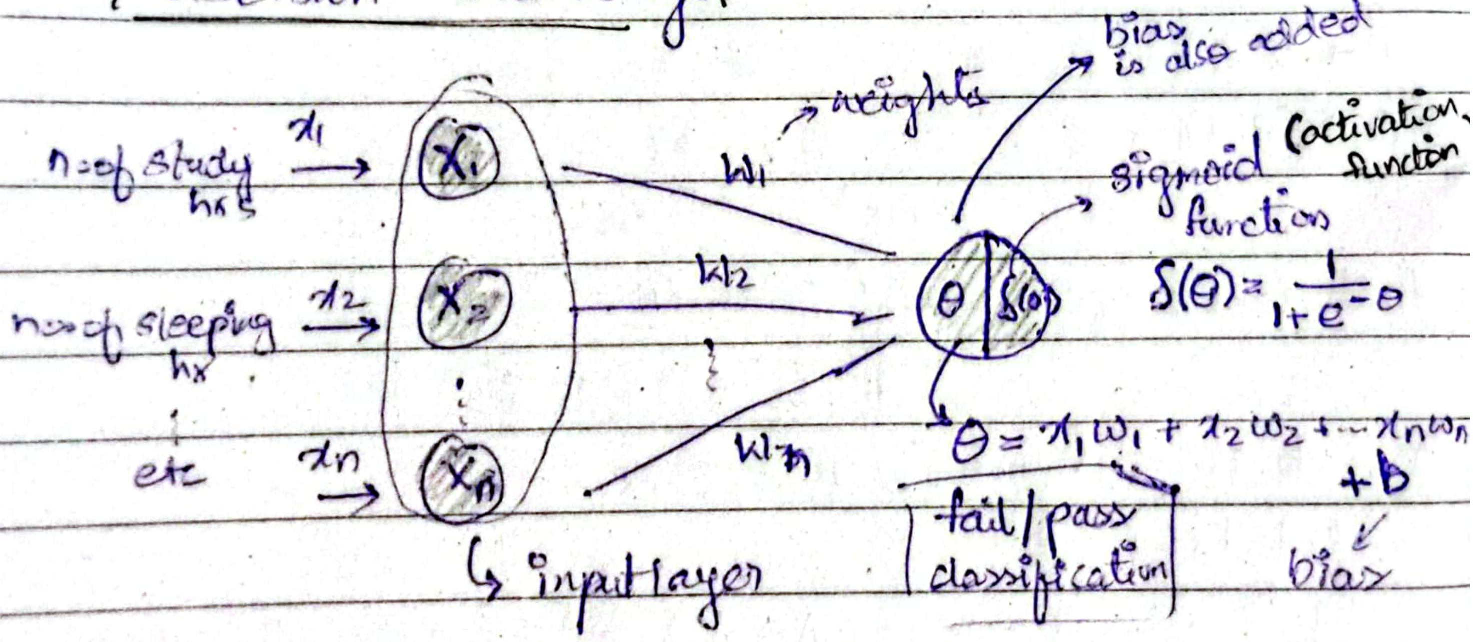
# Neural Network : ANN Feed Forward

## ▷ Single Layer Perceptron

→ It is the simplest form of artifical Neural Network (ANN) and servers as a fundamental building block for understanding more complex networks.

→ It is a linear classifier, it can be used for binary classification based on a linear decision boundary.

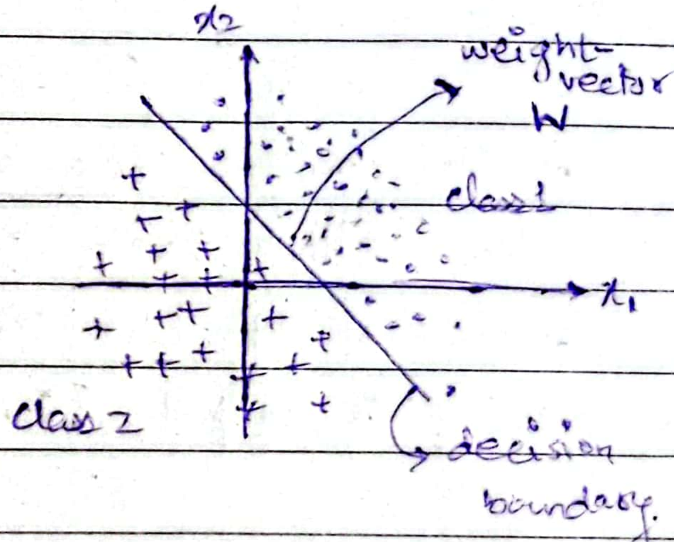



• Components

① Input Layer : A set of input values, each representing a feature of the data

② weights : Each input is assigned a weight that represents it's importance in the decision-making process.

(3) Bias : A constant term added to the weighted sum of inputs, having the model shift the decision boundary.




Comparasion with Naive Bayes:

While Naive Bayes uses a probablistic decision boundary based on the posterior prob of each class

△ The perceptron algo uses a linear decision boundary to seperate the data into classes

NB assumes features are independent

(4) Activation Function

A function that process the sum of weighted inputs and bias to produce an output (the perceptron's decision).

Date ______ Day ______

→ Why Activation Function Needed?

↳ Activation Func introduces non-linearity into the model enabling it to learn complex patterns beyond simple linear relationships.

↳ Without activation Func, a neural network would behive like a basic linear regression model.

↳ Activation func is also used to limit the range of the output. Jese agr simple weights or input ko multiply krein to 10, 20, 1000 kuch bhi answer aaskta hai. Ab agr sigmoid func use krenge to output [0, 1] ki range mein hi aayega.

- Types of Activation Func.
  1. ReLU (Rectified Linear Unit)
  2. Sigmoid → Only this is in our syllabus.
  3. Softmax
  4. Tanh.

Date ______ Day ______

• Sigmoid Activation Function

$$\delta(\theta) = \frac{1}{1+e^{-\theta}}$$

$$\delta'(\theta) = \delta(\theta)\cdot(1-\delta(\theta))$$

Output Range : (0, 1).

Advantage : Usefull in binary classification

Disadvantage : Suffers from Gradient Vanishing Problem.

agr $\delta(\theta)$ ki value boht bari aayega to uske derivative ki value boht small hogi-

let, $\delta(\theta) = 0.994$

$\delta'(\theta) = 0.994(1-0.994)$

$\delta'(\theta) = 0.005964$

Therefore, $\delta(\theta)\uparrow^{big}$ $\delta'(\theta)\downarrow_{small}$

$\delta(\theta)\downarrow_{small}$ $\delta'(\theta)\uparrow^{big}$

• Loss Function

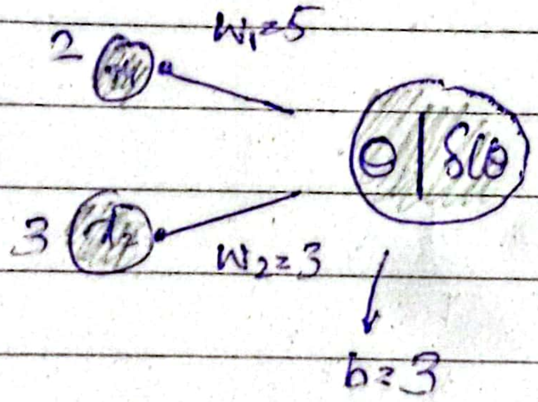
↳ A loss function in a neural network measures how far the predicted output is from actual output

↳ It quantifies the error and helps model improve it's prediction through backpropagation & gradient descent

- Mean Squared Error

→ Squares the error before averaging them
→ Penalizes larger errors more than smaller ones

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (\text{predicted} - \text{actual})^2$$




$\theta = x_1 W_1 + x_2 w_2 + b$

$\theta = 2(5) + 3(3) + 3$

$\theta = 10 + 9 + 3$

$\theta = 22$

$S(\theta) = 1/1+e^{-22} = 0.997$

if actual value = ~~0~~ 1

$$MSE = \frac{1}{1} \left[ (0.997 - 1)^2 \right]$$

$MSE = 7.78 \times 10^{-20}$

Date ______ Day ______

• Multilayer Perceptron (MLP).

→ It is an ANN which consist of at least three layers.

① Input layer
② Hidden layer (one or many)
③ Output layer

↳ Neurons (Node) where computations & learning occurs. Each layer applies activation func

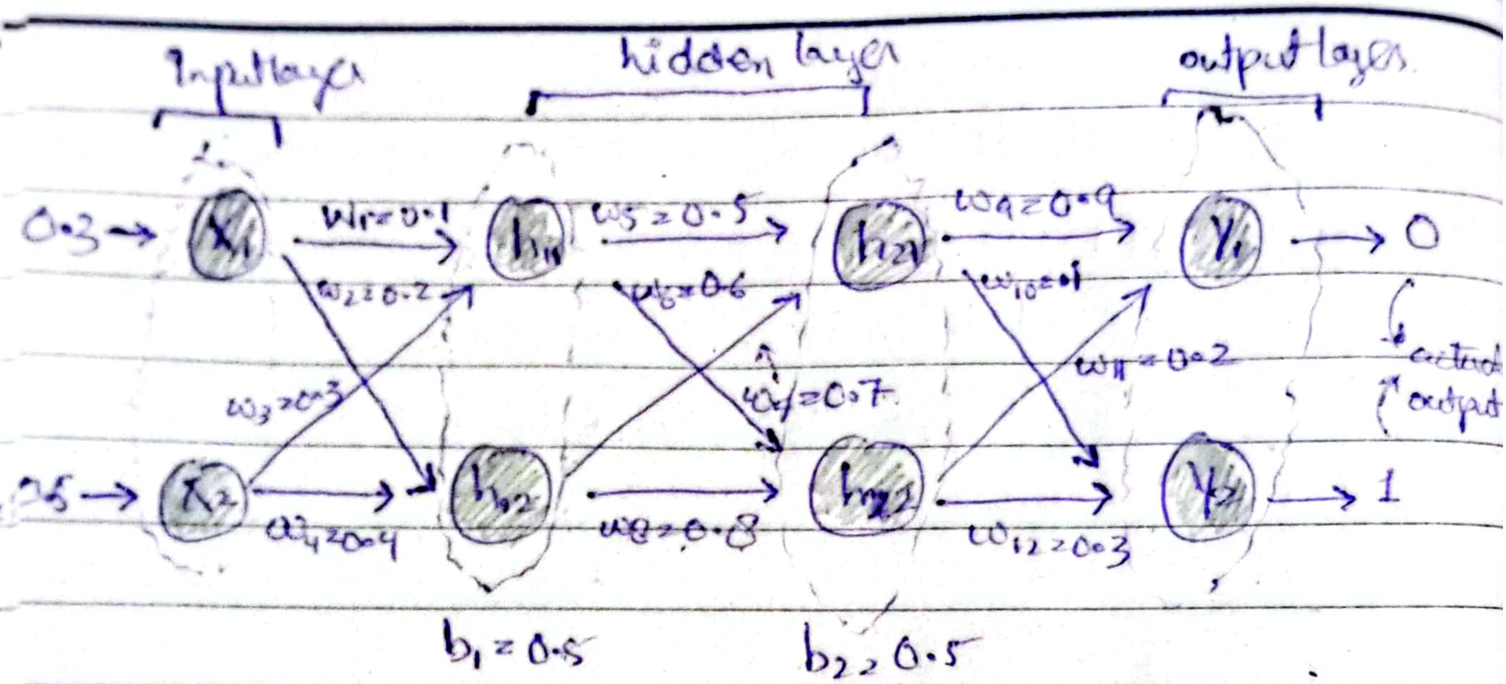
• Feed Forwards

↳ The data flows in one direction from input layer through the hidden layers then to output.

↳ No cycles or loops involved in the network.

↳ Fully Connected: Each neuron in one layer is connected to every neuron in next layer.

Date Day




For $h_{11}$ :

$h_{11} = x_1 w_1 + x_2 w_3 + b_1$

$= 0.3(0.1) + 0.5(0.3) + 0.5$

$h_{11} = 0.68$

$\delta(h_{11}) = 0.66$

For $h_{12}$ :

$h_{12} = x_1 w_2 + x_2 w_4 + b_1$

$= 0.3(0.2) + 0.5(0.4) + 0$

$= 0.76$

$\delta(h_{12}) = 0.68$

For $h_{21}$ :

$h_{21} = \delta(h_{11}) w_5 + \delta(h_{12}) w_7 + b_2$

$= 0.66(0.5) + 0.68(0.7) + 0.5$

$= 1.306$

$\delta(h_{21}) = 0.78$

For $h_{22}$ :

$h_{22} = \delta(h_{11}) w_6 + \delta(h_{12}) w_8 + b_2$

$= 0.66(0.6) + 0.68(0.8) + 0.5$

$= 1.504$

$\delta(h_{22}) = 0.81$

Page No.

Date                                                                 Day

For $y_1$ :

$y_1 = \delta(h_{21})\omega_9 + \delta(h_{22})\omega_{11}$

$= 0.78 \times 0.9 + 0.81 \times 0.2$

$= 0.1636$

$\delta(y_1) = 0.54$

For $y_2$ :

$y_2 = \delta(h_{21})\omega_{10} + \delta(h_{22})\omega_{12}$

$= 0.78(0.1) + 0.81(0.3)$

$y_2 = 0.32$

$\delta(y_2) = 0.58$

MSE :

$$= \frac{1}{2}\left[(0.54-0)^2 + (0.58-1)^2\right]$$

$\boxed{MSE = 0.234 \quad (loss)}$

↳ overall error should be minimum to get optimal weights.