Department of Cyber Security
(End-Semester Examination: Fall 2025)

Subject:	**Artificial Intelligence**	Total Marks:	**100**
Code:	**CS-344**	Date:	
Class:	**BS-CYS-V**	Time:	
Section:	**A**	Duration:	**3 Hours**
		FM Name:	**Dr. M. Imran**

HoD Signatures: _____ FM Signatures: _____

Note:

1. **Attempt all questions**
2. **This examination carries 45% weight towards the final grade**
3. **Make sure to attempt all parts of a question together!**
4. **DO NOT WRITE ANYTHING ON THE QUESTION PAPER, except your ID**
5. **Calculators are ALLOWED**
6. **You may consult the following formulas for answering the questions:**

Entropy for binary classification:

$$H(S) = -P_1 \log_2(P_1) - P_0 \log_2(P_0)$$

where P_1 = proportion of Positive class and P_0 = proportion of negative class

Information Gain for attribute A:

Information Gain of A = Entropy before split − Weighted average entropy after split on A

Squared Loss: $L = \frac{1}{2}(h(x) - y)^2$

Sigmoid function: $\sigma(z) = \frac{1}{1 + e^{-z}}$

Sigmoid derivative: $\sigma'(z) = \sigma(z)(1 - \sigma(z))$

Naïve Bayes theorem: $P(Y|X) = \frac{P(X|Y)\,P(Y)}{P(X)}$

	Q. No 1 (CLO 1)	**40 Marks**
a	State whether the following statements are **True** or **False**. Any cutting/replacement of answer in any statement will result in zero marks for that statement. (i) The solutions returned by minimax and alpha-beta pruning algorithms may or may not be the same. (ii) k-NN is a parametric algorithm that explicitly learns weights during training. (iii) Unsupervised learning algorithms learn by experimentation. (iv) All filters in the convolution layer of a ConvNet have same weights. (v) RNNs are good at learning from sequences but do not do well on spatial data.	5

b	Fill in the blanks. Any cutting/replacement of answer in a statement will result in zero marks for that statement. (i) The function $f(x) = \max(0, x)$ is known as _____ (ii) _____ is able to capture the long-term dependencies in sequential data which can be problematic for RNNs. (iii) In reinforcement learning, the value function estimates the expected _____ _____from a given state or state-action pair. (iv) When a model is too simple, it may have high bias and produce systematically incorrect predictions, which is called _____. (v) In CSP, the most constrained variable heuristic selects the variable which has the _____legal remaining values.	5
c	(i) Briefly describe the phases of parameter learning in a linear regression problem. (ii) Describe the difference between distance metric and linkage criteria in hierarchical clustering	6+4
d	(i) In both linear and logistic regression, the derivatives of loss w.r.t. parameters and update rules for gradient descent have the same general form: $$\frac{\partial J(\theta)}{\partial \theta_j} = \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_j^{(i)}$$ $$\theta_j := \theta_j - \alpha\frac{\partial J(\theta)}{\partial \theta_j}$$ Linear and logistic regression are different in their nature and application. How would you explain the similarity in their loss derivatives and update rules? (ii) Compare and contrast parametric and non-parametric machine learning algorithms.	6+4
e	(i) Distinguish between parameters and hyper-parameters of a machine learning model. (ii) Give two examples of parameters and hyper-parameters for ConvNets.	4+6
	Q. No 2 (CLO 2)	**30 Marks**
a	Consider a fully connected 1-hidden layer neural network with two hidden neurons, three inputs and a single output. All neurons use **sigmoid** activation function. The Loss function is **squared loss** and learning rate is **0.5**. All the weights are initialized to 2 and biases are initialized to 0. Assume that the output is 1 for the following input: $x_1 = 4, x_2 = 0, x_3 = -5$ (i) Draw a diagram for the above neural network and label all the inputs and outputs. Moreover, assign appropriate symbols to any intermediate variables. (ii) Compute the values of all variables and outputs for the forward pass. (iii) Derive and compute the gradient of Loss w.r.t. pre-activation for the output layer.	6+6+6
b	Suppose you have a training dataset consisting of 3 Boolean features X_1, X_2, and X_3 where $Xi \in \{0, 1\}$. Furthermore, assume that the label is defined as $Y=X_1 \vee X_2$ i.e., Y=1 if X_1 =1 or X_2 =1 and Y = 0 otherwise. Suppose that your training dataset contains all of the 8 possible feature vectors as shown. (i) Draw the root node of the Decision Tree for this dataset and calculate the classification error. (ii) On which feature will you split first? Show your working in support of your answer and draw the tree after the split.	3+6+3

X_1	X_2	X_3	Y
0	0	0	0
1	0	0	1
0	1	0	1
1	1	0	1
0	0	1	0
1	0	1	1
0	1	1	1
1	1	1	1

	(iii) Can you decide not to split on X_3 without doing any calculations? Describe briefly.	
	Q. No 3 (CLO 3)	**30 Marks**
a	(i) Suppose we need to devise a ConvNet for an image recognition application. The size of input images is 10x10 pixels. Can we use a 3x3 filter with a stride of 2 in the convolution layer? Explain your answer. (ii) If your answer to part (i) is *No*, propose a way to resolve the issue. Note: you **cannot** change the image size.	4+6
b	(i) Algorithms like k-Means work on continuous numerical features. Propose a way to use k-Means for data that includes both numerical and categorical features. (ii) Explain your proposed method with a small example.	5+5
c	(i) You are given a dataset where y is non-linear with respect to x. Explain why k-NN might outperform linear regression. (ii) Describe a scenario where linear regression might still be preferable.	5+5

********** End of Question Paper **********