21CSG53

USN | | | | | | | | | | |

NITTE

NITTE MEENAKSHI INSTITUTE OF TECHNOLOGY

(AN AUTONOMOUS INSTITUTE AFFILIATED TO VTU, BELAGAVI)

Fifth Semester Mid-Semester Examination BE Degree (MSE-2)

Academic year 2023 -2024

Department of Computer Science and Engineering

Artificial Intelligence and Machine Learning (21CSG53)

SCHEME & SOLUTION.

Duration: 1 Hr **Max. Marks:30**

Instructions

1. *Q1 (6 Marks) & Q6 (4 Marks) - Compulsory questions.*
2. *Q2 - Q5 Choice-based questions for 10 marks each.*
3. *Missing Data (if any) can be suitably assumed.*

Question No		Question	Marks
Answer the following questions.			
1.	a.	Given the below scenario, determine whether it is an example of supervised or unsupervised learning. i) The task in vision recognition that aims to understand and categorize an image as a whole under a specific label., ii) The process of separating markets or customers into smaller, more manageable groups based on shared characteristics., iii) Financial organizations spot fraudulent transactions, iv) Recommendation systems- The method by which data is collected varies greatly depending on the type of products or services sold. Answer: **Each answer carries 0.5 mark—------------------------0.5*4=2 marks** i) The task in vision recognition that aims to understand and categorize an image as a whole under a specific label- supervised learning ii) The process of separating markets or customers into smaller, more manageable groups based on shared characteristics-Unsupervised learning iii) Financial organizations spot fraudulent transactions,- Supervised learning iv) Recommendation systems- The method by which data is collected varies greatly depending on the type of products or services sold- Unsupervised learning	2

b. Illustrate two main differences between overfitting and underfitting in machine learning models.

Answer:

Any two difference, each difference carries 1 mark———---1*2=2 marks

Overfitting	Underfitting
Overfitting occurs when a model learns the detail and noise in the training data to the extent that it negatively impacts the performance of the model on new data.	Underfitting occurs when a model cannot capture the underlying trend of the data. It neither performs well on the training data nor generalizes to new data.
The model shows high accuracy on training data but poor generalization to new, unseen data.	The model shows low performance on both the training data and unseen data.
Often caused by an excessively complex model with too many parameters.	Typically a result of an overly simple model that fails to capture important regularities in the data.

c. Describe the characteristics and purposes of any two activation functions used in ANNs. 2

Answer:

Each activation function carries 1 mark——-----------------1*2=2 marks

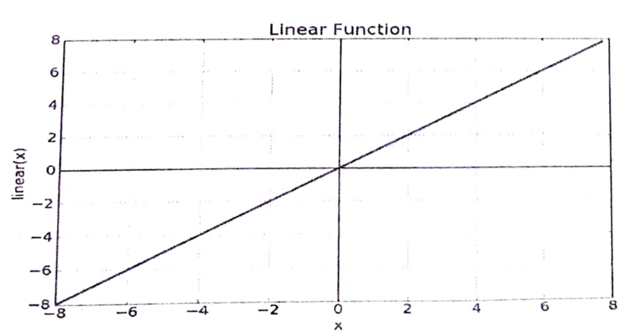
A. Identity Function: Identity function is used as an activation function for the input layer. It is a linear function having the form

$$y_{out} = f(x) = x, \forall x$$

As obvious, the output remains the same as the input.

B. Threshold/step Function: It is a commonly used activation function. As depicted in the diagram, it gives **1 as output** of the input is either 0 or positive. If the input is negative, it gives **0 as output**. Expressing it mathematically,

$$y_{out} = f(y_{sum}) = \begin{cases} 1, x \geq 0 \\ 0, x < 0 \end{cases}$$



The Sigmoid Function curve looks like a S-shape.

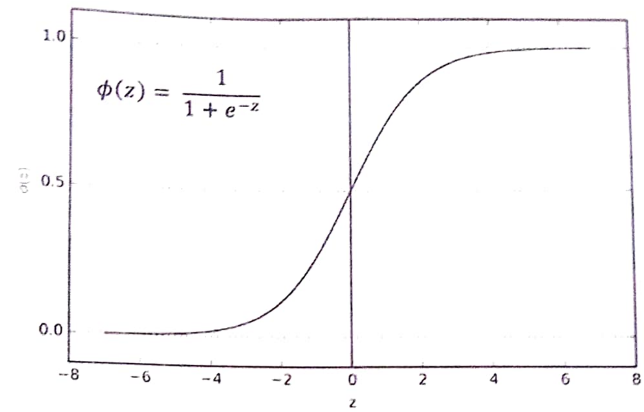


Fig: Sigmoid Function

UNIT-I

2 a. (5 marks)

An optimized model will be sensitive to the patterns in the data, but at the same time will be able to generalize to new data. In this, **both the bias and variance should be low.** Show with equations how bias, variance, and the bias-variance tradeoff are calculated and applied in machine learning.

Answer: **Bias explanation with formula—---------1.5 mark**

Variance explanation with formula——---1.5mark

bias-variance tradeoff explanation with diagram——----2mark

Bias is the difference between the Predicted Value and the Expected Value. To explain further, the model makes certain assumptions when it trains on the data provided. When it is introduced to the testing/validation data, these assumptions may not always be correct.

- Mathematically, let the input variables be X and a target variable Y. We map the relationship between the two using a function f.
- Therefore,
- Y = f(X) + e
- Here 'e' is the error that is normally distributed. The aim of our model f'(x) is to predict values as close to f(x) as possible. Here, the Bias of the model is:
- Bias[f'(X)] = E[f'(X) – f(X)]

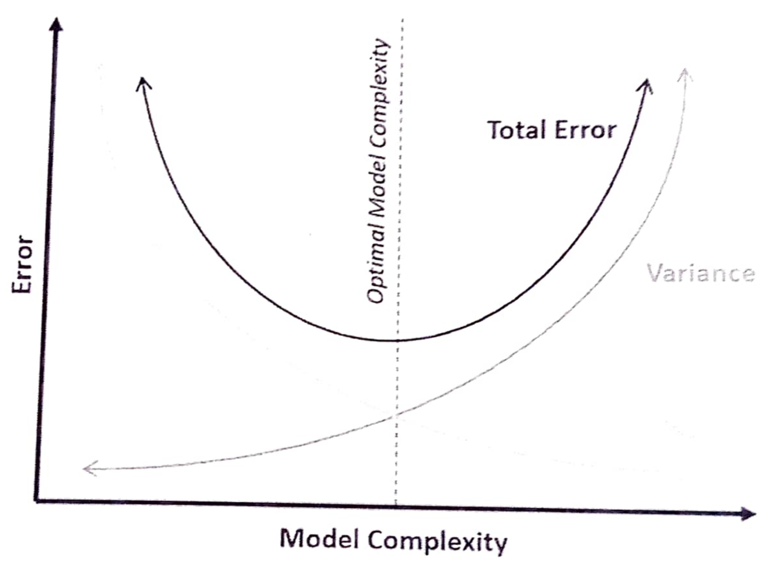
Variance, on the other hand, reflects the model's sensitivity to small fluctuations in the training data. High variance can lead to overfitting, where the model captures noise in the training data and performs poorly on new, unseen data.

Mathematically, the variance error in the model is:

Variance[f(x))=E[X^2]−E[X]^2

Since in the case of high variance, the model learns too much from the training data, it is called overfitting.

The bias-variance tradeoff is a fundamental concept in machine learning and statistics. It refers to the delicate balance between two sources of error in a predictive model: bias and variance.



A model with low bias and high variance predicts points that are around the center generally, but pretty far away from each other. A model with high bias and low variance is pretty far away from the bull's eye, but since the variance is low, the predicted points are closer to each other.

In terms of model complexity, we can use the following diagram to decide on the optimal complexity of our model.

b. Working of Naïve Bayes' Classifier can be understood with the help of the below example: Suppose we have a dataset of **weather conditions** and corresponding target variable "**Play**". 5

Problem: If the weather is sunny, then the Player should play or not, using Naïve Bayes' Classifier.

USN

	Outlook	Play
0	Rainy	Yes
1	Sunny	Yes
2	Overcast	Yes
3	Overcast	Yes
4	Sunny	No
5	Rainy	Yes
6	Sunny	Yes
7	Overcast	Yes
8	Rainy	No
9	Sunny	No
10	Sunny	Yes
11	Rainy	No
12	Overcast	Yes
13	Overcast	Yes

Answer: formula for Baye's theorem

$$P(yes|Sunny) = \frac{P(Sunny|yes) \times P(yes)}{P(sunny)}$$

$$P(No|Sunny) = \frac{P(Sunny|No) \times P(No)}{P(Sunny)}$$

1 M.

Frequency table for Weather Condition

Weather	Yes	No
Overcast	5	0
Rainy	2	2
Sunny	3	2
Total	10	5

Likelihood table weather Condition:

Weather	No	Yes	
Overcast	0	5	5/14 = 0.35
Rainy	2	2	4/14 = 0.29
Sunny	2	3	5/14 = 0.35
All	4/14 = 0.29	10/14 = 0.71	

1M

Applying Baye's theorem

$P(Sunny|yes) = 3/10 = 0.3$

$P(Sunny) = 0.35$

$P(yes) = 0.71$

$P(yes|Sunny) = \frac{0.3 * 0.71}{0.35} = 0.6.$

1M

$P(Sunny|No) = 2/4 = 0.5$

$P(No) = 0.29.$

$P(Sunny) = 0.35$

$P(No|Sunny) = \frac{0.5 * 0.29}{0.35} = 0.41$

1M

from above calculation,

$P(Yes|Sunny) > P(No|Sunny).$

Hence on Sunny day, player can play the game. 1M.

OR

3 a. Apply your understanding of linear and logistic regression to identify which would be more suitable for a data set and justify. (5)

Answer:

consider the following aspects:

1. **Nature of the Dependent Variable:**
   - Use linear regression when the dependent variable is continuous and can take any real value.
   - Use logistic regression when the dependent variable is binary or categorical (e.g., 0/1, Yes/No).

   1 M
2. **Type of Relationship:**
   - Linear regression assumes a linear relationship between the independent and dependent variables.
   - Logistic regression models the log-odds of the probability of an event occurring, allowing for a non-linear relationship.
3. **Distribution of Residuals:**
   - Linear regression assumes normally distributed residuals.
   - Logistic regression does not require normal distribution of residuals; it deals with the log-odds.

   1 M
4. **Assumptions:**
   - Linear regression assumes homoscedasticity and independence of errors.
   - Logistic regression doesn't require homoscedasticity but assumes independence of errors.

   1 M
5. **Model Interpretability:**
   - Linear regression provides coefficients that represent the change in the dependent variable for a one-unit change in the independent variable.
   - Logistic regression provides odds ratios, which represent the change in the odds of an event occurring for a one-unit change in the independent variable.

   1 M

Justification:

- If your dependent variable is continuous and you're interested in predicting a quantity, linear regression would be more appropriate.
- If your dependent variable is binary or categorical, and you want to model the probability of an event occurring, logistic regression is more suitable.

1 M

b. An email service uses a spam detection algorithm. The following confusion matrix represents its performance on a test dataset (5)

	Predicted Spam	Predicted Not Spam
Actual Spam	120	30
Actual Not Spam	40	310

a) Calculate the precision of the spam detection model and discuss the importance of precision in the context of spam detection.
b) Determine the recall of the spam detection model and explain why recall is a critical metric.
c) Calculate the F-score and Discuss the significance of the F-score
d) Calculate the classification accuracy of the model

The different values of the Confusion matrix would be as follows:

- True Positive (TP) = 120, meaning the model correctly classified 120 positive class data points.
- True Negative (TN) = 310, meaning the model correctly classified 310 negative class data points.
- False Positive (FP) = 30, meaning the model incorrectly classified 30 negative class data points as belonging to the positive class.
- False Negative (FN) = 40, meaning the model incorrectly classified 40 positive class data points as belonging to the negative class.

1M.

$$Precision = \frac{TP}{TP + FP} = \frac{120}{120+30} = \frac{120}{150}$$

1M

$$\frac{4}{5} = 0.2 \text{ or } 20\%$$

$$Recall = \frac{TP}{TP + FN} = \frac{120}{120+40} = \frac{120}{160} = 0.75$$

1M

$$= 75\%$$

$$F1 - score = \frac{2}{\frac{1}{Recall} + \frac{1}{Precision}} = 2 * \left[\frac{Precision * recall}{Precision + Recall}\right]$$

USN

$$F\text{-score} = 2 * \left[\frac{0.2 * 0.75}{0.2 + 0.75}\right] = 0.315$$ 1M

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} = \frac{120 + 310}{120 + 310 + 30 + 40}$$

$= 0.86.$

$= 86\%.$ 1M.

UNIT-II

4 a. Artificial Neural Networks (ANNs) can learn and model non-linear and complex relationships, which is really important because in real-life. Describe neural network learning with its characteristics and any four advantages. 5

Answer:

Explanation of 3 characteristics—-------3 marks
Advantages—--------------------------------2 marks

An artificial neuron is characterized by:

1. Architecture (connection between neurons)
2. Training or learning (determining weights on the connections)
3. Activation function

Other advantages include:

1. *Adaptive learning:* An ability to learn how to do tasks based on the data given for training or initial experience.
2. *Self-organization:* An ANN can create its own organisation or representation of the information it receives during learning time.
3. *Real-time operation:* ANN computations may be carried out in parallel, using special hardware devices designed and manufactured to take advantage of this capability.
4. *Fault tolerance via redundant information coding:* Partial destruction of a network leads to a corresponding degradation of performance. However, some network capabilities may be retained even after major network damage due to this feature.

	b.	Implement OR function with bipolar input and target using perceptron learning rule.	5

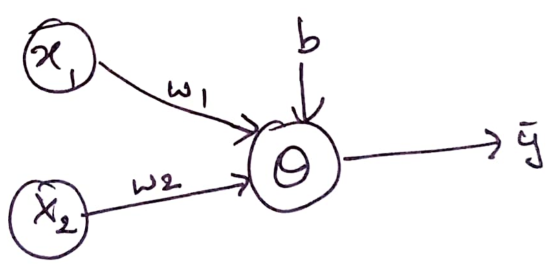
OR function truth table for bipolar input

x_1	x_2	y
$-\phi$	$-\phi$	$-\phi$
$-\phi$	1	1
1	$-\phi$	-1
1	1	1

$w = (w_1, w_2)$ weight vector of i/p

$x = (x_1, x_2)$

Perceptron function can be defined as, $\hat{y} = \Theta(w_1 x_1 + w_2 x_2 + b)$ 1M



for the implementation, consider the weight parameter

$w_1 = 2$, $w_2 = 2$ and bias parameter $b = \cancel{(0.5)}\ 1$. 1M

USN

$x_1 = 1, \quad x_2 = 1$

$w_1 x_1 = 2 * 1 = 2, \quad w_2 x_2 = 2 \times 1 = 2, \; b = 1$

$y = 2 + 2 + 1 = 5 \leq 0.$

$\boxed{y = 1}$

~~wx+b≤0~~

$x_1 = 1, \; x_2 = -1, \; b = 1, \; w_1 = w_2 = 2$

$y = 2 \times 1 + 2 \times (-1) + 1$

$= 2 - 2 + 1 = 1 \leq 0.$ 1M

$\boxed{y = 1}$

$x_1 = -1, \; x_2 = 1, \; b = 1, \; w_1 = w_2 = 2$

$y = 2 \times (-1) + 2 \times 1 + 1$

$= -2 + 2 + 1 = 1 \leq 0.$

$\boxed{y = 1}$

$x_1 = -1, \; x_2 = -1, \; b = 1, w_1 = w_2 = 2$ 1M

$y = 2 \times (-1) + 2 \times (-1) + 1$

$= -2 - 2 + 1 = -3 \nleq 0.$

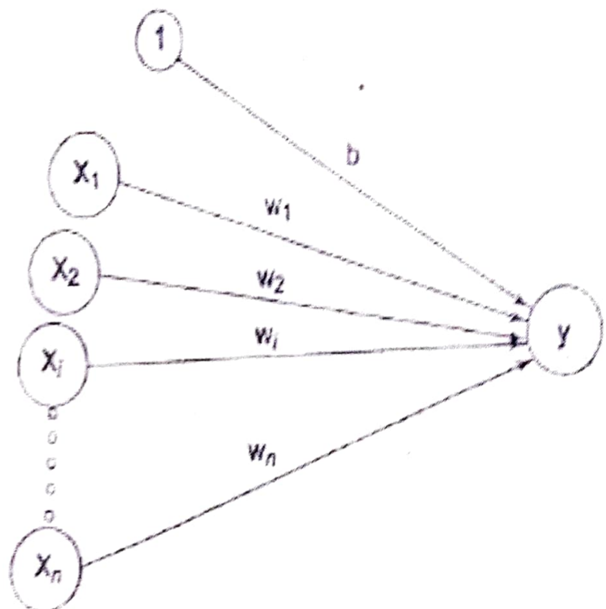
$\boxed{y = -1}$

y values satisfies truthtable 1M

So, $w_1 = w_2 = 2, \; b = 1$

OR

5	a.	Describe the architecture of a single layer perceptron with a neat diagram. **diagram-----------------------2 marks** **Explanation-------------------3 marks**	5



The architecture of the single layer perceptron is shown in Fig. 4.2.

As we have already studied, the perceptron has sensory, associator and response units. The input to the response unit will be the output from the associator unit, which is a binary vector. Since only the weight between the associator and the response unit is adjusted, the concept is limited to single layer network as discussed in Section 2.7.

In the architecture shown in Fig. 4.2, only the associator unit and the response unit is shown. The sensor unit is hidden, because only the weights between the associator and the response unit are adjusted. The input layer consists of input neurons from $X_1 \cdots X_i \cdots X_n$. There always exists a common bias of '1'. The input neurons are connected to the output neurons through weighted interconnections. This is a single layer network because it has only one layer of interconnections between the input and the output neurons. This network perceives the input signal received and performs the classification.

USN

b. Apply the Hebb net to the AND function with bipolar input and targets. Justify the separating line can/can not be drawn. 5

Solution The AND function gives a high '1' if both the inputs are high else returns a low '– 1'. The training patterns are,

1M

Input			Target
x_1	x_2	B	y
1	1	1	1
1	– 1	1	– 1
– 1	1	1	– 1
– 1	– 1	1	– 1

Forming the table, initialize all the weights and the bias to be zero i.e.

$$w_1 = w_2 = 0 \text{ and } b = 0.$$

The weight change is calculated using,

$$\Delta w_i = x_i\, y \text{ and } \Delta b = y.$$

Input			Target	Weight Changes			Weights		
(x_1	x_2	b)	y	Δw_1	Δw_2	Δb	w_1	w_2	B
							Initial (0	0	0)
1	1	1	1	1	1	1	1	1	1
1	– 1	1	– 1	– 1	1	– 1	0	2	0
– 1	1	1	– 1	1	– 1	– 1	1	1	– 1
– 1	– 1	1	– 1	1	1	– 1	2	2	– 2

2M

This completes one epoch of training. The straight line separating the regions can be obtained after pro ing each input pair. Thus,

$$x_2 = -x_1 \frac{w_1}{w_2} - \frac{b}{w_2}$$

After 1st input, $x_2 = -x_1 \frac{1}{1} - \frac{1}{1} = -x_1 - 1$

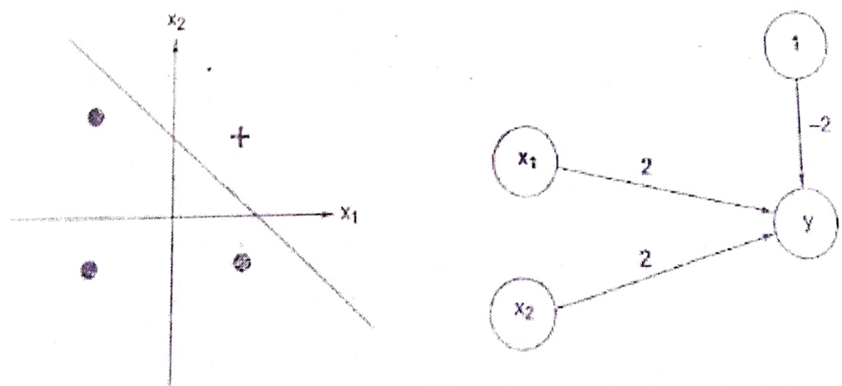
$$x_2 = -x_1 - 1$$

1 M

Similarly after 2nd, 3rd and 4th epochs, the separating lines are,

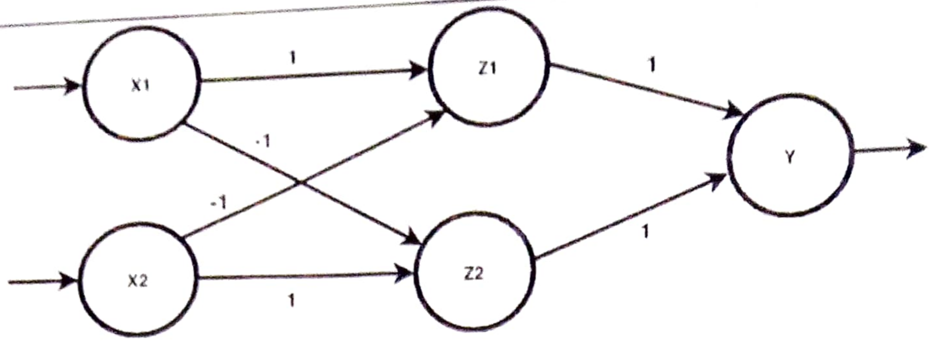
$$x_2 = 0, \quad x_2 = -x_1 + 1, \quad x_2 = -x_1 + 1$$

For the 3rd and 4th epoch the separating line remains the same, hence this line separates the boundary regions as shown in Fig. 3.9.



1M

6 a. Realize the XOR function using McCulloch-Pitts neuron for given diagram below. Assume threshold =1 4



Answer:
Activation function—------------1 mark
Z1 calculation—-----------------1 mark
Z2 calculation—-----------------1 mark
Z1or Z2 calculation—-----------1 mark

Solution XOR function returns a true value if exactly one of the input values is true; otherwise it returns t response as false. The truth table for XOR function is,

x_1	x_2	y
1	1	0
1	0	1
0	1	1
0	0	0

The McCulloch-Pitts neuron model for this is given in Fig 3.6. The threshold of unit y is 1.

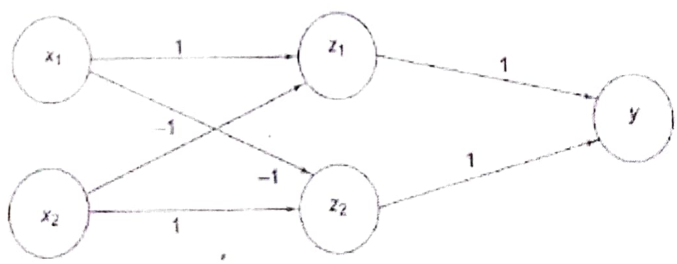


Fig. 3.6 McCulloch-Pitts Neuron for XOR Function

With one layer alone, it was not able to predict the value of the threshold for the neuron to fire, he another layer is introduced.

$$x_1 \text{ XOR } x_2 = (x_1 \text{ ANDNOT } x_2) \text{ OR } (x_2 \text{ ANDNOT } x_1)$$

$$x_1 \text{ XOR } x_2 = z_1 \text{ OR } z_2$$

where, $z_1 = x_1 \text{ ANDNOT } x_2$

and $z_2 = x_2 \text{ ANDNOT } x_1$

The activations of z_1 and z_2 are given as,

$$z_1 = (z_{in-1}) = \begin{cases} 1 & \text{if} \quad z_{in-1} \geq 1 \\ 0 & \text{if} \quad z_{in-1} < 1 \end{cases}$$

$$z_2 = (z_{in-2}) = \begin{cases} 1 & \text{if} \quad z_{in-2} \geq 1 \\ 0 & \text{if} \quad z_{in-2} < 1 \end{cases}$$

The calculation of net input and activations of z_1 and z_2 are shown belc

$z_1 = (x_1 \text{ ANDNOT } x_2)$ $\quad z_{in-1} = x_1 w_1 + x_2 w_2$

x_1	x_2	z_{in-1} $w_1 = 1, w_2 = -1$	z_1
1	1	0	0
1	0	1	1
0	1	−1	0
0	0	0	0

$z_2 = (x_2 \text{ ANDNOT } x_1)$ $\quad z_{in-2} = x_1 w_1 + x_2 w_2$

x_1	x_2	z_{in-2} $w_1 = -1, w_2 = 1$	z_2
1	1	0	0
1	0	−1	0
0	1	1	1
0	0	0	0

The activation for the output unit y is 1.

$$y = f(y_{in}) = \begin{cases} 1 & \text{if} \quad y_{in} \geq 1 \\ 0 & \text{if} \quad y_{in} < 1 \end{cases}$$

Presenting the input patterns (z_1 and z_2) and calculating net input and activations gives output of XOR.

Here, $y_{in} = z_1 w_1 + z_2 w_2$

z_1	z_2	y_{in} $w_1 = 1, w_2 = 1$	$y = z_1$ or z_2
0	0	0	0
1	0	1	1
0	1	1	1
0	0	0	0

Thus the Exclusive-OR function is realized.

Faculty Signature 10/1/24

HOD Signature 10/1/24.