Assignment-3&4

1) What do you mean by Artificial Neural Network (ANN) and write down some applications of Artificial Neural Network?

Ans→ An Artificial Neural Network (ANN) is a computational model inspired by the way biological neural networks in the human brain process information. It is a part of machine learning.

An ANN consists of multiple layer→

i) Input layer → Accepts the input data.

ii) Hidden layer → Layers between the input output that process the data through various transformation.

iii) Output layer→ Produces the final result, such as a prediction or classification.

The applications of Artificial Neural Network (ANN)→

i) Image Recognition → Identifying objects in images.

ii) Natural Language Processing (NLP) → Used in speed recognition, language, translation and chat bots.

iii) Autonomous Vehicles → Self-driving cars use ANNs for object recognition and decision-making.

Q2) Give some idea about the advantages and limitations of Artificial Neural Network.

Ans→ Advantages of Artificial Neural Network (ANN)→

i) Learning from data → Can learn patterns from large datasets without explicit programming.

ii) Flexibility → Suitable for diverse tasks like classification, regression, and pattern recognition.

iii) Fault tolerance → Can handle noisy or incomplete data.

Limitations of Artificial Neural Network→

i) Data hungry → Requires large amount of high-quality data.

ii) Computationally Expensive → Needs significant computational power and time.

iii) Low of interpretability → Hard to understand how decisions are made.

Q3) What is Deep learning? Differentiate between Machine Learning and Deep learning.

Ans→ Deep learning is a subset of machine learning that uses multi-layered neural networks to automatically learn features from large datasets, exelling in tasks like image recognition and language translation. It requires significant data & computational power.

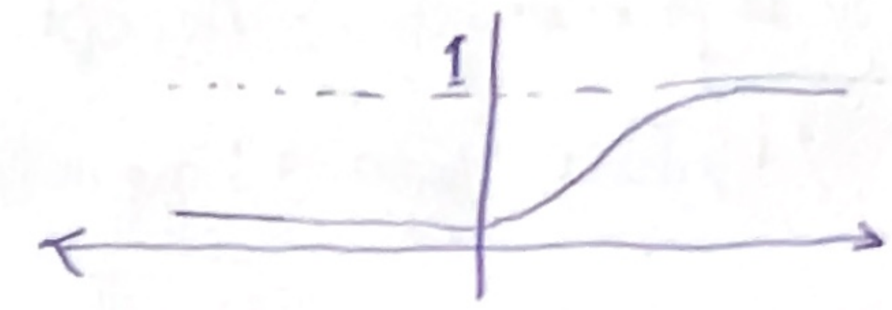
Machine Learning	Deep Learning
i) ML is a subset of AI where algorithms learn from data and improve over time	i) DL is a subset of ML that use neural networks with multiple layers to learn from large datasets

Name:__________________ Regd No.:________________

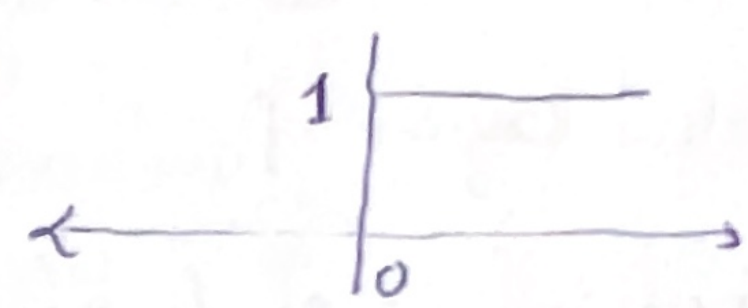
ii) Works well with smaller data sets.	ii) Requires large datasets to perform effectively.
iii) Requires manual feature extraction and section.	iii) Automatically learns features from raw data.
iv) Less computationally intensive.	iv) Requires significant computational resources often with GPUs or TPUs.

Q)4) Describe four different types of activation functions.

Ans → i) Logistic Function → $F(z) = \frac{1}{1+e^{-z}}$, $f(z) = R \to (0,1)$

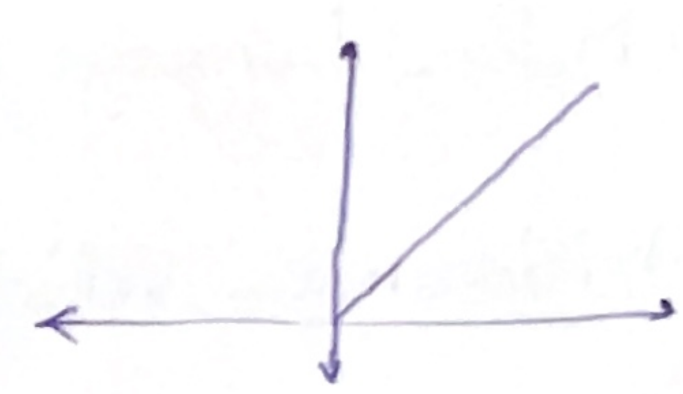


ii) Step function → $f(z) = \begin{cases} 0, & \text{if } z<0 \\ 1, & \text{if } z \geq 0 \end{cases}$



iii) Rectified linear unit function or ReLU (z) →

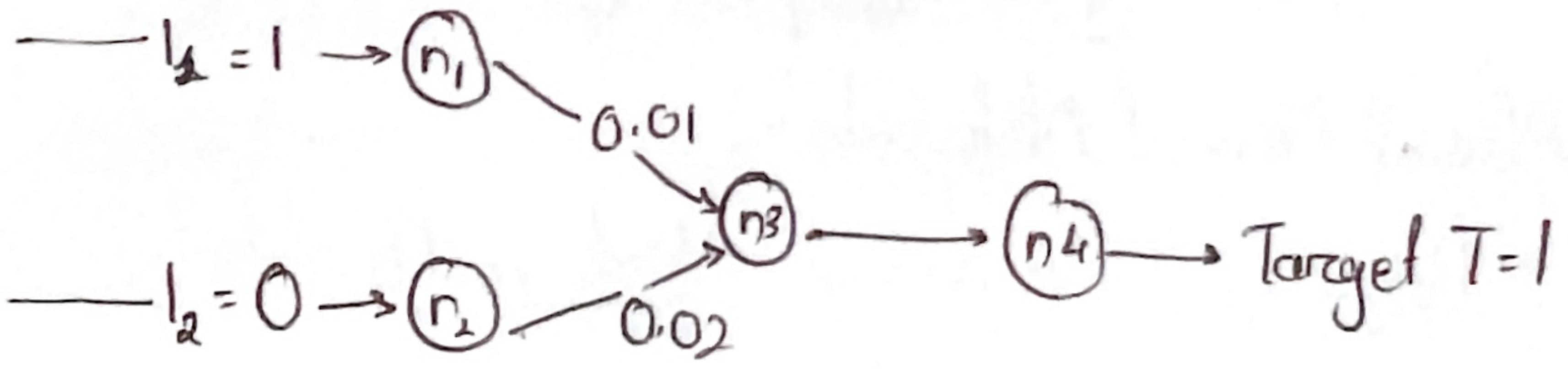
$f(z) = \begin{cases} 0, & \text{if } z<0 \\ z, & \text{if } z \geq 0 \end{cases}$ Range $(0 + 0\ z^{+})$ or R^{+}



iv) Hyperbolic target function →

$\tanh(z) = \frac{e^{z}-e^{-z}}{e^{z}+e^{-z}}$

5) Consider a simple neural network as shown in Figure. The inputs and initial weights are also shown in Figure. Target of the given neural network is T=1. Use back propagation to train the network. The activation function is sigmoid function. Using the learning rate = 0.3.



a) Perform the forward pass and calculate the predicted output.

b) Find out the error of the output layer.

c) Perform the backward pass and update the weights.

Ans → a) $I_1 = 1,\ I_2 = 0,\ \omega = 0.01,\ \omega_2 = 0.02,\ \omega_3 = 0.03,\ T=1$

Net inputs of n3 $= 1 \times 0.01 + 0 \times 0.02 = 0.01$

Output of n3 $= \frac{1}{1+e^{-net\,n3}} = \frac{1}{1+e^{-0.01}} = 0.5025$

Net inputs of n4 $= 0.5025 \times 0.03 = 0.015$

Output of n4 $= 1/1+e^{-0.015} = 0.5038$

b) Error $= \frac{1}{2}(1-0.8038)^2 = 0.1231$

c) Learning rate $(n) = 0.3$

$$w_3\ new = \frac{\partial E_{total}}{\partial w_3} = \frac{\partial E_{total}}{\partial out_{n4}} \times \frac{\partial out_{n4}}{\partial net_{n4}} = \frac{\partial net_{n4}}{\partial w_3}$$

$$\frac{\partial E_{total}}{\partial out_{n4}} = (out_{n4} - Target_{n4}) = 0.5038 - 1 = -0.4962$$

$$\frac{\partial out_{n4}}{\partial net_{n4}} = out_{n4}(1 - out_{n4}) = 0.5038(1-0.5038) = 0.2499$$

$$\simeq 0.25$$

$$\frac{\partial net_{n4}}{\partial w_3} = out_{n3} = 0.5025$$

$$w_3\ new = w_5 - \left(n \times \frac{\partial E_{total}}{\partial w_3}\right) = 0.03 - (0.3 \times -0.4962 \times 0.25 \times 0.5025)$$

$$= 0.03 - (-0.0187)$$

$$= 0.0487.$$

w_1 new

$\Delta w = n \times \delta n_3 \times I$

$\delta(n_3) = out_{n_3} \times (1 - out_{n_3}) = 0.2499 \approx 0.25$

I = Input $I_1 = 1$

$\delta n_3 = 0.124 \times 0.3 \times 0.2499 = 0.000093$

$\Delta w = 0.3 \times 0.00093 \times 1 = 0.00028$

$w_1 new = 0.01 + 0.00028 = 0.01028$

w_2 new

$I_2 = 0$

$\Delta w = n = \delta n_3 \times I_2 = 0$

So, no change $w_2 = new$.

Q6) What are the applications of SVD in machine learning and artificial intelligence?

Ans → Singular value decomposition (SVD) is a popular image transmation technique. It has lot of applications in dimmensionally reduction, image restoration, image compression and object recognization. SVD transform a given matrix A to be factor into matrices U, Σ, V^T.

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

Name:____________ Regd No.:____________

U = Normalised Eigen Values matrix of AA^T.

V^T = Transpose of Normalised Eigen values matrix of A^TA.

Q7) Using singular value decomposition, find the matrices U, V and Σ for the matrix $A = \begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{bmatrix}$

Ans→ Using SVD, we know that $A = U\Sigma V^T$.

For U, AA^T

$$AA^T = \begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{bmatrix} \begin{bmatrix} 3 & 2 \\ 2 & 3 \\ 2 & -2 \end{bmatrix} = \begin{bmatrix} 17 & 8 \\ 8 & 17 \end{bmatrix}$$

$\lambda_1 = 25$, $\lambda_2 = 9$

$(x - \lambda I)\begin{pmatrix} P_1 \\ P_2 \end{pmatrix} = 0$

For $\lambda_1 = 25$

$$\begin{bmatrix} 17 & 8 \\ 8 & 17 \end{bmatrix} - \begin{bmatrix} 25 & 0 \\ 0 & 25 \end{bmatrix} \cdot \begin{pmatrix} P_1 \\ P_2 \end{pmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow \begin{bmatrix} -8 & 8 \\ 8 & -8 \end{bmatrix}\begin{pmatrix} P_1 \\ P_2 \end{pmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$-8P_1 + 8P_2 = 0$

$\Rightarrow 8P_1 = 8P_2$

$\Rightarrow P_1 = P_2$

Eigen vector for $\lambda_1 = 25 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

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

For $(\lambda_2 = 9)$

$$\begin{bmatrix} 17 & 8 \\ 8 & 17 \end{bmatrix} \begin{bmatrix} 9 & 0 \\ 0 & 9 \end{bmatrix} \cdot \begin{bmatrix} P_1 \\ P_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow \begin{bmatrix} 8 & 8 \\ 8 & 8 \end{bmatrix}\begin{pmatrix} P_1 \\ P_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$8p_1 + 8p_2 = 0$

$x_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$

$\Rightarrow P_2 = -P_1$

Normalise eigen vector $= \begin{pmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$

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

$\Sigma \to \sigma_1 = \sqrt{\lambda_1} = \sqrt{25} = 5$

$\sigma_2 = \sqrt{\lambda_2} = \sqrt{9} = 3$

$\Sigma = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix}$

$V = A^T A$

Eigen values are $\lambda_1 = 25, \lambda_2 = 9, \lambda_3 = 0$

$$A^TA = \begin{bmatrix} 3 & 2 \\ 2 & 3 \\ 2 & -2 \end{bmatrix} \begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{bmatrix} = \begin{bmatrix} 13 & 12 & 2 \\ 12 & 13 & -2 \\ 2 & -2 & 8 \end{bmatrix}$$

For $\lambda_1 = 25$ $(x - \lambda_1 I)\begin{pmatrix} P_1 \\ P_2 \\ P_3 \end{pmatrix} = 0$

$$\begin{bmatrix} -12 & 12 & 2 \\ 12 & -12 & -2 \\ 2 & -2 & -17 \end{bmatrix}\begin{pmatrix} P_1 \\ P_2 \\ P_3 \end{pmatrix} = 0$$

$-12p_1 + 12p_2 + 2p_3 = 0$

$\Rightarrow 12p_1 - 12p_2 - 2p_3 = 0 - (1)$

$\Rightarrow 2p_1 - 2p_2 - 17p_3 = 0 - (3)$

$P_3 = 0$

$P_1 = P_2$

$x_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$ $N(x_1) = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{pmatrix}$

For $\lambda_2 = 9$

$$\begin{bmatrix} 4 & 3 & -7 \\ 3 & 4 & -11 \\ -7 & -11 & -1 \end{bmatrix}\begin{pmatrix} P_1 \\ P_2 \\ P_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$4P_1 + 3P_2 - 7P_3 = 0 - (1)$ $3P_1 + 4P_2 - 11P_3 = 0 - 2$

$-7P_1 - 11P_2 - P_3 = 0$

$\Rightarrow P_3 = 7P_1 - 11P_2$

$\Rightarrow -7P_2 - 65P_3 = 0$

$\Rightarrow -7P_2 = 65P_3$

$\Rightarrow P_2 = \frac{-65}{7}P_3$ $P_1 = \frac{61}{7}P_3$

$x_2 = \begin{bmatrix} 61/7 \\ -65/7 \\ 1 \end{bmatrix}$ $N(x_2) = \begin{bmatrix} 61/89 \\ -65/89 \\ 0 \end{bmatrix}$

For $\lambda_3 = 0$ $\begin{bmatrix} 13 & 12 & 2 \\ 12 & 13 & -2 \\ 2 & -2 & 8 \end{bmatrix}\begin{pmatrix} P_1 \\ P_2 \\ P_3 \end{pmatrix} = 0$

$P_1 = P_2$ $P_3 = 0$ $x_3 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$ $N(x_3) = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{pmatrix}$

$$V = \begin{bmatrix} 1/\sqrt{2} & 61/89 & 1/\sqrt{2} \\ 1/\sqrt{2} & -65/89 & 1/\sqrt{2} \\ 0 & 0 & 0 \end{bmatrix}$$

Name:______________ Regd No.:______________

Q8) What is Principal Component Analysis and how does it work?

Ans→ It works on the condition that while the data in a higher dimension space is mapped to data in a lower dimension space the variance of the data in the lower dimensional space should be maximum.

- It is a statical produces that uses an orthogonal transaction that consists of a set of co-related variable to set of an co-related variable.
- It is an unsupervised learning algorithm used to examine the interrelation between the set of relations.
- The main goal of principal component analysis is to reduce the dimensity of data set.

Q9) Given data x_1: {2, 3, 4, 5, 6, 7} and x_2: {1, 5, 3, 6, 7, 8}. Compute the principal component using PCA algorithm.

Ans→ $\bar{x} = \frac{2+3+4+5+6+7}{6} = 4.5$, $\bar{y} = \frac{1+5+3+6+7+8}{6} = 1.66$

Here, the co-variance matrix is.

$$S = \begin{bmatrix} cov(x, x) & cov(x, y) \\ cov(y, x) & cov(y, y) \end{bmatrix}$$

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

$$= \frac{1}{6-1} [(2-4.5)^2 + (3-4.5)^2 + (4-4.5)^2 + (5-4.5)^2 + (6-4.5)^2 + (7-4.5)^2]$$

$$= 1/5 (2.25 + 0.25 + 0.25 + 0.25 + 2.25 + 6.25)$$

$$= \frac{11.5}{5} = 2.3$$

$$cov(x, y) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

$$= 1/5 [(2-4.5)(1-1.66) + (3-4.5)(5-4.5) + (4-4.5)(3-1.66) + (5-4.5)(6-1.66) + (6-4.\cancel{50})(7-1.6) + (7-4.5)(8-1.66)]$$

$$= 1/5 [(-2.5)(-0.66) + (-0.5)(3.34) + (0.5)(1.34) + (0.5)(4.34) + (1.5)(5.34) + (2.5)(6.34)]$$

$$= 1/5 (26.87) = 5.374.$$

$$cov(y, x) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = 5.375$$

$$cov(y, y) = 20.18672$$

Name:________________ Regd No.:________________

Hence, $S = \begin{bmatrix} 2.3 & 5.374 \\ 5.374 & 20.18672 \end{bmatrix}$

$|S - \lambda I| = 0$

$\left| \begin{bmatrix} 2.3-\lambda & 5.374 \\ 5.374 & 20.1867 \end{bmatrix} \right| = 0$

$\Rightarrow (2.3-\lambda)(20.1867-\lambda) - (5.374)^2 = 0$

$\Rightarrow \lambda^2 - 22.4867\lambda + 17.55 = 0$

$\lambda_1 = 21.6780, \lambda_2 = 0.8096$

Q10) What is the significance of Eigen vector and Eigen values in PCA?

Ans→ Eigen vectors define the new axes (Principal components) that maximise variance in the data. Eigen values measures the amount of variance along each eigen vector.

The largest eigen values and their corresponding eigen vectors capture the most important feature of the data, enabling dimensionality reduction and pattern recognition.

Q11) What is the purpose of pooling layer in CNN?

Ans→ Pooling is a down-sampling operation that reduces the dimensionally of a feature map. It is applied to rectify feature map to get pooled feature map.

Q12) Describe the term padding in CNN and the purpose of using padding in image processing.

Ans→ Padding in CNN refers to adding extra pixels around the edges of an image before applying convolution.

Purpose of padding in Image Processing →

i) ~~Preserve~~ Preserve Dimensions → Ensure the output size of the feature map remains the same as the input size, which is important for deeper networks.

ii) Avoid Information Loss → Without padding, edge ~~pixel~~ axis are used less frequently in convolutions, leading to the loss of edge information.

iii) Control Output size → Padding allows ~~a~~ flexibility in determining the size of the output feature map.

iv) Improve Performance → Helps maintain spatial information at the edges of the input image.

13) a) Feature map(a) = $\begin{bmatrix} 33 & 47 \\ 18 & 44 \end{bmatrix}$

b) Padded Input → $\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 4 & 7 & 1 & 5 & 0 \\ 0 & 2 & 5 & 8 & 0 & 6 & 0 \\ 0 & 3 & 6 & 9 & 2 & 7 & 0 \\ 0 & 0 & 1 & 2 & 3 & 8 & 0 \\ 0 & 1 & 0 & 3 & 4 & 9 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

Convolution with strides →

Feature map (b) = $\begin{bmatrix} 10 & 25 & 20 \\ 15 & 27 & 38 \\ 2 & 13 & 24 \end{bmatrix}$

c) Pooled Feature map (c) = [19.25]

14) What is Natural Language Processing and how does it differ from Natural Language Understanding (NLU) and Natural Language Generation (NLG)?

Ans → Natural Language Processing (NLP) is a ~~satisfied~~ field of Artificial Intelligence that focuses on enabling computers to process, analyze and generate human language. It encompasses a wide range of tasks, such as text, analysis, language translation, speech recognition and sentiment analysis.

NLP	NLU	NLG
i) Processing language data	i) Understanding language	i) Generating language
ii) Text or speech	ii) Text or speech	ii) Structured data or intent.
iii) Structured data	iii) Contextual meaning	iii) Human-readable language
iv) Sentimental Analysis	iv) Intent recognitition	iv) Chatbot responses.

15) Given the following corpus of 3 documents.

i) "The apple is red and sweet."

ii) "The orange is orange and tangy."

iii) "The fruit basket contains apples, oranges and bananas."

a) Calculate the TF for the term "apple" and "sweet" for the document (i). Also, calculate IDF for the term "orange" and "fruit".

b) Find term-document matrix for documents (i) and (ii).

Ans → a) $TF = \dfrac{\text{No. of occurrences of term in documents}}{\text{Total no. of terms in documents}}$

Total term in documents = ["the", "apple", "is", "red", "and", "sweet"]

= 6

$TF(\text{apple}) = 1/6 = 0.1667 = TF(\text{sweet})$

$IDF(\text{orange}) = \log(3/2) = \log(1.5) = 0.1761$

$IDF(\text{fruit}) = \log(3) = 0.4771$

b) Unique terms = ["the", "apple", "is", "red", "and", "sweet", "orange", "tangy"]

We calculate TF values for each term for both documents.

Term	Documents (i)	Documents (ii)
the	$1/6 = 0.1667$	$1/6 = 0.1667$
apple	$1/6 = 0.1667$	0
is	$1/6 = 0.1667$	$1/6 = 0.1667$
red	$1/6 = 0.1667$	0
and	$1/6 = 0.1667$	$1/6 = 0.1667$
sweet	$1/6 = 0.1667$	0
orange	0	$2/6 = 0.1667$
tangy	0	$1/6 = 0.1667$