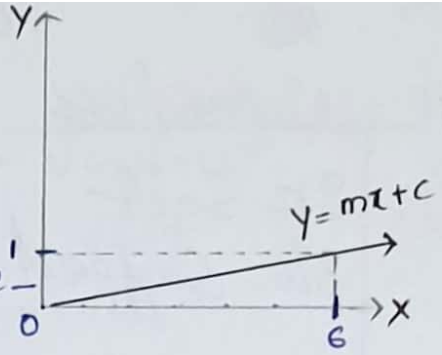
Thus we can say $mx = 0 \Rightarrow y = c$ is referred as 'Y' Intercept.

Ex : $m = 0,\ x = 6,\ y = 1$

$y = mx + c$

$1 = 0 + c$

$c = 1$, So the line Eqn will be



Similarly in Neural Networks we represent $y = mx + c$ as $W.I + b$

where W → Weight

I → Input → May be Image or text data

b → Bias (Some what)

Activation function :- it is a function, that added into an artificial neural network in Order to help the network learn Complex patterns in the data.

$$\alpha(W.I + b)$$

Here — α → Activation function.

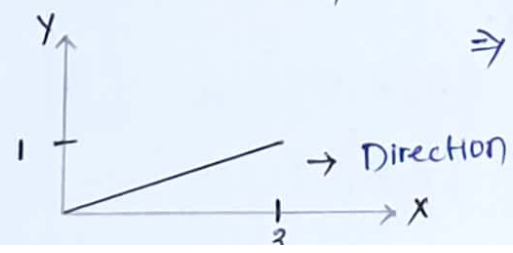
Generally the values lies in between (0-1) in activation function.

Input → I_0^1

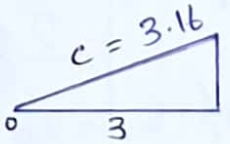
Vector :- Vectors are Used in NLP / NLU

Vectors are represent direction and magnitude

Ex ① $\vec{V} = [3, 1]$ (x, y)



⇒ hypotenuse



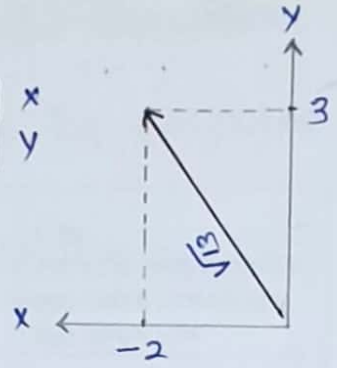
⇒ $c = \sqrt{3^2 + 1^2}$

$= \sqrt{9+1}$

$= 3.16$

Ex -②

$$\vec{u} = \begin{bmatrix} -2 \\ 3 \end{bmatrix} \begin{matrix} x \\ y \end{matrix}$$



$$\Rightarrow \quad c = \sqrt{(-2)^2 + (3)^2}$$

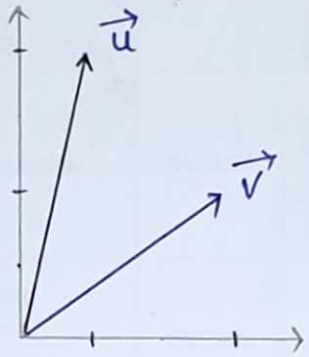
$$= \sqrt{4+9}$$

$$c = \sqrt{13}\ //$$

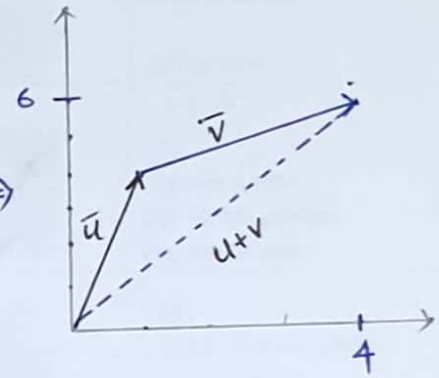
Vector Addition :-

$$\vec{v} = [3, 2]$$

$$\vec{u} = [1, 4] \quad \Rightarrow$$



$$\Rightarrow$$



(u+v)

$$v+u = \begin{bmatrix} 3 \\ 2 \end{bmatrix} + \begin{bmatrix} 1 \\ 4 \end{bmatrix}$$

$$= \begin{bmatrix} 4 \\ 6 \end{bmatrix}.$$

$v+u \Rightarrow$

Vector Multiplication :-

Scalar Vector multiplication :-

Where multiply by Scalar value to the vector.

$$\overline{v} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \Rightarrow 2.\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$$

↳ Scalar value

Dot product :- $u.v = u_1v_1 + u_2v_2 + -- + u_mv_m.$

Ex: $u = [2, 4, 6]$

$v = [1, 3, 5]$

$\Rightarrow u.v = 2\times1 + 4\times3 + 6\times5$

$= 2 + 12 + 30$

$= 44\ //$

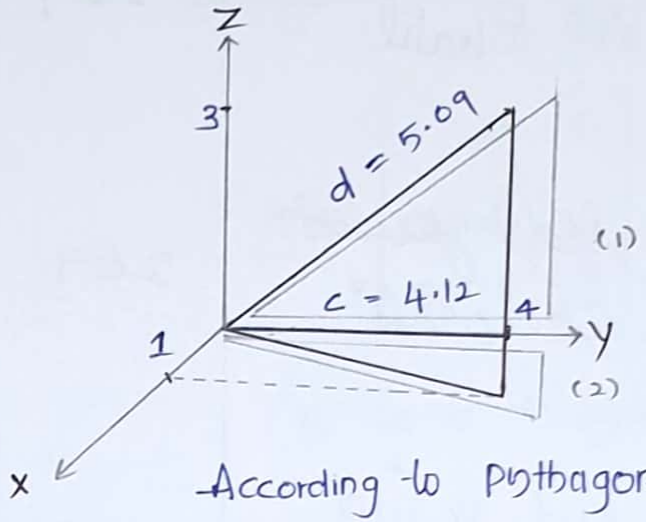
Note :- if the dimentions are not matching then dot Product can not be computed.

Properties of vector :-

1) $u+v = v+u$
2) $(u+v)+w = u+(v+w)$
3) $u+0 = u$
4) $u+(-u) = 0$
5) $c[u+v] = c.u+c.v.$

3-Dimention Vector :-

$$\overline{V} = \begin{bmatrix} 1 \\ 4 \\ 3 \end{bmatrix} \begin{matrix} x \\ y \\ z \end{matrix} \Rightarrow$$



Here we devided this in-to 2 right angle triangles.

Intially we find the 'c' value

Then 'd' value

According to Pythagorean

$$c = \sqrt{4^2+1^2} = \sqrt{17} = 4.12$$

$$d = \sqrt{(4.12)^2+3^2}$$

$$d = 5.09$$

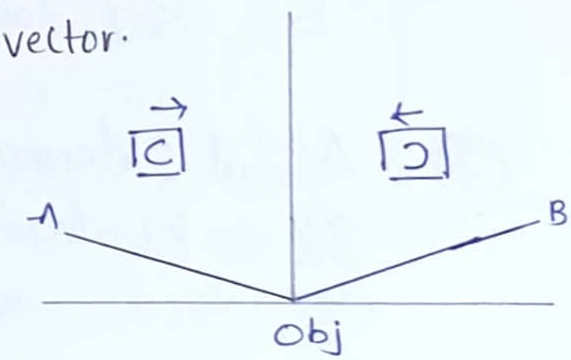
$$\|d\| = \sqrt{1^2+4^2+3^2} = \sqrt{26} = 5.09\,\|$$

i) if the Dot product between two vectors is zero then we represent that vector as Orthogonal vector.

$u = [4,3]$; $v = [-3,4]$

$u.v = 4(-3) + 3(4)$

$u.v = -12+12 = 0$ ‖ (Orthogonal)



Matrix :-

A set of numbers arranged in rows and columns. So as to form a rectangular array.

Types of matrix :-

1) Square matrix
2) Diagonal matrix
3) Scalar matrix
4) Identity matrix
5) Symmetric matrix
6) Skew Symmetric matrix
7) Orthogonal matrix

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

Row →

3 × 3

Column →

A is a Square matrix

$A \Rightarrow$ No of rows = No. of columns

$$B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

'B' is a Diagonal matrix

$$C = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

'C' is a Scalar matrix

$$D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

'D' is a Identity matrix

$$A = \begin{bmatrix} -1 & -1 & 2 \\ -1 & 4 & -2 \\ 2 & -2 & 3 \end{bmatrix} \qquad A^T = \begin{bmatrix} -1 & -1 & 2 \\ -1 & 4 & -2 \\ 2 & -2 & 3 \end{bmatrix}$$

When transpose a matrix There is no change.

Such matrix called as Symmetric matrix

$$A^T = A$$

The transpoce of a matrix, resultant in Negation Such matrix are called as Skew Symmetric matrix.

$$A^T = -A$$

Ex:

$$A = \begin{bmatrix} 0 & 1 & 2 \\ -1 & 0 & 4 \\ -2 & -4 & 0 \end{bmatrix} ; A^T = \begin{bmatrix} 0 & -1 & -2 \\ 1 & 0 & -4 \\ 2 & 4 & 0 \end{bmatrix}$$

Orthogonal matrix :- The multiplication of matrix "A" and transpoce of 'n' matrix, resultant would be Identical Such matrix called as Orthogonal matrix.

$$A^T . A = I$$

Matrix Multiplication :-

$$A = \underset{1\times3}{\begin{bmatrix} 1 & 4 & 3 \end{bmatrix}} \qquad B = \underset{3\times2}{\begin{bmatrix} 4 & 3 \\ 2 & 5 \\ 6 & 1 \end{bmatrix}}$$

Before going to multiplication process we have to check Dimentionality of an matrix

1) $A \times B = [1\times3][3\times2] = [1\times2]$ the result;

In form of $m\times n ; n\times p$ then only multiplicatin possible

2) $A \times B = [1\times3][2\times3] \Rightarrow 3 \neq 2$ Dimention Error.

3) $A = \begin{bmatrix} 3 & 4 \\ 7 & 2 \\ 5 & 1 \end{bmatrix}_{3\times 2}$ $\quad B = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}_{2\times 3}$

$A \cdot B$ = 3×3 as result.

$$A \cdot B = \begin{bmatrix} 3\times1+4\times4 & 3\times2+4\times5 & 3\times3+4\times6 \\ 7\times1+2\times4 & 7\times2+2\times5 & 7\times3+2\times6 \\ 5\times1+1\times4 & 5\times2+1\times5 & 5\times3+1\times6 \end{bmatrix}$$

$$= A \cdot B = \begin{bmatrix} 19 & 26 & 33 \\ 15 & 24 & 33 \\ 9 & 15 & 21 \end{bmatrix}_{3\times3}$$

Addition and Substraction of matrix :-

$$A, B = \begin{bmatrix} 1 & -2 \\ 3 & 4 \end{bmatrix}, \begin{bmatrix} 5 & -1 \\ 6 & -2 \end{bmatrix}$$

$$A+B = \begin{bmatrix} 1+5 & -2+(-1) \\ 3+6 & 6+(-2) \end{bmatrix} \quad ; \quad A-B = \begin{bmatrix} 1-5 & -2-(-1) \\ 3-6 & 4-(-2) \end{bmatrix}$$

$$A+B = \begin{bmatrix} 6 & -3 \\ 9 & 4 \end{bmatrix} \quad ; \quad A-B = \begin{bmatrix} -4 & -1 \\ -3 & 6 \end{bmatrix}$$

Determinant :- When division of matrix, we use determinant

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \Rightarrow \boxed{\det|A| = ad-bc}$$

$$A^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

$$B = \begin{bmatrix} 6 & -1 \\ 3 & -2 \end{bmatrix} \Rightarrow |B| = 6\cdot(-2) - (-1(3) = (-12) - (-3) = -9 //$$

Matrix Inverse :-

1) The Inverse Matrix (A^{-1}) is defined Such as -

$$A \cdot A^{-1} \text{ is "I"}$$

2) Not every matrix has an inverse, if no inverse exist Then the matrix is called as Singular (Non-invertible)

3) if a is non Singular, so is A^{-1}

4) if A, B, are non-Singular, then the AB is also non Singular and $(AB)^{-1} = B^{-1} \cdot A^{-1}$ (Reversed Order)

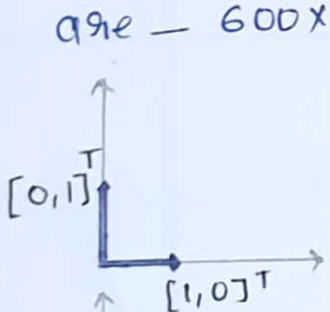
5) if A is non Singular, then so is its transpoce and

$$(A^T)^{-1} = (A^{-1})^T$$

Matrices : Scaling, Rotation, Identity :-

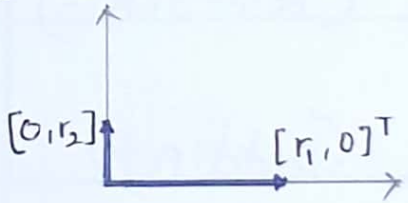
Matrices Scaling is more used in Image Scaling.

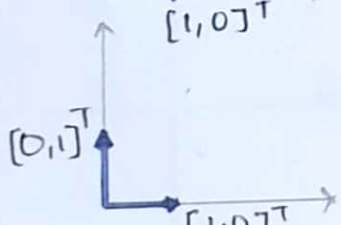
Ex: if we have an Image of high-<h> and width <w> has pixels size = 300 x 300 after Scaling the Size of pixels are - 600 x 600, This process called as Interpolation.



$$\begin{bmatrix} r_1 & 0 \\ 0 & r_2 \end{bmatrix}$$

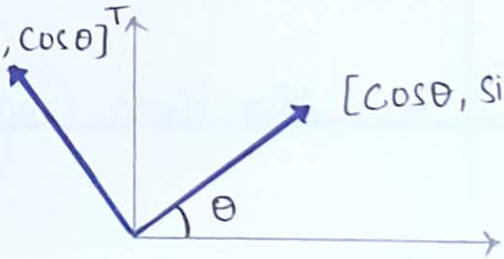
Scalling →





$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

Rotation →



Introduction to function :

When we building a models, we are essentially Creating a Surface, in mathematical Space.

The Surface could be a line, plane or hyperplane, these are used to do classification or to predict some number.

The Surface is called function and represented as "f"

$$y = f()$$

The Surfaces are called functions, are inturns called models.

Artificial Neural Networks :-

Neural Networks :- These are new method of programming Computers. These are exceptionally good at performing patten recognization and other tasks. that are very difficult to program using conventional Techniques.

An NN Configured for a Specific application such as pattern recognization or data classification. through learning process.

Learning in biological System involves adjustments to the Synaptic connections that exist between the neurons.

Deeplearning Used in wide range of applications. —

→ Automated feature Selection

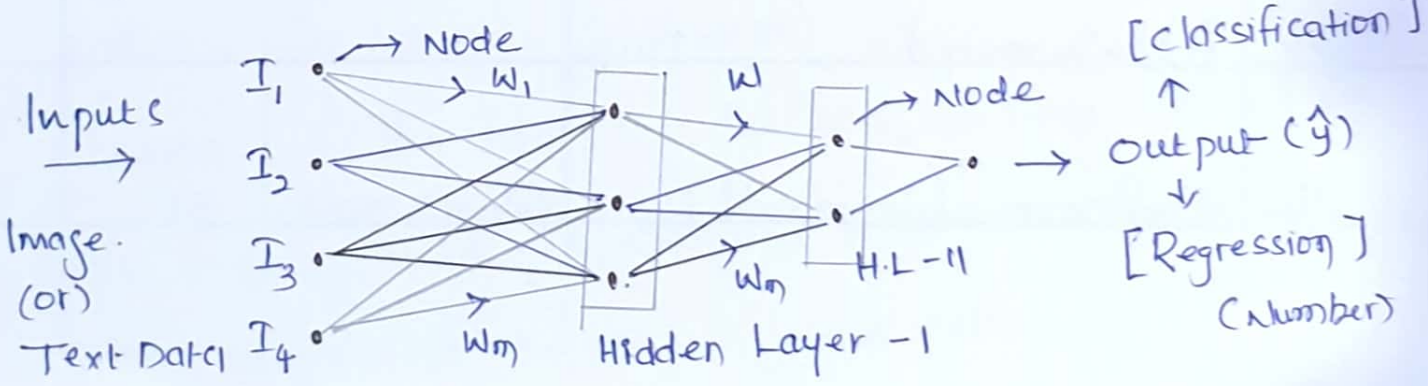
→ Architecture innovation and flexibility.

→ Scalable & Robust

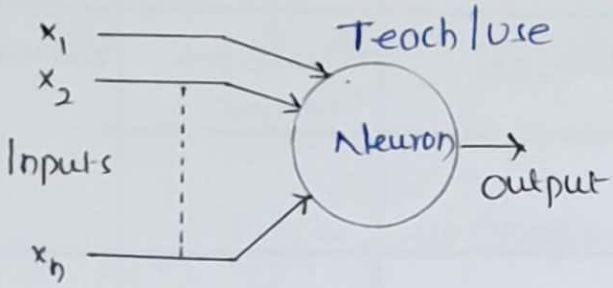
Building Components of Neural Network :—

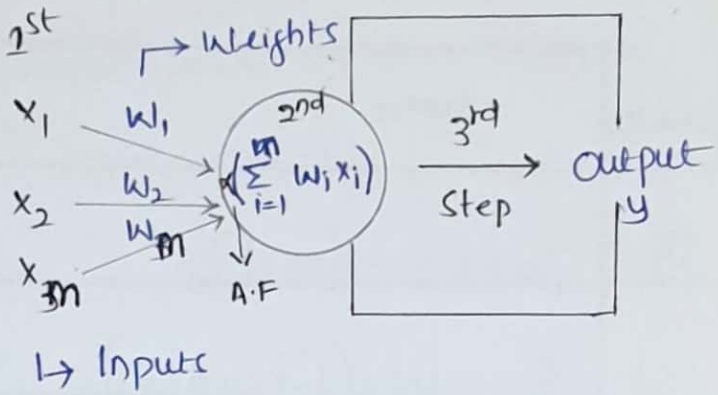
1) Layer 2) Nodes 3) Activation functions.

4) Optimizers 5) Back propogation.



A) Simple Neuron :-





An artificial neuron is a device with many inputs and outputs. its working functionality based on human neuron system like it is a biological process, this neurons also similarly activate it self like human brains.

The neuron has '2' mode of operations — ① Training mode
② Using mode.

In training mode the neuron can be trained to fire(or not) for particular input patterns.

In the using mode, when a thought input pattern is detected at the input, is associated output becomes the current O/P. if the i/p pattern doesnot belonging in the thought list of i/p patterns. the firing rule is used to determine whether fire or not.

Neuron :- The Neuron is the basic working unit of the brain, a specialized cell, designed to transmit information to other Nerve cells, muscle or gland cells.

A Simple Neural Network :-

As above we discuss about neural networks, now we have to focus on network layers mainly consisting of -

① input layer ② Hidden layer ③ Output layer.

Input Layer :- the activity of a input units represents the raw information that is fed into the network.

Hidden Layer :- the activity of each hidden unit is activates the input units and weights on the connections b/w the input and the hidden units.

Output Layer :- The behaviour of the Output units depends on the activity of the hidden units and the weights b/w the hidden and output units.

Network Structure :- the no of layers and neurons depend on the specific task. In practise this issue is solved by trial and error.

Two types of adaptive algorithms can be used.

i) Start from large network and successively remove some neurons and links until network performance degrades.

ii) Begin with small network and introduce new neurons untill performance is satisfactory.

Training of Neural Network :-

* The most basic method of training a neural network is trial and error.
* if the network isn't behaving the way it should, change the weighting of a random link by random amount. if the accuracy of a network declines, undo the change and make a different one.
* it takes a time but, trial and error method does produce result.

Network parameters :-

Network parameters depends on - ① weights intialization. ii) depends on hidden layers and neurons. iii) training sets.

Weights :- In general Intial weights are randomly choosen, with typical values b/w (-1.0 and 1.0) or (-0.5 and 0.5)

There are two types of Neural Networks - ① Fixed ② Adaptive

① In fixed Networks weights are fixed.

② In Adaptive Networks where the weights are changed. to reduce prediction error.

the no of training examples should be. atleast five to ten times the no of weights of the network.

Different types of Neural Networks :-

Feed forward Networks :-

Feed forward networks allows signals to travel one way only, from input to output, there is no-feedback (loops ←), ie the output of any layer doesn't affect the same layer.

this tend to be straight-forward. they are mostly used in pattern recognization.

Feedback Networks :-

This networks can have signal travelling in both directions by introducing loops in the network. these are dynamic thier state is changing continiously until they reach an equilibrium point. and these are single layer organizations.

Disadvantages of Neural Network :-

i) The individual relation between the input and output variables are not developed by engineering judgement so that the model tends to be a block box or input/output table without analytical basis.

ii) the sample size has to be large.

iii) Requires a lot of trial and error. So training can be time consuming.