		Marks	BL	CO	PO
b.	Differentiate ReLu from a Leaky ReLu with its advantages and disadvantages.	8	3	4	2

PART – C (1 × 15 = 15 Marks)
Answer **ANY ONE** Question

		Marks	BL	CO	PO
26.i.	Why do we prefer CNN over ANN for image data as input? Discuss your view on this.	5			
ii.	Can we use CNN to perform dimensionality reduction? If yes then which layer is responsible for dimensionality reduction particularly in CNN. Explain your answer with an example.	10			
27.i.	Are decision trees affected by outliers. Discuss your answer.	5			
ii.	Is it possible to process long sequences using regular RNNs? If not, then why not?	5			
iii.	In what situations would you prefer to use LSTMs over simple neural nets?	5			

* * * * *

Reg. No.														

B.Tech/ M.Tech (Integrated) DEGREE EXAMINATION, MAY 2023
Fourth Semester

21AIC202J – NEURAL NETWORKS AND MACHINE LEARNING
(For the candidates admitted from the academic year 2022-2023 onwards)

Note:
(i) **Part - A** should be answered in OMR sheet within first 40 minutes and OMR sheet should be handed over to hall invigilator at the end of 40th minute.
(ii) **Part - B** and **Part - C** should be answered in answer booklet.

Time: 3 Hours Max. Marks: 75

PART – A (20 × 1 = 20Marks)
Answer **ALL** Questions

	Question	Marks	BL	CO	PO
1.	_______ refers to a set of assumptions made by a learning algorithm inorder to generalize a limited set of observations. (A) Inductive bias (B) Incremental learning (C) Hypothesis (D) Bias	1	2	1	1
2.	An importance trait of a concept learning algorithm should be able to ______ the choice of hypothesis found. (A) Back track (B) Consider (C) Progress with (D) Restrict	1	2	1	2
3.	_______ refers to a sample belonging to the negative class being classified correctly. (A) True positive (B) True negative (C) False positive (D) False negative	1	2	1	2
4.	Inductive learning takes examples and generalized rather than starting with _______ knowledge. (A) Inductive (B) Existing (C) Deductive (D) Implicit	1	2	1	2
5.	In the case of a decision tree, ________ is referred as inductive bias (A) The depth of the tree (B) A leaf node (C) Root node (D) Middle layer	1	2	2	2
6.	_______ is biased towards multivalued attributes. (A) Information gain (B) Gain ratio (C) Entropy (D) Gini index	1	2	2	2
7.	______ handles the outliers automatically and are robust to outliers (A) Naïve Bayes (B) Linear regression (C) Decision trees (D) Logistic regression	1	2	2	2
8.	Adding new data points to a decision tree may lead to the _______ of the overall decision tree. (A) Stability (B) Regeneration (C) Learning ability (D) Increased complexity	1	2	2	2

9. Give a hypothesis space H, a hypothesis ‘h’ ϵH is said to ______ the training data there exists some alternative hypothesis h' ϵH such that h has a smaller error than h' over the entire distribution of instances. 1 3 3 2
(A) Underfit (B) Overfit
(C) Validate (D) Accept

10. The entropy value is ______ when the sample is completely homogenous and ______ for the sample equally divided between different classes. 1 3 3 2
(A) 0, 1 (B) 1, 0.5
(C) 0.5, 0 (D) 0.5, 0.5

11. ______ performs well for both high dimensional data and low dimensional data 1 2 3 2
(A) Kernelized SVM (B) SVM
(C) Linear regression (D) Decision trees

12. The softmax activation function output the values between ______ and ______. 1 3 2 2
(A) 1 and –1 (B) 0 and 1
(C) 0.5 and –0.5 (D) –1.0 and $\propto$

13. ______ has the ability to escape from local minima and converge to global minima 1 3 4 2
(A) Batch gradient percent (B) Stochastic gradient percent
(C) Mini-batch gradient percent (D) Gradient descent

14. Vanishing gradient can be solved by _______ activation function. 1 3 4 2
(A) Sigmoid (B) Tanh
(C) ReLu (D) Binary step function

15. The _______ computes the error of all observations. 1 3 4 2
(A) Loss function (B) Cost function
(C) Tanh function (D) Sigmoid function

16. Given an array of real numbers: $\begin{bmatrix} 8 \\ 5 \\ 0 \end{bmatrix}$ application of soft-max function transforms the array into $\begin{bmatrix} - \\ - \\ - \end{bmatrix}$ 1 3 4 2

(A) $\begin{bmatrix} 0.5121 \\ 0.7101 \\ 0.6132 \end{bmatrix}$ (B) $\begin{bmatrix} 0.9522 \\ 0.0471 \\ 0.0003 \end{bmatrix}$

(C) $\begin{bmatrix} 0.0010 \\ 0.2120 \\ 0.1121 \end{bmatrix}$ (D) $\begin{bmatrix} 0.9511 \\ 0.4111 \\ 0.3211 \end{bmatrix}$

17. _______ was designed to handle handwriting recognition. 1 2 5 2
(A) Lenet (B) Alexnet
(C) Resnet (D) Gazlenet

18. Vanishing gradient can be resolved by _______ activation function. 1 2 5 2
(A) Sigmoid (B) Tanh
(C) ReLu (D) Binary step function

19. _______ handles handwriting and speech recognition effectively. 1 2 5 2
(A) CNN (B) Neural N/WS
(C) Feed forward N/WS (D) RNN

20. _______ has the ability to process temporal data. 1 3 5 2
(A) CNN (B) RNN
(C) Alexnet (D) Resnet

PART – B (5 × 8 = 40 Marks)
Answer ALL Questions

Marks BL CO PO

21. a. Elucidate the steps in FIND ‘S’ algorithm with suitable examples. Also, write the pseudo code of the algorithm. 8 2 1 2

(OR)

b. With suitable examples, justify the steps in the candidate elimination algorithm. 8 2 1 2

22. a. With respect to decision tree, discuss about hypothesis space search and inductive bias. 8 2 2 1

(OR)

b. Write the steps involved in ID3 algorithm for constructing a decision tree. 8 2 2 2

23. a. Compute the Eigen values and Eigen vectors of: 8 3 2 2

$$\begin{pmatrix} 2 & 0 & 0 \\ 0 & 3 & 4 \\ 0 & 4 & 9 \end{pmatrix}$$

(OR)

b. What is the main goal of the SVM algorithm? Differentiate a linear SVM from a non-linear SVM. 8 3 2 2

24. a. Implement a perceptron for the logic gate “AND” with 2-bit binary inputs. 8 3 3 2

(OR)

b. List out the types of gradient descent used for optimization and explain the SGD-Stochastic Gradient Descent with its advantages. 8 3 3 2

25. a. Discuss the architecture of the Alex net with a neat sketch. 8 3 4 2

(OR)