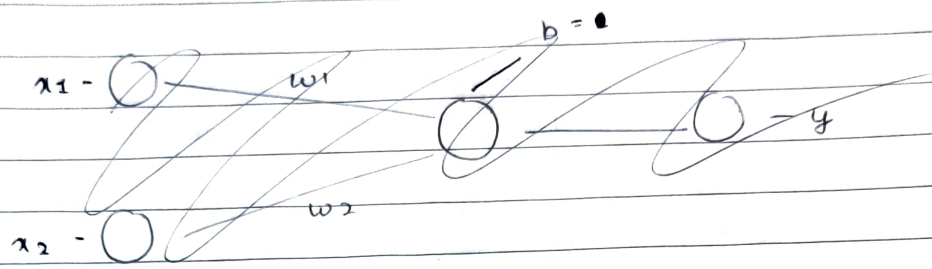
ML

Module - 4

Q. (Ex-1) Hebb Learning :

→



AND f^n			
x_1	x_2	b	y
1	1	1	1
-1	1	1	-1
1	-1	1	-1
-1	-1	1	-1

OR f^n			
x_1	x_2	b	y
1	1	1	-1
-1	1	1	1
1	-1	1	1
-1	-1	1	1

$w_1 = w_2 = b = 0$

$w(new) = w(old) + \Delta w$

$\Delta w = xy$

$\Delta b = y$

$b(new) = b(old) + \Delta b$

First I/P :

$[x_1 \quad x_2 \quad b] = [1 \quad 1 \quad 1]$

$y = 1$

$\Delta w_1 = x_1 y = 1 \times 1 = 1$

$\Delta w_2 = x_2 y = 1 \times 1 = 1$

$\Delta b = 1$

$w_1(new) = 0 + 1 = 1$

$w_2(new) = 0 + 1 = 1$

$new(b) = 0 + 1 = 1$

$new [1 \quad 1 \quad 1]$

Second I/P :-

$[x_1 \; x_2 \; b] = [-1 \quad 1 \quad 1]$

$y = -1$

$\Delta w_1 = x_1 y = -1 \times -1 = 1$

$\Delta w_2 = x_2 y = 1 \times -1 = -1$

$\Delta b = -1$

$w_1(new) = old(w_1) + \Delta w_1$

$= 1 + 1$

$w_1(new) = 2$

$w_2(new) = 1 + (-1)$

$= 0$

$b(new) = old(b) + \Delta b$

$= 1 + (-1)$

$= 0$

new $[2 \ 0 \ 0]$

Third I/p :

$[1 \ -1 \ 1]$

$y = -1$

$\Delta w_1 = 1 \times -1 = -1$

$\Delta w_2 = -1 \times -1 = 1$

$\Delta b = -1$

$new(w_1) = old(w_1) + \Delta w_1$

$= 2 + (-1)$

$= 1$

$new(w_2) = old(w_2) + \Delta w_2$

$= 0 + 1$

$= 1$

$new(b) = old(b) + \Delta b$

$= 0 + (-1)$

$= -1$

new $[1 \ 1 \ -1]$

fourth I/P :

$[-1 \ -1 \ 1]$, $y = -1$

$\Delta w_1 = -1 \times -1 = 1$

$\Delta w_2 = -1 \times -1 = 1$

$\Delta b = -1$

$new(w_1) = old(w) + \Delta w_1$

$= 1 + 1$

$= 2$

$new(w_2) = old(w) + \Delta w_2$

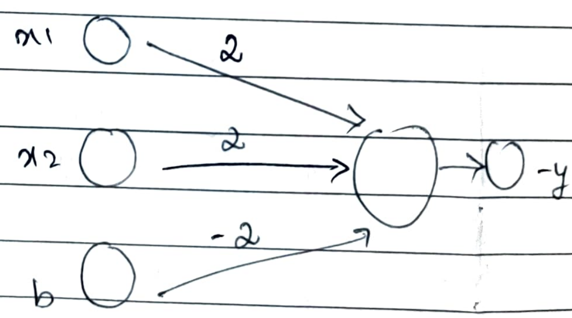
$= 1 + 1$

$= 2$

$new(b) = (-1) + (-1)$

$= -2$

new $[2 \ 2 \ -2]$



Q. Ex-2

OR - f^n

x_1	x_2	b	y	
1	1	1	~~1~~	(1)
1	-1	1	~~1~~	(1)
-1	1	1	~~1~~	(1)
-1	-1	1	~~1~~	(-1)

sol :-

$$w_1 = w_2 = b = 0$$

$$new(w_1) = old(w_1) + \Delta w_1$$

$$new(w_2) = old(w_2) + \Delta w_2$$

$$new(b) = old(b) + \Delta b$$

$$\Delta w_1 = x_1 y$$

$$\Delta w_2 = x_2 y$$

$$\Delta b = y.$$

first I/p :-

$$\begin{bmatrix} x_1 & x_2 & b \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}$$

$$y = 1$$

$$\Delta w_1 = 1 \times 1 = 1$$

$$\Delta w_2 = 1 \times 1 = 1$$

$$\Delta b = y = 1$$

$$new(w_1) = old(w_1) + \Delta w_1$$

$$= 0 + 1$$

$$new(w_1) = 1$$

$$new(w_2) = 0 + (1) = 1$$

$$new(b) = 0 + 1 = 1$$

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

Second I/P :-

[1 -1 1]

$y = 1$

$\Delta w_1 = \cancel{old + \Delta w_1}$

$= x_1 y$

$= 1 \times 1 = 1$

$\Delta w_2 = 1 \times 1 = -1$

$\Delta b = 1$

$new(w_1) = old(w) + \Delta w_1$

$= 1 + 1$

$= 2$

$new(w_2) = 1 + (-1)$

$= 0$

$new(b) = 1 + 1$

$= 2$

new [2 0 2].

Third I/P :-

[-1 1 1]

$y = 1$

$\Delta w_1 = -1 \times 1 = -1$

$\Delta w_2 = 1 \times 1 = 1$

$\Delta b = 1$

$new(w_1) = 2 - 1$

$= 1$

$new(w_2) = 0 + 1$

$= 1$

$new(b) = 2 + 1$

$= 3$

= [1 1 3]

Fourth I/P :-

[-1 -1 1]

$y = -1$

$\Delta w_1 = -1 \times -1 = 1$

$\Delta w_2 = -1 \times -1 = 1$

$\Delta b = -1$

$new(w_1) = 1 + 1 = 2$

$new(w_2) = 1 + 1 = 2$

$new(b) = 3 + (-1) = 2$

[2 2 2]

Ex-3

+	+	+
	+	
+	+	+

'I'

+	+	+
+		+
+	+	+

'O'

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	b	y
I	1	1	1	-1	1	-1	1	1	1	1	1
O	1	1	1	1	-1	1	1	1	1	1	-1

sol:- $w_1 = w_2 = w_3 = w_4 = w_5 = w_6 = w_7 = w_8 = w_9 = b = 0$.

First I/P :- 'I' , y = 1

$\Delta w_1 = 1 \times 1 = 1$

$\Delta w_2 = 1 \times 1 = 1$

$\Delta w_3 = 1 \times 1 = 1$

$\Delta w_4 = -1 \times 1 = -1$

$\Delta w_5 = 1 \times 1 = 1$

$\Delta w_6 = -1 \times 1 = -1$

$\Delta w_7 = 1 \times 1 = 1$

$\Delta w_8 = 1 \times 1 = 1$

$\Delta w_9 = 1 \times 1 = 1$

$\Delta b = 1$

new$(w_1) = 0 + 1 = 1$

new$(w_2) = 0 + 1 = 1$

new$(w_3) = 0 + 1 = 1$

new$(w_4) = 0 + -1 = -1$

new$(w_5) = 0 + 1 = 1$

new$(w_6) = 0 + -1 = -1$

new$(w_7) = 0 + 1 = 1$

new$(w_8) = 0 + 1 = 1$

new$(w_9) = 0 + 1 = 1$

new$(b) = 0 + 1 = 1$

Second I/P = 'O' , y = -1

$\Delta w_1 = 1 \times -1 = -1$

$\Delta w_2 = 1 \times -1 = -1$

$\Delta w_3 = 1 \times -1 = -1$

$\Delta w_4 = 1 \times -1 = -1$

$\Delta w_5 = -1 \times -1 = 1$

$\Delta w_6 = 1 \times -1 = -1$

$\Delta w_7 = 1 \times -1 = -1$

$\Delta w_8 = 1 \times -1 = -1$

$\Delta w_9 = 1 \times -1 = -1$

$\Delta b = -1$

new$(w_1) = 1 + -1 = 0$

new$(w_2) = 1 + (-1) = 0$

new$(w_3) = 1 + (-1) = 0$

new$(w_4) = -1 + (-1) = -2$

new$(w_5) = 1 + 1 = 2$

new$(w_6) = -1 + (-1) = -2$

new$(w_7) = 1 + (-1) = 0$

new$(w_8) = 1 + (-1) = 0$

new$(w_9) = 1 + (-1) = 0$

new$(b) = 1 - 1 = 0$

Q. Hebbian learning :-

-> • Hebbian learning rule is also known as hebb learning rule.

• It is the easiest learning rule in neural network.

• It has single layer network i.e one input layer & one output layer.

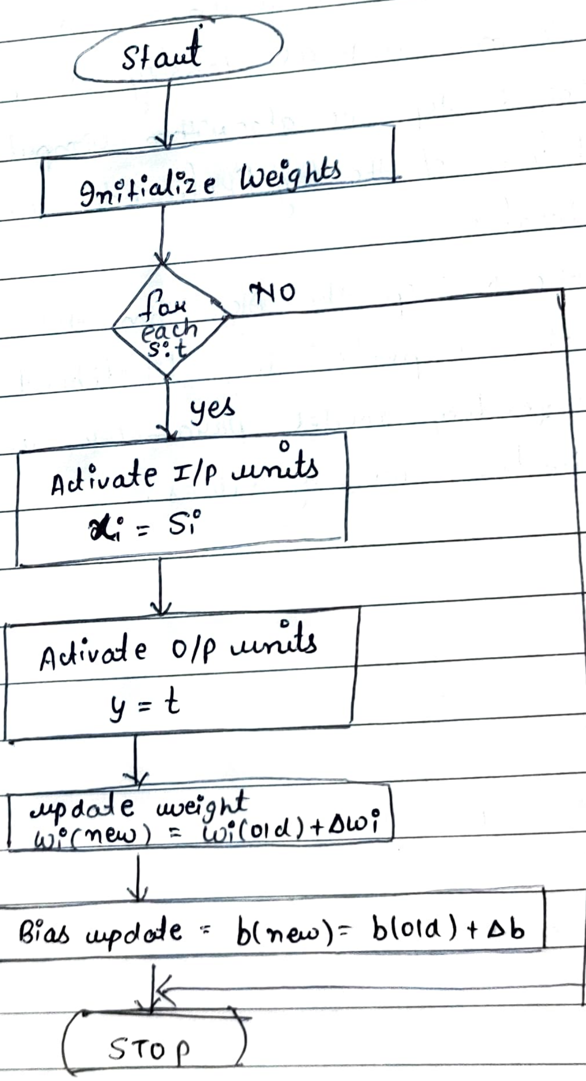
• Input layer can have n units.

• Output layer has one unit.

• The formula to update weight is

$w_i(new) = w_i(old) + \Delta w_i$

where $\Delta w_i = x_i y$.



Q. Explain Expectation Maximization Algorithm ?

-> • The EM Algorithm is an iterative approach that cycles betn two modes.

• The first mode attempt to estimate the missing variables, called the estimation-step or E-Step.

• E-Step basically estimate the missing variables in the dataset.

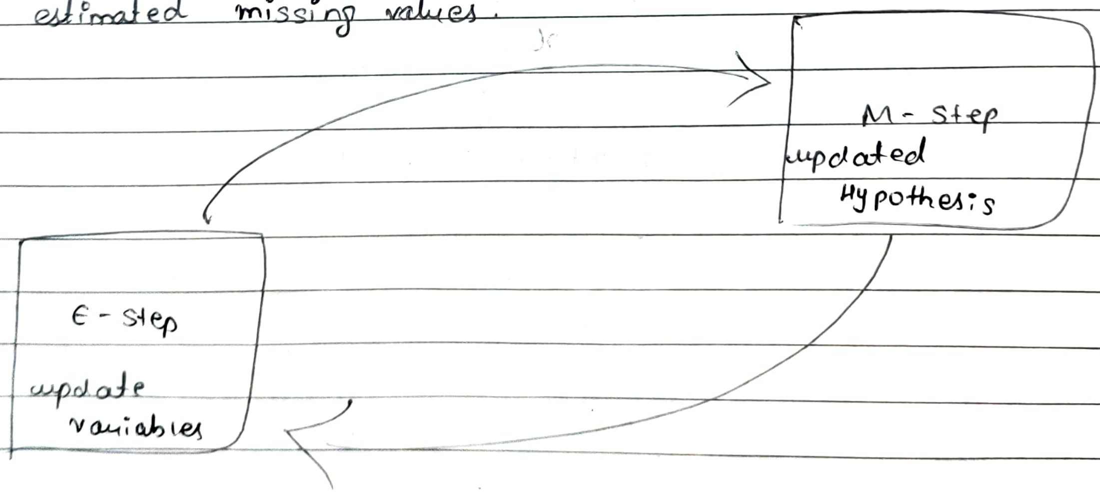
• The second mode attempt to optimize the parameter of the model to best explain the data, called the maximum-step or M-Step.

• M-Step basically Maximize the parameters of the model in the presence of the data.

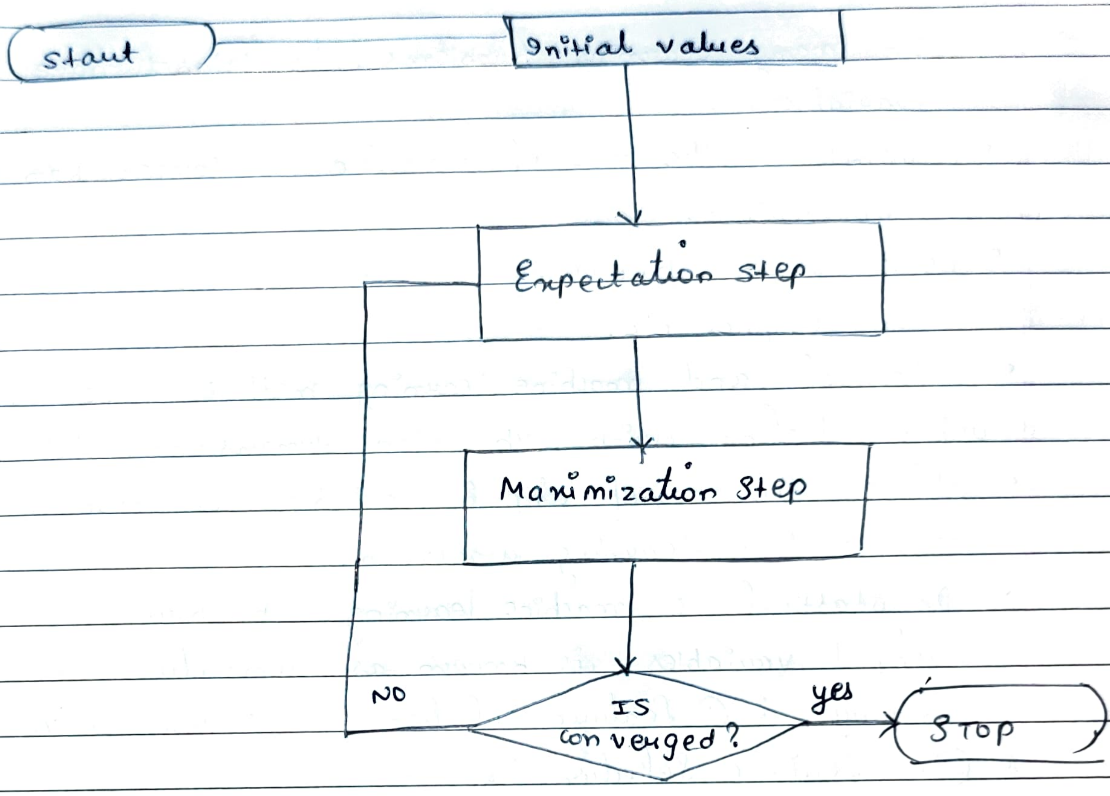
• The EM Algorithm can be applied quite widely, its most well known in Machine learning for use in unsupervised learning problems such as density estimation & clustering

• In the E-Step the algorithm computes the missing variables i.e expectation of the log likelihood using the current parameter estimates.

• In the M-Step the Algorithm determines the parameter that maximize the expected log-likelihood obtained in the E-Step & corresponding model parameter are updated based on the estimated missing values.



- EM flowchart :-



① incomplete data send karto System la

- Initially, a set of initial values of parameters are considered. A set of in-complete observed data is given to the system with the assumption that the observed data comes from a specific model.

② E-step karto Expect karte ke incomplete data madhe value ky fill karnar asnar

- The next step is known as "Expectation" - step or E-step. we use the observed data in order to estimate or guess the values of the missing or incomplete data. It is basically used to update the variables.

③ M-step E-step data optimize karna & data la calculate karto

- The next step is known as "Maximization" - step or M-step. In this step, we use the compute data generated in the preceding "Expectation" - step in order to update the values of the parameters. It is basically ~~to~~ used to update the hypothesis.

④ optimal jo parameter nahi step ② & step ③ repeate karnar calculate karnar sagla iteration ① cha

- Now, in the fourth step, it is checked whether the values are converging or not, if yes then stop otherwise repeate step 2 & step 3.

Module - 6

Q. Why dimensionality reduction is very important step in ML?

→ • In many learning problems, the dataset have large no. of variables.

• Sometimes the no. of variable is larger than the no. of observation.

• Eg :- Image processing, time series analysis, internet search engines etc.

• Statical and machine learning methods have some difficulty while dealing with such high dimensional data.

• Hence, the no. of input variables is reduced before applying the machine Learning algorithm.

• In Statical & machine learning, The process of reducing the no. of input variables is known as dimentional reduction.

• 2 ways : ① feature Selection ② feature Extraction.

• ①. feature Selection :

In feature selection we are intrested in finding k of total n feature that provide most of the information & discarding (n - k) dimention.

②. feature Extraction :

In feature Extraction we are intrested in finding a new set of k feature that are the combination of original n feature.

Q. Concept of feature Selection ~~& Extraction~~ & Extraction ?

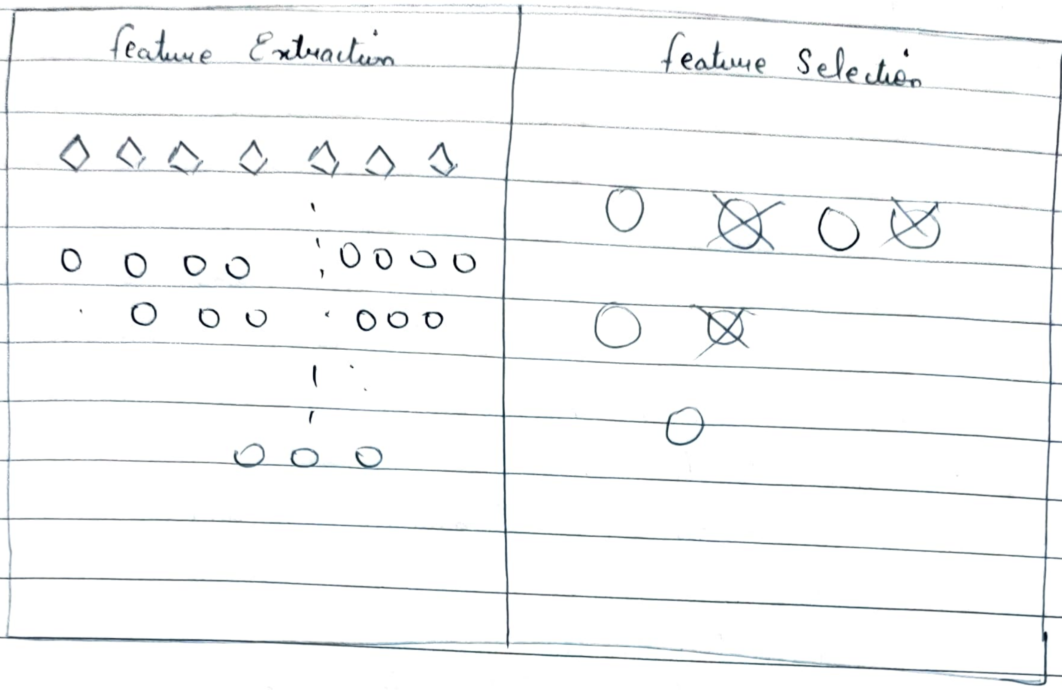
→ ① feature Selection :-

- In feature selection we are intrested in finding k of total n feature that provide most of the information and discarding (n-k) dimensions.
- feature Selection is a way of selecting the subset of most relevent feature from the original ~~feature Extraction :-~~ feature sets by removing the redundancy irrelevant or noisy features.
- It is a way of reduming the input variable for the model.
- Two main Techniques :-
  - Superived feature selection Technique — consider Target var, uses labeled data
  - Unsupervied feature Sel^n Tech^n — uses unlabelled data, Ignore Target var

~~the feature~~

② feature Extraction :-

- In the feature extraction, we are intrested in finding a new set of k features that are the combination of original n features.
- These methods can be Supervised or unsupervied depending on whether or they use the output info or not
- The best known & widely used feature extraction method are PCA (~~Pun~~ Principal component App^n) & linear discriminent analysis. ~~LDA~~ (LDA)
- Basically, feature extraction is process that involves identifying & extracting relevant feature from raw data.



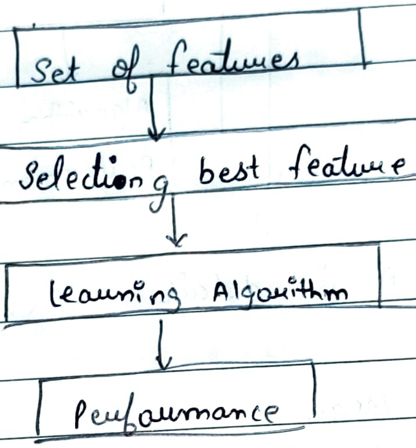
Q. Feature Selection Method used for dimentionality reduction.

→

- Feature Selection Technique
  - Supervised feature selection
    - Filter Method
      - missing value
      - infogain
      - chi-square Test
      - fisher's score
    - Embedded method
      - Regularization L_1, L_2
      - Random forest importance
    - Wrappers methods
      - forward feat^n sel^n
      - Backward f^n sel^n
      - Exhaustive f^n sel^n
      - Recursive f^n sel^n
  - unsupervised feature selection

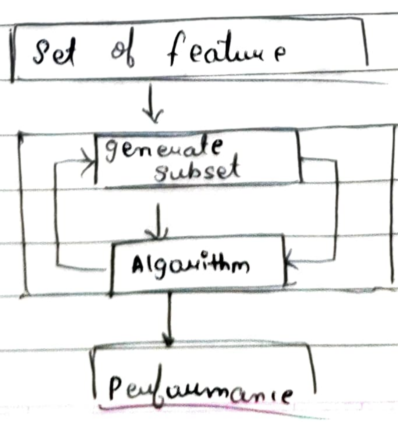
1. Filter Methods :-

- In filter method, features are selected on the basis of Stastical measures.
- This method does not depend on the learning Algorithm & chooses feature as pre-processing step.
- The filter method filter out the irrelevant feature & redundant columns from the model by using different metrices through ranking.



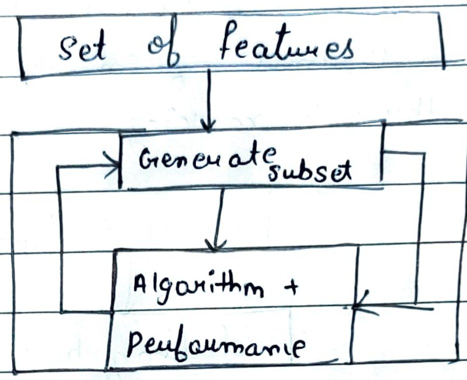
2. Wrapper Method :-

- In wrapper methodology, Selection of feature is done by considering it as a search problem.
- In which different combinations are made, evaluated, & compared with other Combinations.
- It trains the Algorithm by using the subset of features iteratively.



3. Embedded Method :-

- Embedded methods combined the advantage of both filter and wrapper methods by considering the ~~iter~~ interaction of features along with low computational cost.
- ~~These are fast pure~~



Q. Explain PCA in detail -

→ Steps :-

1. Calulate Mean
2. Calculate Covariance Matrix
3. Calculate Eigen value of covariance Matrix
4. Compute Eigen vectors
5. compute 1^{st} principal component
6. Geometrical meaning of 1^{st} Principal component.

Ex. 1 PCA :-

F	Ex 1	Ex 2	Ex 3	Ex 4
x_1	4	8	13	7
x_2	11	4	5	14

Step 1 : calculate Mean :-

$N = 4$

$$\bar{x}_1 = \frac{4+8+13+7}{4} = 8$$

$$\bar{x}_2 = \frac{11+4+5+14}{4} = 8.5$$

Step 2 : calculate the covariance matrix :-

$$S = \begin{bmatrix} cov(x_1, x_1) & cov(x_1, x_2) \\ cov(x_2, x_1) & cov(x_2, x_2) \end{bmatrix}$$

$$cov(x_1, x_1) = \frac{1}{N-1} \sum_{k=1}^{N} (x_{1k} - \bar{x}_1)(x_{1k} - \bar{x}_1)$$

$$= \frac{1}{3} \left[(4-8)^2 + (8-8)^2 + (13-8)^2 + (7-8)^2 \right]$$

$$= \frac{1}{3} \left[16 + 0 + 25 + 1 \right]$$

$$= \frac{1}{3} \left[42 \right]$$

$$cov(x_1, x_1) = \underline{14}$$

$$cov(x_1, x_2) = \frac{1}{N-1} \sum_{k=1}^{N} (x_{1k} - \bar{x}_1)(x_{2k} - \bar{x}_2)$$

$$= \cancel{\frac{1}{3} \left[(4-8) + (8-8) + (13-8) + (7-8) \right]}$$

$$= \frac{1}{3} \left[(4-8)(11-8.5) + (8-8)(4-8.5) + (13-8)(5-8.5) + (7-8)(14-8.5) \right].$$

$$= -11$$

$cov(x_2, x_1) = -11$

$$cov(x_2, x_2) = \frac{1}{N-1} \sum_{k=1}^{N} (x_{2k} - \bar{x}_2)(x_{2k} - \bar{x}_2)$$

$$= \frac{1}{3}\left[(11-8.5)^2 + (4-8.5)^2 + (5-8.5)^2 + (14-8.5)^2\right]$$

$$= \frac{1}{3}\left[6.25 + 20.25 + 12.25 + 30.25\right]$$

$$= \frac{1}{3}\left[69\right]$$

$cov(x_2, x_2) = 23$.

$\therefore \quad S = \begin{bmatrix} 14 & -11 \\ -11 & 23 \end{bmatrix}$ .. covariance matrix.

Step 3 : <u>Eigen values of the covariance matrix</u> :-

$$\begin{bmatrix} 14-\lambda & -11 \\ -11 & 23-\lambda \end{bmatrix} \quad .. \quad \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

$$= \lambda^2 - (a+d)\lambda + [ad - bc] = 0$$

$$= \lambda^2 - 37\lambda + 201 = 0$$

$\lambda_1 = 30.38$

$\lambda_2 = 6.61$

Step 4 : Eigen vectors computation :-

$$U = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

$$\begin{bmatrix} 14-\lambda & -11 \\ -11 & 23-\lambda \end{bmatrix} \times \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

$$\begin{bmatrix} 14-\lambda\ u_1 & -11\ u_2 \\ -11\ u_1 & 23-\lambda\ u_2 \end{bmatrix}$$

Take any one eq. from above 2 eq.

$$(14-\lambda)u_1 - 11 u_2 = 0$$

$$\therefore (14-\lambda) u_1 = 11 u_2$$

$$\therefore \frac{u_1}{11} = \frac{u_2}{14-\lambda}$$

$\therefore u_1 = 11$

$\therefore u_2 = 14-\lambda$

$$U_1 = \begin{bmatrix} 11 \\ 14-\lambda \end{bmatrix}$$

Take Biggest or greatest λ .. $\lambda_1 = 30.38$

$$\|U\| = \sqrt{11^2 + (14-\lambda)^2}$$

$$= \sqrt{11^2 + (14-30.38)^2}$$

$$\|U\| = 19.730$$

$$e_1 = \begin{bmatrix} 11/\|U_1\| \\ (14-\lambda_1)/\|U_1\| \end{bmatrix}$$

$$= \begin{bmatrix} 11/19.734 \\ 14-30.38/19.734 \end{bmatrix}$$

$$e_1 = \begin{bmatrix} 0.5574 \\ -0.8303 \end{bmatrix}$$

Similarly,

$$e_2 = \begin{bmatrix} 0.8303 \\ 0.5574 \end{bmatrix}$$ — Mostly e_2 is not asked.

Step 5 :- Computation of first principal :-

$$Ex_1 = e_1^T \begin{bmatrix} x_{1k} - \bar{x}_1 \\ x_{2k} - \bar{x}_2 \end{bmatrix}$$

$$[0.5574(4-8) - 0.8303(11-8.5)]$$

$$\begin{bmatrix} 0.5574 & -0.8303 \end{bmatrix} \times \begin{bmatrix} 4-8 \\ 11-8.5 \end{bmatrix}$$

$Ex_1 = -4.30535$

Similarly

$Ex_2 = 3.7361$

$Ex_3 = 5.6928$

$Ex_4 = -5.1238$.

Module - 2

Q. Explain SVD & its application.

-> • The Singular value decomposition of a matrix is a factorization of that matrix into three matrices.

• It has some interesting algebraic properties & conveys important geometrical and theoretical insights about linear transformation.

• It is a very well known matrix decomposition method used in reducing a matrix in fragments to have simpler calculation for certain matrix.

• The fact states that every matrix A belong to R has an SVD.

$$A = U \Sigma V^T$$

• Here, A is a $m \times n$ matrix

• U is an $m \times n$ orthogonal matrix

• V is a $m \times n$ orthogonal matrix.

* Application :-

① Calculation of Pseudo Inverse - The Pseudo Inverse is the generalization of the matrix that may not be invertible. It is denoted by A^+.

② Rank - The rank of matrix M can be calculated from SVD by the no. of non zero singular value.

③ Curve fitting - The SVD can be used to minimize the least square error.

④ The SVD & Pseudo inverse can also be used in Digital Signal processing & image processing.

Q. SVD Numerical :-

→ $SVD = U\Sigma V^T$

$$A = \begin{bmatrix} 1 & -1 & 3 \\ 3 & 1 & 1 \end{bmatrix}$$

→ Step 1 : $V = A^T A$.

$$= \begin{bmatrix} 1 & 3 \\ -1 & 1 \\ 3 & 1 \end{bmatrix} * \begin{bmatrix} 1 & -1 & 3 \\ 3 & 1 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 10 & 2 & 6 \\ 2 & 2 & -2 \\ 6 & -2 & 10 \end{bmatrix}$$

Step 2 : Calulate Eigen values & Eigen vectors.

Standard Eqn = $\lambda^3 - S_1\lambda^2 + S_2\lambda - |A| = 0$

S_1 = Sum of dig. element

S_2 = Sum of minar elements.

$\therefore S_1 = 10 + 2 + 10 = 22$

$\therefore S_2 =$

$$\begin{bmatrix} 2 & -2 \\ -2 & 10 \end{bmatrix} + \begin{bmatrix} 10 & 6 \\ 6 & 10 \end{bmatrix} + \begin{bmatrix} 10 & 2 \\ 2 & 2 \end{bmatrix}$$

$= (20 - 4) + (100 - 36) + (20 - 4)$

$S_2 = \underline{96}$.

$|A| = 0$.

$= \lambda^3 - 22\lambda^2 + 96\lambda - 0 = 0$

$\lambda_1 = 16, \lambda_2 = 6, \lambda_3 = 0$.

$x_3 = 0$

$$\begin{bmatrix} 10 & 2 & 6 \\ 2 & 2 & -2 \\ 6 & -2 & 10 \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0$$

$$\begin{bmatrix} 10 & 2 & 6 \\ 2 & 2 & -2 \\ 6 & -2 & 10 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0$$

$$10x_1 + 2x_2 + 6x_3 = 0$$
$$2x_1 + 2x_2 - 2x_3 =$$

$$\frac{x_1}{\begin{vmatrix} 2 & 6 \\ 2 & -2 \end{vmatrix}} = \frac{-x_2}{\begin{vmatrix} 10 & 6 \\ 2 & -2 \end{vmatrix}} = \frac{x_3}{\begin{vmatrix} 10 & 2 \\ 2 & 2 \end{vmatrix}}$$

$$\frac{x_1}{-16} = \frac{\cancel{-}x_2}{\cancel{-}32} = \frac{x_3}{16}. \quad \begin{bmatrix} -16 \\ 32 \\ 16 \end{bmatrix} \div 16 = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}.$$

Normalize $\sqrt{(-1)^2 + (2)^2 + (1)^2}$

$= \sqrt{1 + 4 + 1}$

$= \sqrt{6}.$

$$\begin{bmatrix} -1/\sqrt{6} \\ 2/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix}.$$

S

9

scrib

on

$$V = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{3} & -1/\sqrt{6} \\ 0 & 1/\sqrt{3} & 2/\sqrt{6} \\ 1/\sqrt{2} & -1/\sqrt{3} & 1/\sqrt{6} \end{bmatrix}.$$

$$V^T = \begin{bmatrix} 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 1/\sqrt{3} & 1/\sqrt{3} & -1/\sqrt{3} \\ -1/\sqrt{6} & 2/\sqrt{6} & 1/\sqrt{6} \end{bmatrix}.$$

$$\Sigma = \begin{bmatrix} \sqrt{16} & 0 & 0 \\ 0 & \sqrt{6} & 0 \\ 0 & 0 & \sqrt{0} \end{bmatrix} = \begin{bmatrix} 4 & 0 & 0 \\ 0 & \sqrt{6} & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Step : Calculate $u = AA^T$.

$$\begin{bmatrix} 1 & -1 & 3 \\ 3 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ -1 & 1 \\ 3 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 11 & 5 \\ 5 & 11 \end{bmatrix}$$

$\lambda^2 - 22\lambda + 96 = 0$.

$\lambda_1 = 16$, $\lambda_2 = 6$.

$\lambda = 16$

$$\begin{bmatrix} 11 & 5 \\ 5 & 11 \end{bmatrix} - \begin{bmatrix} 16 & 0 \\ 0 & 16 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0$$

$$\begin{bmatrix} -5 & 5 \\ 5 & -5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0$$

- Eigen vector :

$(A - \lambda I) = 0$

$\lambda_1 = 16$

$$\begin{bmatrix} 10 & 2 & 6 \\ 2 & 2 & -2 \\ 6 & -2 & 10 \end{bmatrix} - \begin{bmatrix} 16 & 0 & 0 \\ 0 & 16 & 0 \\ 0 & 0 & 16 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0$$

$$\begin{bmatrix} -6 & 2 & 6 \\ 2 & -14 & -2 \\ 6 & -2 & -6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0$$

Take the eq which are unique i.e different from each other.

$-6x_1 + 2x_2 + 6x_3 = 0$

$2x_1 - 14x_2 - 2x_3 = 0$

$$\frac{x_1}{\begin{vmatrix} 2 & 6 \\ -14 & -2 \end{vmatrix}} = \frac{-x_2}{\begin{vmatrix} -6 & 6 \\ 2 & -2 \end{vmatrix}} = \frac{x_3}{\begin{vmatrix} -6 & 2 \\ 2 & -14 \end{vmatrix}}$$

$$= \frac{x_1}{-4+84} = \frac{-x_2}{12-12} = \frac{x_3}{84-4}$$

$$= \frac{x_1}{80} = \frac{-x_2}{0} = \frac{x_3}{80} = \begin{bmatrix} 80 \\ 0 \\ 80 \end{bmatrix} \div 80 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

Normalize $= \sqrt{1^2 + 0^2 + 1^2}$

$= \sqrt{2}$

$$= \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix}$$

$\lambda_2 = 6$.

$(A - \lambda I) = 0$.

$$\begin{bmatrix} 10 & 2 & 6 \\ 2 & 2 & -2 \\ 6 & -2 & 10 \end{bmatrix} - \begin{bmatrix} 6 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0.$$

$$= \begin{bmatrix} 4 & 2 & 6 \\ 2 & -4 & -2 \\ 6 & -2 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0$$

$$= \begin{cases} 4x_1 + 2x_2 + 6x_3 = 0 \\ 2x_1 - 4x_2 - 2x_3 = 0 \\ 6x_1 - 2x_2 + 4x_3 = 0 \end{cases}$$

$4x_1 + 2x_2 + 6x_3 = 0$

$2x_1 - 4x_2 - 2x_3 = 0$.

$$\frac{x_1}{\begin{vmatrix} 2 & 6 \\ -4 & -2 \end{vmatrix}} = \frac{-x_2}{\begin{vmatrix} 4 & 6 \\ 2 & -2 \end{vmatrix}} = \frac{x_3}{\begin{vmatrix} 4 & 2 \\ 2 & -4 \end{vmatrix}}$$

$$\frac{x_1}{-4+24} = \frac{-x_2}{-8-12} = \frac{x_3}{-16-4}$$

$$\frac{x_1}{20} = \frac{x_2}{20} = \frac{x_3}{-20} \qquad \begin{bmatrix} 20 \\ 20 \\ -20 \end{bmatrix} \div 20 = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$$

Normalize $= \sqrt{(1)^2 + (1)^2 + (-1)^2}$

$= \sqrt{3}$.

$$\begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ -1/\sqrt{3} \end{bmatrix}$$

$-5x_1 + 5x_2 = 0$ — ①

$5x_1 - 5x_2 = 0$ — ②

Assume $x_1 = 1$ in eq ①

$-5 + 5x_2 = 0$

$5x_2 = 5$

$x_2 = \frac{\cancel{5}^1}{\cancel{5}^1}$

$x_2 = 1$ $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$

Normalize $= \sqrt{(1)^2 + (1)^2}$

$= \sqrt{2}$

$\begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$

$\lambda = 6$

$$\left(\begin{bmatrix} 11 & 5 \\ 5 & 11 \end{bmatrix} - \begin{bmatrix} 6 & 0 \\ 0 & 6 \end{bmatrix} \right) \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0$$

$$\begin{bmatrix} 5 & 5 \\ 5 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0$$

$5x_1 + 5x_2 = 0$ — ③

$5x_1 + 5x_2 = 0$ — ④

Assume, $x_1 = 1$ in ③

$5 + 5x_2 = 0$

$5x_2 = -5$

$x = -1$ $\begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$

$N = \sqrt{(1)^2 + (1)^2}$

$= \sqrt{2}$

$\therefore SVD = U \Sigma V^T$

$$= \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 4 & 0 & 0 \\ 0 & \sqrt{6} & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1\sqrt{2} & 0 & 1/\sqrt{2} \\ 1/\sqrt{3} & 1/\sqrt{3} & -1/\sqrt{3} \\ -1/\sqrt{6} & 2/\sqrt{6} & 1/\sqrt{6} \end{bmatrix}$$

Module: 1

Q. What is Machine Learning? What are steps in developing a ML Application? (AI)

→ • Machine Learning is a subset of Artificial intelligence that involves the development of Algorithm & Statical model that enable computer to perform task without explicit instruction.

• Instead, these Algorithm learn and improve from experience & data.

• The primary goal of Machine learning is to enable computers to automatically learn & improve from experience without being explicitly programmed to do so.

STEPS :-

There are major 7 Steps to develop a ML app^n.

1. Collecting the data.
2. Preparing the data
3. Choosing the model
4. Training the model
5. Evaluating the model
6. Parameter Tuning
7. Making prediction.

1. Collecting the data :-

• It is important to collect reliable data so that your ML model can find the correct patterns.

• The quality of the data that you feed to the machine will determine how accurate your model will be.

• If you provide outdate or incorrect data, you will have wrong outcome or predictions.

Step ②. Preparing the data :-

- Cleaning the data to remove unwanted data, missing value, duplicate values etc.
- visualize the data to understand how it is structured & understand the relationship between various variables & classes.
- Spliting the cleaned data into 2 sets :-
  Training set & Testing set
  Training set to train the model Testing set to check the accuracy of the model.

Step ③ Choosing the Model :-

- A ML model determines the O/P after running the machine learning Algo.
- It is important to choose the model which is relevant to the task.
- The model Should be Suited for different tasks such as image recognition, speech recognition etc.
- The model you choose Should be Suited for numerical or categorial data.

Step ④ Training the Model :-

- Training is most important Step in ML.
- In Training you pass the prepared data to the ML model to find patterns & make prediction.
- The model learns from the data to accomplish the task & with training the model gets better at predicting.

Step ⑤ Evaluating the Model :-

- After training your model you have to check how is it Performing.
- The performance of the model is tested using the unseen data. ~~will not get on~~
- If testing was done on the Same data that

was used for training the you will not get an accurate measure or your model will not predict accurate results.

* Step ⑥ : Parameter Tuning

- Once you have created and evaluated your model, see if its ~~a~~ ~~an~~ accuracy can be improved.
- This is done by ~~training~~ tuning the parameter in your model.
- Parameter tuning refer to finding the values whose accuracy will be max.

Step ⑦ Making prediction

At the end you can use your model on unseen data to make prediction accurately.

* Application of ML :-

① Image Recognition -

- Image Recognition is one of the most common app^n of ML
- It is used to identify objects, preson, places, digital image etc.
- Eg :- Instagram provides the feature of auto friend tagging suggestion.

② Speech Recognition -

- While using Google we get an option of 'search by voice' it comes under speech recognition.
- Speech recognition is a process of converting the voice instruction to text & it is also known as speech to text or 'computer speech Recognition'.
- Eg :- Google Asistant, siri, Bixsiby.

③ Product recommendation :-

- Machine learning is widely used by various ecommerce & entertainment companies such as Amazon, Netflix etc.
- for Product recommendation to the user.
- When we use Netflix, we find some recommendation for entertainment series, movies etc

④ Self-driving car :-

- ML plays a significant role in self-driving cars.
- It uses unsupervised learning method to train the car model to detect people & objects while driving.
- Tesla the most popular car manufactory company is working on self-driving car.

⑤ Email Spam fitting :-

- Whenever we recieve a new email, it is filtered automatically as important normal & spam.
- We always receive an important mail in our inbox with the important symbol & spam email in the spam box.
- Eg: Email.

⑥ Stock Market Prediction :-

- Machine learning is widely used in stock market prediction.
- In the stock market there is always a risk of up & down in shares so for this machine learning ~~long~~ long-short term memory neural network used for the prediction of stock market trends.
- Eg :- Groww, Stake etc.

Q. Diffⁿ betⁿ Supervised & Unsupervised

->

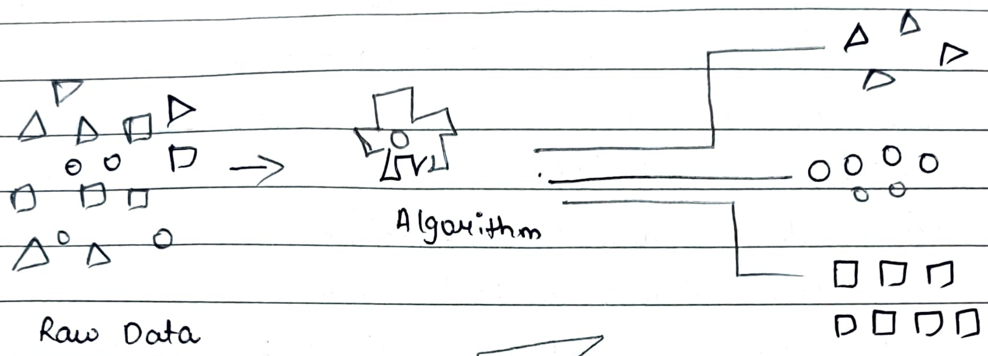
	Parameter	Supervised ML	Unsupervised ML
I.	Input Data	Algorithm are trained using labeled data.	Algorithm are trained ~~used~~ using unlabeled data.
C.	~~Computational~~ complexity	Simpler method	Complex method.
A.	Accuracy	Highly acurate	Less accurate
N.	No. of classes	No. of classes is known	No. of classes is not known
D.	Data analysis	uses offline analysis	uses real time analysis
O.	output	Desired output is given	Dired o/p is not given.
T.	Training data	use training data to infer method	No training data is ~~use~~
T.	~~Model~~ Testing model	we can test our model	we can not test our model
C.	called as	called as classification	called as clustering
E.	Example	Eg: Random forest	Eg: ~~Le~~ Linear & Logistic R.
S.	Supervision	Supervised learning needs supervision to train the model.	Unsupervised learning does not need any Supervision to train the model.

CANDOTTEES

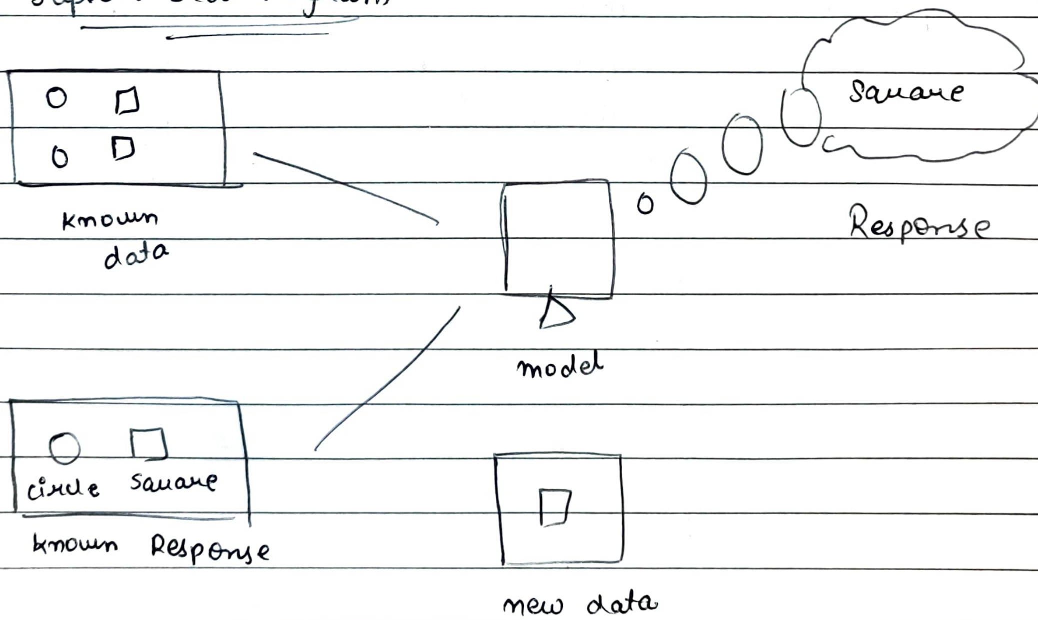
Q. Supervised & unsupervised Learning?

-> • Learning is the process converting experience into expertise.

• Learning can be broadly classified into 3 categories, based on nature of learning & the interaction bet^{n} learner & the environment.



Supervised Diagram



Example :

		Predicted		
		Positive	Negative	Total
Actual	Positive	10 (TP)	22 (FN)	32
	Negative	8 (FP)	60 (TN)	68
	Total	18	82	100

- Accuracy $= \dfrac{TP + TN}{Total}$

$= \dfrac{10 + 60}{100}$

$= \dfrac{70}{100} = \underline{\underline{0.7}}$

- Recall $= \dfrac{TP}{TP + FN}$

$= \dfrac{10}{10 + 22} = \underline{\underline{0.31}}$

- Precision $= \dfrac{TP}{TP + FP}$

$= \dfrac{10}{10 + 8} = \underline{0.56}$

- Specificity $= \dfrac{TN}{TN + FP}$

$= \dfrac{60}{60 + 8} = \underline{\underline{0.73}}$

- F1 Score $= 2 \times \left[\dfrac{Precision \times Recall}{Precision + Recall}\right] = 2 \times \left(\dfrac{0.56 \times 0.31}{0.56 + 0.31}\right)$

$= 2 \times \left(\dfrac{0.1736}{0.87}\right)$

$= \approx \underline{\underline{0.4}}$

$\int$ Norms :-

$$||x||_p = \left(\sum_i^n |x_i|^p\right)^{1/p} = \left(|x_1|^p + |x_2|^p \ldots + |x_n|^p\right)^{1/p}$$

eg :- $v = [\,1 \quad 0 \quad 4 \quad 9\,]$

$$l_0 = |1|^0 + |0|^0 + |4|^0 + |9|^0.$$
$$= 1 + 0 + 1 + 1$$
$$l_0 = \underline{\underline{3}}$$

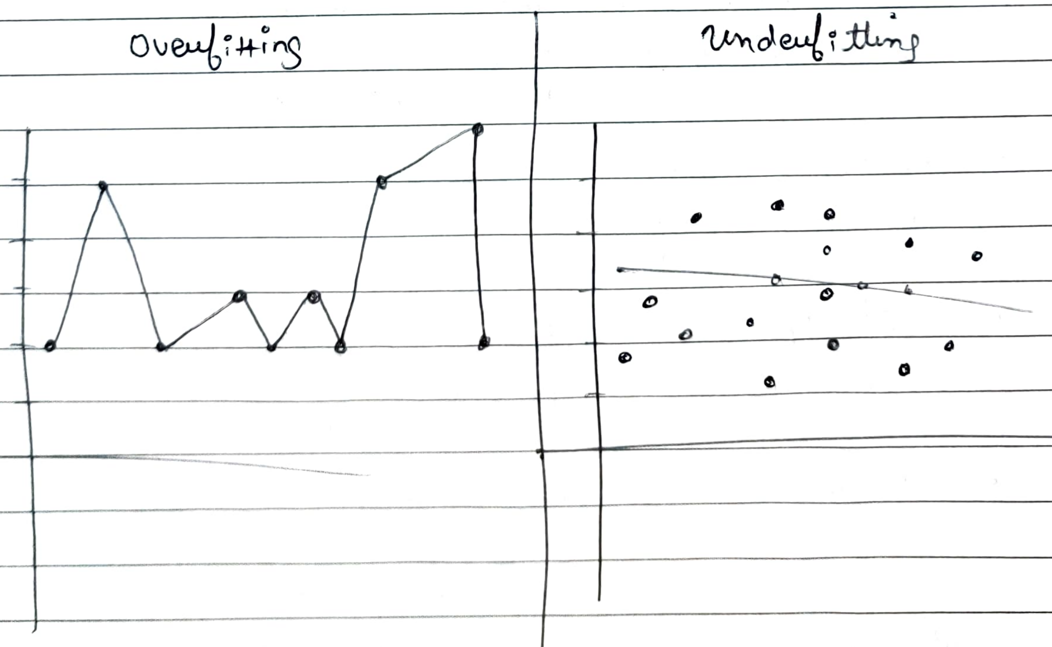
$$l_1 = |1|^1 + |0|^1 + |4|^1 + |9|^1$$
$$= 1 + 0 + 4 + 9$$
$$= 14$$

$$l_2 = |1|^2 + |0|^2 + |4|^2 + |9|^2$$
$$= 1 + 0 + 16 + 81$$
$$l_2 = \underline{\underline{\sqrt{98}}}$$

Q. Overfitting & Underfitting in ML.
->

* Overfitting :
- Overfitting occurs when a ML model tries to cover all the data points or more than the required data points present in the given dataset.
- Because of this, the model starts caching noise & inaccurate values present in the dataset, & all these factors reduce the efficiency & accuracy of the model.
- The chances of occurances of overfitting increase as much we provide training to our model.
- It means the more we train our model, the more chances of occuring the overfitted model.

* Underfitting :
- underfitting occurs when our ML model is not able to capture the underlying trend of the data.
- In the case of underfitting, the model is not able to learn enough from the training data, & hence it reduces the accuracy & products unreliable preditions.



Q. How to calulate performance Measures by Measuring Quality of models 5

→
- When evaluating the performance of a ML model, several metrices can be used to understand its effectiveness.
- Confusion Matrix is a table used to describe the performance of a classification model.
- It shows the no. of correct & incorrect predictions made by the model compared to the actual outcomes.
- True Positives (TP) : correctly Predicted Positive cases. 6
- True Negative (TN) : Correctly predicted negative cases
- False Positives (FP) : Incorrectly predicted positive case
- False Negatives (FN) : Incorrectly predicted negative case.
- The Matrix looks like this :-

	Predicted Positive	Predicted Negative
Actual Positive	TP	FN
Actual Negative	FP	TN

7

- Accuracy is the ratio of correctly predicted instances to the total instances. Accuracy = $\frac{TP + TN}{Total}$
- Recall measures the ability of the model to correctly identify all Positive instances. Recall = $\frac{TP}{TP + FN}$
- Precision measures the accuracy of the positive preditions. 8
  Precision = $\frac{TP}{TP + FP}$
- Specificity measures the ability of the model to correctly identify all negative instances. Specificity = $\frac{TN}{TN + FP}$
- F1 Score is the harmonic mean of precision & recall 9
  F1 Score = $2 \times \frac{Precision \times Recall}{Precision + Recall}$
- Root Mean Squared Error (RMSE) :-
  Measures prediction errors [illegible]

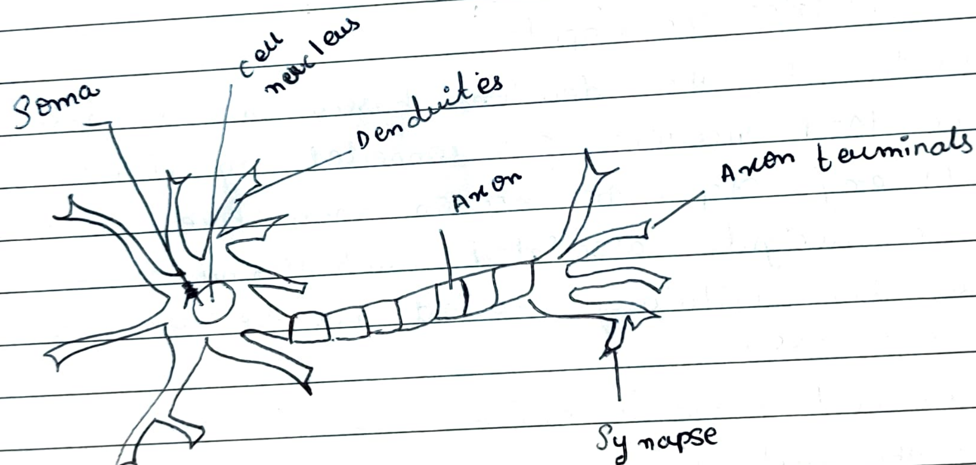
Module - 6

Q. Draw & explain Biological neural networks & compare them with artificial neural networks.

→ • A biological neural network is a group of nerve cell (neurons) in the brain & nervous System that communicate with each other to process information.

• These neurons connect through Synapses, where they Send electrical and chemical Signals to help us think, feel, move & member.

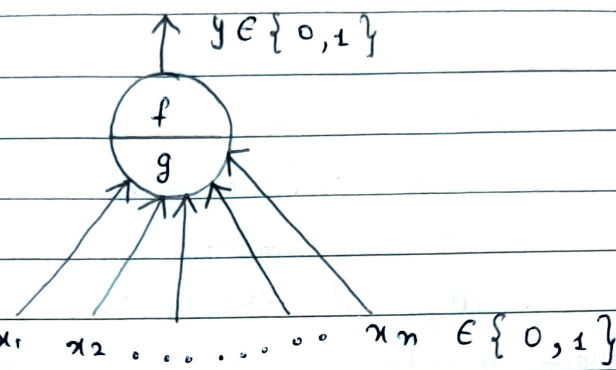
• Essentially, its the natural circuitry that makes our brain works.



	Biological neuron	Artificial Neuron
	• It is made of cell.	• The cells correspond to neuron.
	• It has dendrites which are interconnection betn cell body.	• The connection weight correspond to dendrites.
Learning	• very precise Structure & formatted data	• They can tolerate ambiguity
Processor	• complex, high speed.	• Simple, low speed.
memory	• Separate from a processor	• Integrated into processor.
	• localized	• distributed
computing	• centralized	• Integrated into processor.
Realiblity	• very vulnerable	• Robust.
Expertise	• numerical & symbolic	• Perceptual

Q. Explain MP neuron model :-

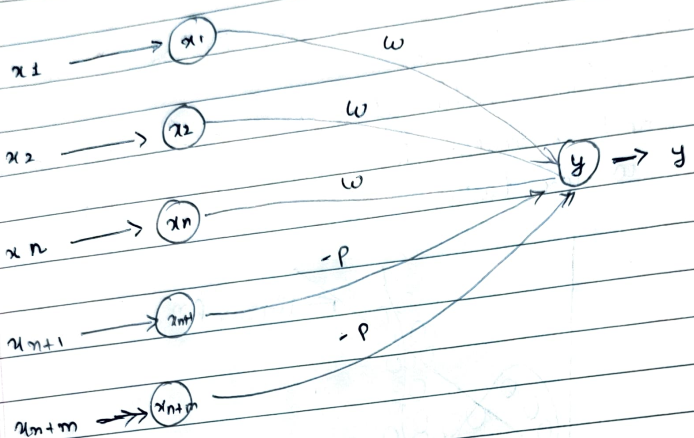
→ • Maculloch (neurosientist) and pitts (logican) proposed a highly Simplified computational model of the neuron (1943)



- g aggregates the inputs.
- function f takes decisions based on this aggregation.
- The M-P neurons are connected by directed weighted paths.
- At any step, the neuron may fire or may not fire.
- The weights associated with communication links may be ~~exit~~ ~~exi~~ excitatory (weight are +ve) or ~~inhi~~ inhibitory (weights are -ve).
- Threshold plays a major role in MP neuron.
- If the net input to the neuron is greater than the Threshold then the neuron fires.
- The M-P neurons are mostly widely used in the case of logic functions.

Architecture :-

The simple MP neuron is shown :-



- The activation f^n here is defined as,

$$f(y_{in}) = \begin{cases} 1 & y_{in} \geq \theta \\ 0 & y_{in} < \theta \end{cases}$$

- The threshold with activation function should satisfy the following condition :

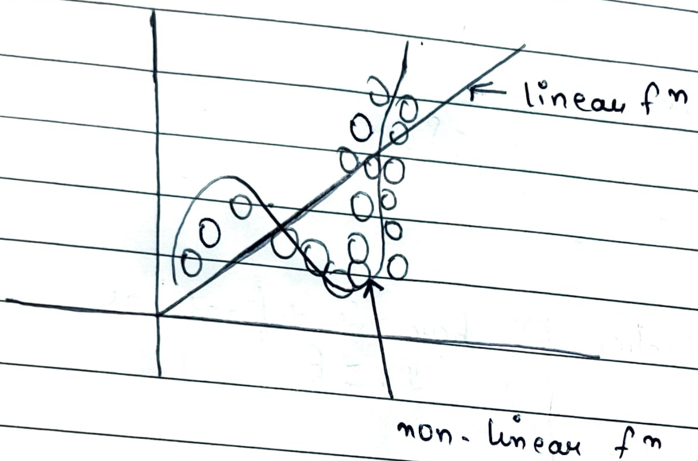
$$\theta \geq nw - p$$

Q. What are Activation fn? Explain Binary, Bipolar, continuous & Ramp Activation fn.

→ • A neural net^n activation function is a function that introduces non linearity into the model.

• A neural network without an activation fn is just a linear regression model. The

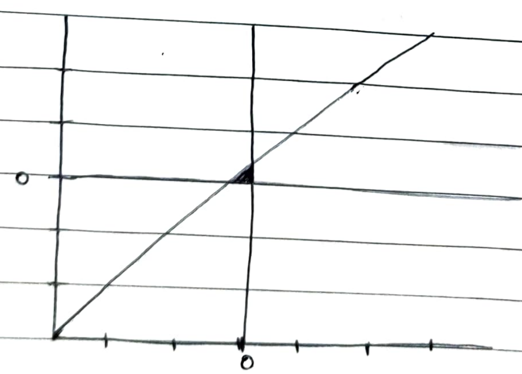
• The activation function does the non-linear transformation to the input, making it capable of learning & performing more complex tasks.



Types of Activation fn :-

① Linear function / Identity fn.

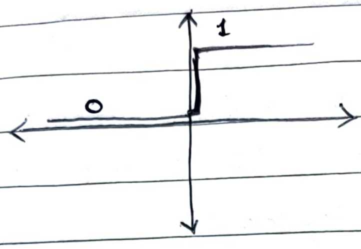
→



In linear process, activation is directly proportional to the input.

② Binary Step function :-

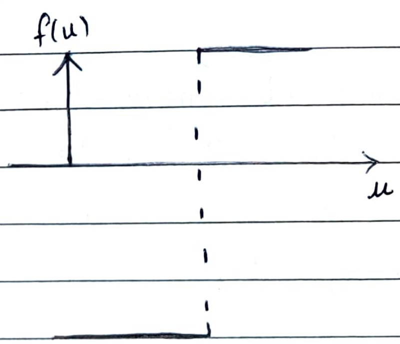
- This function is used in single layer network to convert the net input to output
- The output is binary i.e 0 or 1.
- The θ represent the threshold value.



$$f(x) = \begin{cases} 1 & \text{if } x \geq \theta \\ 0 & \text{if } x < \theta \end{cases}$$

③ Bipolar Step fn :-

- In the Bipolar Step fn, if the value of y is above a certain value known as the threshold, the o/p is +1 and if less than the threshold then the o/p is -1.
- If the Bipolar o/p (+1 to -1).
- It can be utilized in single-layer networks.



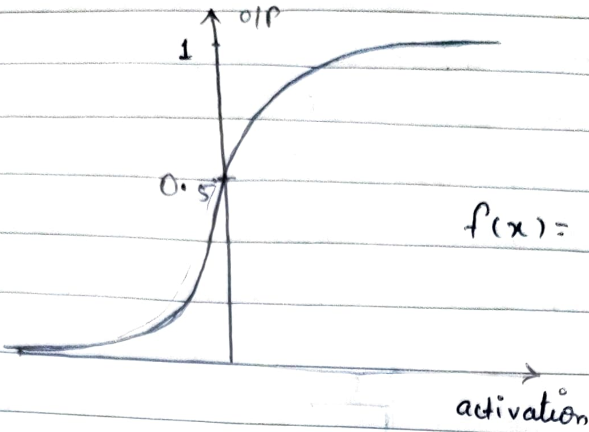
$$f(x) = \begin{cases} 1 & \text{if } x \geq \theta \\ -1 & \text{if } x < \theta \end{cases}$$

④ Sigmoid fn :-

- It is also known as S-Shaped fn. logistic & hyperbolic tangent fn is commonly used in Sigmoid fn.

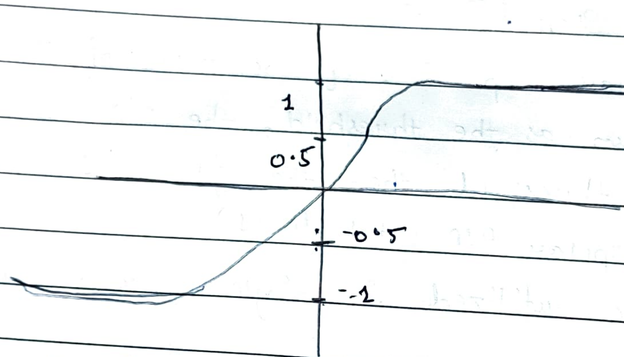
There are two types of Sigmoid fn :-

- Binary Sigmoid fn (Logistic fn)
- Binary Sigmoid fn is a logistic fn where the o/p values are either binary or vary from 0 to 1.



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

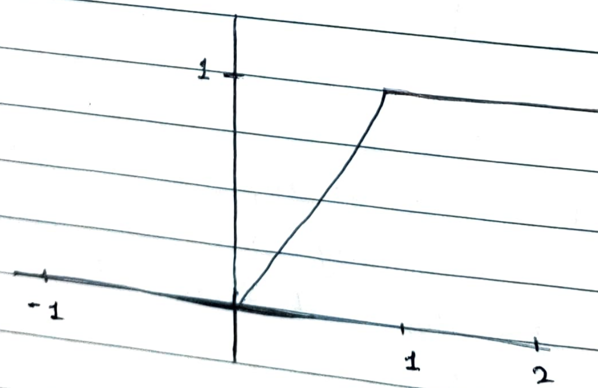
- Bipolar Sigmoid f^n (or Hyperbolic Tangent f^n) :-
  - It is a logistic f^n where the o/p value varies from -1 to 1.
  - It is also known as ~~Bipolar Sig~~ Hyperbolic Tangent f^n,
  - The o/p centers around 0 & sigmoid around 0.5.



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

⑤ Ramp function :

Relu stands for the rectified linear unit (Relu).
It is most used activation f^n in the world.
It represent o/p '0' for negative values of x.
This is known as ramp f^n.



$$f(x) = \begin{cases} 1 & \text{if } x \\ 0 & \text{if } x < \\ x & \text{if } 0 \le \end{cases}$$

Q. Draw Delta Learning Rule (LMS . widrow Hoff) model & explain it with training process flowchart.

→ • A network with single linear unit is called Adaptive linear neuron (Adaline).

• Adaline neuron can be trained using Delta rule or ~~windrow~~ widrow-hoff least Mean Square (LMS) rule or rule.

• The delta rule is derived from gradient descent method.

• The delta rule update the weights betn the connection so to minimize the differences between the net input to the output unit and the Target value.

• The aim is to minimize the error over all training Patterns.

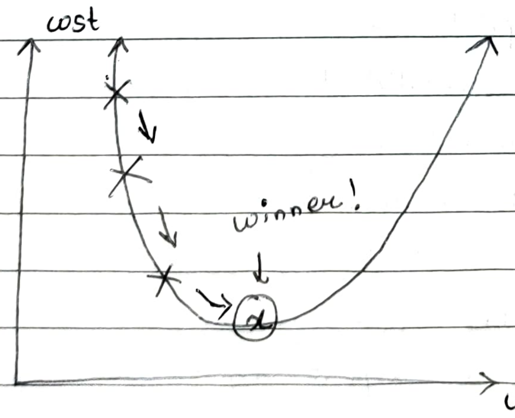
• This is done by reducing the error for each pattern, one at a time.

• The delta rule for adjusting the weight of ith Pattern

$\Delta w_i = \alpha (t - y_{in}) x_i$ OR $\Delta w_i = \alpha T x_i$

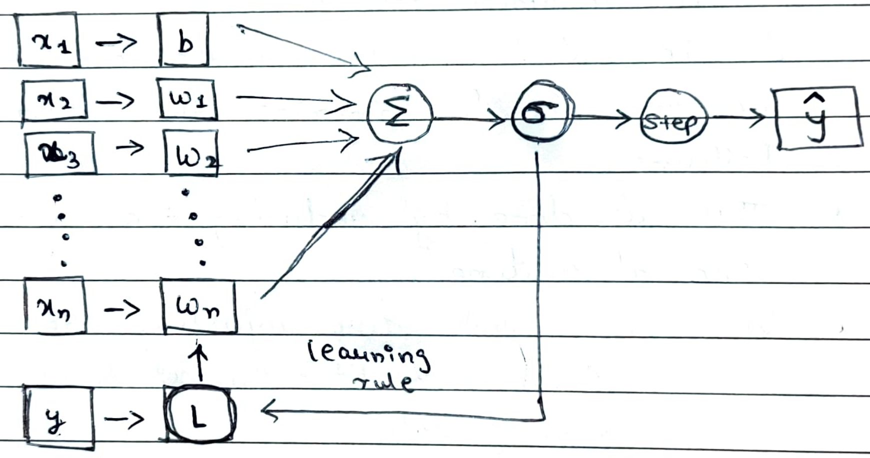
α = learning rate
x = input
y_{in} = Σ

α = learning rate
T = Target
x_i = Input



- The widrow-Hoff learning rule, also known as the Delta learning rule or least Mean square (LMS) rule.
- It is a popular Algorithm used in supervised learning for Training neural networks.
- It is an iterative method that adjust the weights of a network to minimize the mean Squared error betn the network's O/P and the ~~di~~ desired O/P.

Widrow Hoff Algo :-



1. Initialization : Initialize weights randomly.
2. Iterative update :
   - Compute the predicted O/P by taking the dot product of the I/P features and the current weights.
   - Calulate the error betn the predicted O/P & the true O/P.
   - update the weights by adding the Learning rate multiplied by the error & the input features (Gradient descent Step)
3. Convergence : Repeat Step 2 for multiple epochs or until convergence.

Q. Short note :

① Artificial Neural Network :-

- ANN is an information processing model that is inspired by the biological nervous system, such as brain, which process information.
- This model tries to replicate most basic function of the brain.
- ANN is composed of large no. of highly interconnected processing elements working in ~~unis~~ union to solve specific ~~program~~ problem.
- The ANN has 3 main layer :-

① Input layer - The input patterns are fed to the I/P layers. There is one input layer.

② Hidden layer - The hidden layer refines the input by removing redundant information & send the information to the next hidden layer for further processing.

③ Output layer - The hidden layer is connected to the 'output layer' where the output is shown.