Q. Code: 606362

Reg. No. | | | | | | | | | | | | | |

B.E. / B.TECH. DEGREE EXAMINATIONS, MAY 2024

Fourth Semester

EC22408 – MACHINE LEARNING: THEORY AND PRACTICE

(Electronics and Communication Engineering)

(Regulation 2022)

TIME: 2 HOURS **MAX. MARKS: 60**

COURSE OUTCOMES	STATEMENT	RBT LEVEL
CO 1	Explain the basic concepts of machine learning.	**2**
CO 2	Construct supervised learning models.	**3**
CO 3	Construct unsupervised learning algorithms.	**3**
CO 4	Evaluate and compare different models.	**4**
CO 5	Evaluate the machine learning experiments.	**4**

PART- A (10 x 2 = 20 Marks)

(Answer all Questions)

		CO	RBT LEVEL
1.	Discuss how the VC dimension of the hypothesis space might influence the amount of training data required to achieve good performance with the model.	**1**	**2**
2.	Distinguish between over-fitting and under-fitting. How it can affect model Generalization?	**1**	**2**
3.	Explain the basic principle behind Bayes linear regression and how it differs from classical linear regression?	**2**	**2**
4.	Describe the significance of Kernal functions in SVM. List any two kernel functions.	**2**	**2**
5.	Differentiate between bagging and boosting.	**3**	**2**
6.	State advantage and disadvantage of the K-Nearest Neighbors (K-NN) algorithm.	**3**	**2**
7.	Calculate the output y of a three input neuron with bias. The input feature vector is $(x_1, x_2, x_3) = (0.8, 0.6, 0.4)$ and weight values are $[w_1, w_2, w_3, b] = [0.2, 0.1, -0.3, 0.35]$. Use binary Sigmoid function as activation function.	**4**	**3**
8.	With suitable equations, explain any two types of activation functions used in neural networks.	**4**	**3**

9.		Explain the concept of K-fold Cross Validation and how it works?	**5**	**2**	
10.		What is the significance of the McNemar's test statistic, and how is it interpreted?	**5**	**2**	

PART- B (3 x 10 = 30 Marks)

			Marks	CO	RBT LEVEL
11. (a)	**(i)**	Explain the types of machine learning techniques with examples.	**(6)**	**1**	**2**
	(ii)	Choose a specific real-world application of machine learning (e.g., healthcare, finance, autonomous vehicles) and describe in detail how machine learning techniques are utilized in solving the problem.	**(4)**	**1**	**2**
		(OR)			
(b)	**(i)**	Discuss the role of Occam's razor in shaping the inductive bias of machine learning algorithms. How does it relate to the preference for simpler models?	**(6)**	**1**	**2**
	(ii)	Short note on a) Bias and variance b) Bias and variance trade off	**(4)**	**1**	**2**
12. (a)		a) Identify the attributes for decision tree by using ID3 algorithm for the following dataset	**(10)**	**2**	**3**

Major	Experience	Tie	Hired?
CS	programming	pretty	NO
CS	programming	pretty	NO
CS	management	pretty	YES
CS	management	ugly	YES
business	programming	pretty	YES
business	programming	ugly	YES
business	management	pretty	NO
business	management	pretty	NO

(OR)

(b)		Consider a dataset of emails labeled as either "spam" or "not spam," with features representing the presence or absence of certain keywords. Using the Naive Bayes algorithm, classify a new email with the following	**(10)**	**2**	**3**

features:

- Contains the word "free"
- Contains the word "offer"
- Does not contain the word "buy"

Given the following training data:

Email	Contains "free"	Contains "offer"	Contains "buy"	Label
1	Yes	Yes	No	Spam
2	Yes	Yes	Yes	Spam
3	Yes	No	No	Not Spam
4	No	Yes	No	Not Spam
5	No	Yes	Yes	Spam
Use α=1.				

13. (a)	Consider a dataset of 10 data points in a two-dimensional space: {(2,3), (4,7), (6,5), (8,8), (10,6),(12,10),(14,8),(16,4),(18,9),(20,12)} Apply the K-Means algorithm with k=3 clusters, starting with the initial centroids at (2,3) (2,3), (8,8) (8,8), and (16,4) (16,4). Perform two iterations of the algorithm and show the resulting cluster assignments and centroids.	**(10)**	**3**	**3**
	(OR)			
(b)	**(i)** Explain the concept of ensemble models in machine learning. Discuss the different types of ensemble techniques.	**(6)**	**3**	**3**
	(ii) Compare and contrast the performance of ensemble models with individual base learners and discuss scenarios where ensemble techniques are particularly effective.	**(4)**	**3**	**3**

PART- C (1 x 10 = 10 Marks)

(Q.No.14 is compulsory)

		Marks	CO	RBT LEVEL
14.	Consider a single-layer perceptron neuron with two input features (x_1 and x_2) and a step activation function. The neuron's weights are $w_1 = 0.5, w_2 = -0.2$, and the bias is $b = 0.1$. Draw a diagram of the perceptron neuron with the activation function and calculate the output for the input feature vector (x_1 and x_2) = (0.7,0.9) and compute the error.	**(10)**	**4**	**4**

****