U.S.N.										

B.M.S. College of Engineering, Bengaluru-560019

Autonomous Institute Affiliated to VTU

September / October 2023 Supplementary Examinations

Programme: B.E **Semester: VI**

Branch: ES – Cluster Elective **Duration: 3 hrs.**

Course Code: 19ET6CE1DS **Max Marks: 100**

Course: Data Science

Instructions: 1. Answer any FIVE full questions, choosing one full question from each unit.
2. Missing data, if any, may be suitably assumed.

UNIT - I

1 a) Examine membership operators and identity operators with an example each in python **06**

b) Explain functions with suitable code for the following cases in python **08**
(i) function with args
(ii) function with *args
(iii) function with kwargs
(iv) function with **kwargs

c) What is anonymous function in python? Illustrate how do you implement anonymous functions in python with an example code **06**

UNIT - II

2 a) For the list given in x=[1,2,3,1,4,5,1,3,2,4,5,1,1], find **04**
i. mean
ii. median
iii. quantile (x,25%)
iv. mode

b) Define the following **10**
i. Range
ii. Variance
iii. Standard deviation
iv. Covariance
v. Correlation coefficient

c) Consider a family with two children. If we assume that: Each child is equally likely to be a boy or a girl and the gender of the second child is independent of the gender of the first child. Then simulate the following probability using a suitable python code. **06**
(i) Both children are girls conditional on the event “the older child is a girl”
(ii) Both children are girls conditional on the event “at least one of the children is a girl”.

UNIT - III

3 a) What is Gradient Descent? Examine the estimation of gradient using python codes with respect to **08**
   (i) Difference quotient
   (ii) Partial difference quotient

b) Differentiate between gradient descent and stochastic gradient descent algorithm. Also list the various options available for choosing the step size **06**

c) Find the correlation matrix for the data given below. **06**

$$data = \begin{bmatrix} 1 & 3 & -1 \\ 2 & 2 & -2 \\ 3 & 1 & 3 \end{bmatrix}$$

UNIT - IV

4 a) Describe choice of model in data science with respect to following **06**
   (i) Trade-off between Precision and Recall
   (ii) Bias- Variance trade off

b) For the dataset given below, Apply KNN algorithm to predict the possible sport for the query (Angelina, 5 years, female) using k=3. Assume Male=0 and Female =1 **08**

Name	Age (years)	Gender	Sport
Ajay	32	Male	Foot ball
Mark	40	Male	Neither
Sara	16	Female	Cricket
zaira	34	Female	Foot ball
Sachin	55	Male	Neither
Rahul	40	Male	Cricket

c) Describe the mathematics behind a spam filter with an example **06**

OR

5 a) Discuss the working of Naive Bayes classifier with necessary equations. **06**

b) Consider the data set given below. Apply Naive bayes classifier to determine whether the statement "*Players will play if weather is sunny* "is correct or not. **06**

Weather	Play
Sunny	No
Overcast	Yes
Rainy	Yes
Sunny	Yes
Sunny	Yes
Overcast	Yes
Rainy	No
Rainy	No
Sunny	Yes
Rainy	Yes
Sunny	No
Overcast	Yes
Overcast	Yes
Rainy	No

c) Consider the model for simple regression and the data set as shown below. **08**
$y = \beta x + \alpha$ where y is the dependent variable and x is the independent variable.

x	y
1	1
2	3
4	3
3	2
5	5

Find the following

(i) β

(ii) α

(iii) Predicted values of y

(iv) RMS Error

(v) R Squared

UNIT - V

6 a) Consider the dataset given below. Apply Logistic Regression to answer the following **10**

x1	x2	y (class)
2.78	2.55	0
1.47	2.36	0
3.39	4.40	0
1.38	1.85	0
3.06	3.00	0
7.63	2.76	1
5.33	2.09	1
6.92	1.77	1
8.67	-0.24	1
7.67	3.51	1

i) Calculate the probability/prediction of the first training instance that belongs to class 0

ii) Update the coefficients for Epoch 1. Assume learning rate=0.3

Suppose the coefficients after 10 Epochs are a=-0.41, b1=0.85 and b2=-1.10 find the predictions and predicted class for above data set

b) The (x, y) coordinates of few points in 2 dimensional space belonging to 2 different class groups are as given below. Here x is the point along x axis and y is the point along y axis. **10**

Class 1	(1,1), (2,1), (1,-1), (2,-1)
Class 2	(4,0), (5,1), (5,-1), (6,0)

Apply SVM algorithm to answer the following.

i. Plot the points on a 2-D plane with different representations for class1 & class2.

ii. Identify the support vectors

iii. Estimate the hyper plane coefficients α with respect to each

support vector

iv. Estimate the hyper plane equation along with weight vector and bias

v. Draw the classifier line and hyper plane on given sample points in 2-D plane.

OR

7 a) Answer the following with respect to Neural Network **12**

i. Draw the block diagram of ANN and describe how the ANN learns?

ii. Perceptron Model

iii. Write a python function to implement a perceptron model.

iv. Implement NOT gate using perceptron model

b) Plot a dendogram using Agglomerative clustering for the following data elements **08**

Item	A	B	C	D	E	F
A	0					
B	0.71	0				
C	5.66	4.95	0			
D	3.61	2.92	2.24	0		
E	4.24	3.54	1.41	1.00	0	
F	3.20	2.50	2.50	0.50	1.12	0

******