Air University **Student ID:_____**

Department of Cyber Security

(End-Semester Examination: Fall 2025)

Subject:	**Artificial Intelligence**	Total Marks:	**100**
Code:	**CS-344**	Date:	
Class:	**BS-CYS-Eve-V**	Time:	
Section:	**A&B**	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) Supervised learning can be described as learning by example. (ii) Minimax is a parametric machine learning algorithm. (iii) In k-NN algorithm, changing the sequence of inputs affects the classification accuracy. (iv) k-means algorithm can only cluster data with continuous values. (v) In an RNN with only one hidden layer, all time steps have the same hidden layer weights.	5

CamScanner

b	Fill in the blanks. Any cutting/replacement of answer in any statement will result in zero marks for that statement. (i) In Naïve Bayes classification, the posterior probability P(Y|X) is computed using the likelihood P(X|Y) and _____ P(Y) (ii) In ConvNets the _____ layer is used to reduce spatial dimensions (iii) A _____ agent tries to reach the goal in the best possible way (iv) The ratio of correctly classified samples of class A to the total samples classified as class A is called _____ (v) Greedy search is an example of _____ search strategy	5
c	(i) Briefly describe the factors that resulted in boom of deep learning in the last few years. (ii) Can we use K-NN for regression? Explain your answer with the help of an example.	5 + 5
d	Consider the following linear hypothesis function and the respective loss function: $h(x) = wx + b$ $L = \frac{1}{2}(h(x) - y)^2$ (i) Derive the gradients of L with respect to w and b (ii) Assume α to be the learning rate. Derive the update rules for w and b	4+4
e	(i) Describe the factors affecting the size of feature maps in convolution layer of ConvNets. (ii) Can step functions be used in machine learning algorithms? Explain. (iii) Briefly describe the steps in machine learning workflow.	4+4+4
	Q. No 2 (CLO 2)	**30 Marks**
a	Consider the linear hypothesis function and loss function given in Q 1 (d). Assume α to be 0.1, $w = 0$, $b = 0$. Suppose that you have been provided the following data points: $\{(1, 3), (2, 3)\}$. Find the updated values of parameters w and b after iterating over the two data points.	10
b	An organization operates a network intrusion detection system (IDS) that flags packets as **malicious** or **benign** based on statistical analysis of past network traffic. From historical data, the security team has determined that **13%** of all network packets observed on the network are malicious, while **87%** are benign. The IDS also maintains statistics on how frequently certain source IP addresses appear in malicious and benign traffic. For a particular source IP address k, the likelihood of observing this IP in a malicious packet is p, while the likelihood of observing it in a benign packet is q. During live monitoring, the IDS captures a packet with source IP address k. (i) State the condition, in the form of an inequality involving p and q, under which the IDS should classify this packet as **malicious** rather than benign. (ii) During a security investigation, the analyst determines that for this source IP address k, the likelihoods are $p = 0.20$ and $q = 0.04$. Using the condition from part (i), determine how the IDS should classify the packet. Show the key comparison used to arrive at your decision.	5+5

CamScanner

c	Suppose we need to devise a ConvNet for an image recognition application. The size of input images is 10x10 pixels. Assume that we use three 3x3 filters with a stride of 1 in the convolution layer. (i) How much memory will be required for feature maps in the convolution layer? Assume a pixel is 1-bit. (ii) How many weights will be required in the convolutional layer? Assume each filter has a bias term. (iii) In the pooling layer we'll use a non-overlapping 2x2 window with max pooling. What will be the memory requirements after the pooling has been done?	4+3+3
	Q. No 3 (CLO 3)	**30 Marks**
a	You are making a classifier that can distinguish between birds of two species i.e., crows and ravens. You have a training set with 100 crows and 100 ravens, and a test set with 20 crows and 20 ravens. (i) The 1-nearest-neighbors algorithm gets 100% accuracy on the training set, but only 60% accuracy on the test set. The 3-nearest neighbors algorithm gets only 90% accuracy on the training set, but it gets 70% accuracy on the test set. If you want your algorithm to work well for birds you've never seen before, should you choose $k = 1$ or $k = 3$? Explain the reason for your answer. (ii) Suppose that, instead of 100 crows and 100 ravens, your training set has only 3 crows and 3 ravens. The crows are named Alpha, Beta and Charlie, and they are 12, 18, and 13 inches long, respectively. The ravens are named Alfred, Bingo, and Chillem, and they are 24, 16, and 22 inches long, respectively. You want to guess if the Bird X, which is 19 inches long, is a crow or a raven. Will the value of k have any effect on the output of k-NN algorithm for the bird X? Explain your answer by using two different values of k. Also provide the k nearest neighbors for each of these values of k.	5+5
b	Suppose you have to derive the Decision Tree for the data given below for 5 samples where A1, A2 and A3 are binary inputs (attributes) and Y is output. (i) Calculate the entropy of the root node. (ii) On which attribute should the tree be split first? Show complete working based on Information Gain to support your answer.	2+8
	(i) Malware detection (detecting if a given program is malicious or benign) and malware classification (determining the type of malware to which the given programs belongs) are important cyber security problems. Relate these problems with the machine learning concepts that you studied in the class. (ii) Suppose we have a collection of programs which include malware as well as benign applications. Propose ways to use (a) ConvNets and (b) RNNs for malware detection/classification.	4 + 6

Example	A1	A2	A3	Y
x1	1	0	0	0
x2	1	0	1	0
x3	0	1	0	0
x4	1	1	1	1
x5	1	1	0	1

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

CamScanner