CBCS SCHEME

USN

21AD62

Sixth Semester B.E./B.Tech. Degree Examination, June/July 2025
Data Science and its Applications

Time: 3 hrs. Max. Marks: 100

Note: ***Answer any FIVE full questions, choosing ONE full question from each module.***

Module-1

1 a. With the help of python code compute mean and median of a data set by considering odd and even cases. **(08 Marks)**

b. Describe dispersion and variance. Show the python code to compute the variance. **(08 Marks)**

c. Write a function to create a matrix and retrieve an element from it. **(04 Marks)**

OR

2 a. What are random variables? Explain normal distribution of data. **(08 Marks)**

b. Discuss conditional probability with an example. **(06 Marks)**

c. Illustrate Simpson's paradox with an example. **(06 Marks)**

Module-2

3 a. Illustrate the concept of gradient descent and its purpose in fitting models. Also demonstrate how you would choose an appropriate step size for the gradient descent algorithm with the help of a python program. **(10 Marks)**

b. What is web scrapping? Explain the steps involved in scraping data from the website using python. **(10 Marks)**

OR

4 a. Explain statistical hypothesis testing. How level of significance is important for decision making during hypothesis testing. **(10 Marks)**

b. Describe how data can be manipulated by considering an example. **(06 Marks)**

c. Write a note on dimensionality reduction. **(04 Marks)**

Module-3

5 a. From the given dataset deduce the confusion matrix and calculate the FI score. Write corresponding python program Refer Table Q5(a).

Index	1	2	3	4	5	6	7	8	9	10
Actual	Dog	Dog	Dog	Not Dog	Dog	Not Dog	Dog	Dog	Not Dog	Not Dog
Predicted	Dog	Not Dog	Dog	Not Dog	Dog	Dog	Dog	Dog	Not Dog	Not Dog

Table Q5(a) **(10 Marks)**

b. With suitable examples, explain linear and logistic regression models. **(10 Marks)**

OR

6 a. Develop python program to build a nearest neighbor model that can predict the class from the IRIS dataset. **(10 Marks)**

b. Explain overfitting and underfitting in detail with an example. **(10 Marks)**

Module-4

7 a. Neurons have the sigmoid activation function to perform forward and backward pass on the given network. Assume that the actual putout of y is 0.5 and learning rate is 1. Perform back propagation for 1 epoch to update weights with a neat block diagram, $W_{2,3}$ and $W_{1,3}$. Refer Fig.Q7(a).

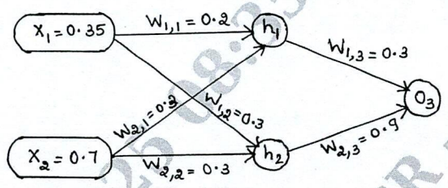

Fig.Q7(a) **(12 Marks)**

b. Realize AND gate using a perceptron if $W_1 = 1.2$ and $W_2 = 0.6$ and threshold is 1. Assume learning rate to be 0.5. **(08 Marks)**

OR

8 a. For the given dataset find the clusters using bottom-up approach. Use Euclidean distance and draw the dendrogram to identify the number of clusters. Refer Table Q8(a).

	X	Y
P1	0.40	0.53
P2	0.22	0.38
P3	0.35	0.32
P4	0.26	0.19
P5	0.08	0.41
P6	0.45	0.30

Table Q8(a) **(12 Marks)**

b. Explain feed forward neural network and realize XOR gate. **(08 Marks)**

Module-5

9 a. Find the most significant node in the following network using eigen vector centrality. Refer Fig.Q9(a).

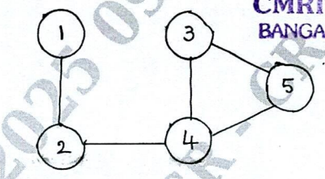

Fig.Q9(a) **(08 Marks)**

b. Explain how Gibbs sampling helps in language processing. **(06 Marks)**

c. For the given network find betweeness centrality for node A. Refer Fig.Q9(c).

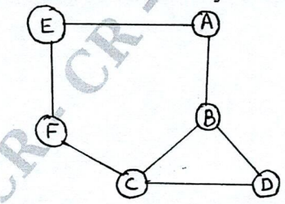

Fig.Q9(c) **(06 Marks)**

OR

10 a. Discuss user-based and item-based collaborative filtering with examples. **(08 Marks)**

b. Describe directed graphs and page rank. **(06 Marks)**

c. Explain n gram model. How it helps in predicting the next word. **(06 Marks)**

* * * * *