2022

COMPUTER SCIENCE AND ENGINEERING

Paper : CSEL-0919

(Elective - II)

[Machine Learning]

Full Marks : 70

The figures in the margin indicate full marks.

Candidates are required to give their answers in their own words as far as practicable.

Answer ***question nos.* 1, 2** and ***any four*** from the rest.

1. Answer ***any five*** questions : 2×5

(a) Distinguish between regression and classification.

(b) What is a logistic function?

(c) Consider a learning problem with 2D features. How are the decision tree and 1-nearest neighbour decision boundaries related?

(d) What will be the correlation between two independent variables, X and Y?

(e) Which one of the following is a supervised algorithm? —Linear regression, Logistic regression, k-means clustering, kNN. Justify your answer.

(f) What is cross-validation? How does it improve the accuracy of the outcomes?

2. Answer ***any five*** questions : 4×5

(a) Computer Accuracy, Specificity and Sensitivity considering the following confusion matrix obtained as a result of logistic regression on a dataset :

		Actual	
		True	False
Predicted	True	20	3
	False	5	18

(b) What are the different terminating conditions in the generation of decision tree?

(c) List the steps to solve any machine learning problem.

Please Turn Over

(d) Describe a method used for assessing the normality of errors in linear regression.

(e) What type of machine learning problem is addressed by logistic regression? Can linear regression address the same problem? Justify your answer.

(f) Define the expected loss function, clearly mentioning the random variables. Write the formula for convergence in expectation, clearly defining each variable.

(g) How can you show or prove that a dataset is not linearly separable?

(h) What are Loss Functions and Cost Functions? Explain the critical difference between them.

3. (a) Define Mean Squared Error (MSE).

(b) Explain the method of deciding the optimal number of parameters in linear regression. 3+7

4. (a) What types of problem are solved using k-means and kNN algorithm?

(b) What is k in k-means, and what is the effect of choosing a low or high k value in k-means?

(c) What is k in kNN, and how do you select an appropriate k? 2+4+4

5. (a) Find eigenvalues and eigenvectors for matrix $A=\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$.

(b) Why is it required in machine learning? 7+3

6. (a) What is meant by feature extraction? Write an algorithm to extract the features from a given dataset.

(b) Critically comment on a recommender system. Briefly explain the different phases of a recommendation system. 5+5

7. (a) What are the significances of PCA?

(b) Is it important to standardize the data before applying PCA?

(c) What are the fundamental steps of PCA? Write down their importance.

(d) Critically comment on the pros and cons of PCA. 2+2+4+2

8. Consider the dataset below to learn a decision tree that predicts if people pass machine learning (Yes or No), based on their previous GPA (High, Medium, or Low), and whether or not they studied.

(For this problem, you can write your answers using $\log_2$*, but it may be helpful to note that* $\log_2 3 \approx 1.6$*)*

GPA	Studied	Passed
L	F	F
L	T	T
M	F	F
M	T	T
H	F	T
H	T	T

(a) What is the entropy H(Passed)?

(b) What is the entropy H(Passed | GPA)?

(c) What is the entropy H(Passed | Studied)?

(d) Draw the entire decision tree that would be learnt for this dataset. 2+2+2+4