

U.S.N.								
--------	--	--	--	--	--	--	--	--

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

Autonomous Institute Affiliated to VTU

January / February 2025 Semester End Main Examinations

Programme: B.E.

Semester: VII

Branch: Electronics & Telecommunication Engineering

Duration: 3 hrs.

Course Code: 22ET7PE3DS

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			CO	PO	Marks
Important Note: Completing your answers, compulsorily draw diagonal cross lines on the remaining blank pages. Revealing of identification, appeal to evaluator will be treated as malpractice.	1	a)	What are the datatypes available in python? Give one example code for each.			CO1	-	10
		b)	Explain membership operators and identity operators with an example each in python.			CO1	-	05
		c)	Demonstrate object oriented programming in python with respect to creating a class, initializing the members and creating an object.			CO2	PO1	05
			OR					
2	a)	Develop python functions for the following operations in linear algebra : (i) vector addition (ii) vector subtraction (iii) vector element by element addition (iv) scalar multiplication (v) dot product				CO2	PO1	10
	b)	Write the output of the following code:				CO4	PO4	05
		<pre> x=[0,1,2,3,4,5,6,7,8,9] print(x) print("\n") print("x[0]=",x[0]) print("\n") print("x[1]=",x[1]) print("\n") print("x[-1]=",x[-1]) print("\n") print("x[-2]=",x[-2]) print("\n") print("x[:3]=",x[:3]) print("\n") </pre>						

	c)	Discuss list comprehensions in python with suitable programming examples	CO2	PO1	05
		UNIT - II			
3	a)	Define the following with an example for each of the following : i. Mean ii. Median iii. quantile iv. mode	CO1	-	06
	b)	What is dispersion? What are the statistical tools to compute dispersion?	CO1	-	06
	c)	for the list given in $x=[1,2,3,1,4,5,1,3,2,4,5,1,1]$, find i. Range ii. Variance iii. Standard deviation iv. Interquartile range	CO3	PO2	08
		OR			
4	a)	What is Simpsons paradox, explain with an example.	CO1	-	05
	b)	Discuss correlational caveats with suitable examples.	CO1	-	05
	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 find the probability of (i) No girls (ii) One girl, one boy (iii) Two girls (iv) Both children are girls conditional on the event “the older child is a girl”. (v) Both children are girls conditional on the event “at least one of the children is a girl”.	CO3	PO2	10
		UNIT - III			
5	a)	What is Gradient Descent? Explain the estimation of gradient using python codes with respect to: (i) Difference quotient (ii) Partial difference quotient	CO2	PO1	06
	b)	With respect to Gradient Descent algorithm : (i) List the steps involved in minimizing a function. (ii) Develop a python code for Gradient Descent algorithm.	CO2	PO1	08
	c)	Find the correlation matrix for the data given below: $data = \begin{bmatrix} 1 & 3 & -1 \\ 2 & 2 & -2 \\ 3 & 1 & 3 \end{bmatrix}$	CO3	PO2	06
		OR			
6	a)	For the dataset given below:	CO3	PO2	08

		<table border="1"> <tr> <td>Perso n</td><td>Height (inches)</td><td>Height (centimetres)</td><td>Weight (pounds)</td></tr> <tr> <td>A</td><td>63</td><td>160</td><td>150</td></tr> <tr> <td>B</td><td>67</td><td>170.2</td><td>160</td></tr> <tr> <td>C</td><td>70</td><td>177.8</td><td>171</td></tr> </table>	Perso n	Height (inches)	Height (centimetres)	Weight (pounds)	A	63	160	150	B	67	170.2	160	C	70	177.8	171									
Perso n	Height (inches)	Height (centimetres)	Weight (pounds)																								
A	63	160	150																								
B	67	170.2	160																								
C	70	177.8	171																								
		<ul style="list-style-type: none"> (i) Find B's nearest neighbour with respect to [height (inches),weight] (ii) Find B's nearest neighbour with respect to [height (centi meters),weight] (iii) Write python code to demonstrate (i) and (ii) (iv) Develop python code to demonstrate Rescaling required. 																									
	b)	Apply PCA to demonstrate dimension reduction for the following dataset. Show all steps and provide the results for following.	CO3	PO2	12																						
		<table border="1"> <tr> <td>x</td> <td>y</td> </tr> <tr> <td>2.5</td> <td>2.4</td> </tr> <tr> <td>0.5</td> <td>0.7</td> </tr> <tr> <td>2.2</td> <td>2.9</td> </tr> <tr> <td>1.9</td> <td>2.2</td> </tr> <tr> <td>3.1</td> <td>3.0</td> </tr> <tr> <td>2.3</td> <td>2.7</td> </tr> <tr> <td>2</td> <td>1.6</td> </tr> <tr> <td>1</td> <td>1.1</td> </tr> <tr> <td>1.5</td> <td>1.6</td> </tr> <tr> <td>1.1</td> <td>0.9</td> </tr> </table> <ul style="list-style-type: none"> (i) Mean values (ii) Covariance matrix (iii) Eigen values (iv) Eigen vectors (v) Transformed data with both principal components (vi) Transformed data with first principal component 	x	y	2.5	2.4	0.5	0.7	2.2	2.9	1.9	2.2	3.1	3.0	2.3	2.7	2	1.6	1	1.1	1.5	1.6	1.1	0.9			
x	y																										
2.5	2.4																										
0.5	0.7																										
2.2	2.9																										
1.9	2.2																										
3.1	3.0																										
2.3	2.7																										
2	1.6																										
1	1.1																										
1.5	1.6																										
1.1	0.9																										
		UNIT - IV																									
7	a)	Discuss the following in data science with suitable examples <ul style="list-style-type: none"> (i) Modelling (ii) Machine learning (iii) Overfitting (iv) Underfitting 	CO1	-	06																						
	b)	Following are the statistics of a test conducted for a particular disease in a laboratory. TP=70, FP=4930, FN=13930, TN=981070. Calculate the following. Also comment on the results. <ul style="list-style-type: none"> (i) Accuracy (ii) Precision (iii) Recall (iv) F1 Score 	CO3	PO2	06																						
	c)	For the dataset given below, Apply KNN algorithm to predict the possible classification for the query (3 seconds, 7 kg/square meter)	CO3	PO2	08																						

		using k=3.																																											
		<table border="1"> <thead> <tr> <th>Acid durability (seconds)</th> <th>Strength (kg/square meter)</th> <th>Classification</th> </tr> </thead> <tbody> <tr> <td>7</td> <td>7</td> <td>Bad</td> </tr> <tr> <td>7</td> <td>4</td> <td>Bad</td> </tr> <tr> <td>3</td> <td>4</td> <td>Good</td> </tr> <tr> <td>1</td> <td>4</td> <td>Good</td> </tr> </tbody> </table>	Acid durability (seconds)	Strength (kg/square meter)	Classification	7	7	Bad	7	4	Bad	3	4	Good	1	4	Good																												
Acid durability (seconds)	Strength (kg/square meter)	Classification																																											
7	7	Bad																																											
7	4	Bad																																											
3	4	Good																																											
1	4	Good																																											
		OR																																											
8	a)	Describe the mathematics behind a spam filter with an example.	<i>CO2</i>	<i>PO1</i>	06																																								
	b)	Consider the data set given below. Predict if a given fruit is a 'banana' or 'orange' or 'other' when only three features [long, sweet and yellow] are known using naive bayes classifier.	<i>CO3</i>	<i>PO2</i>	08																																								
		<table border="1"> <thead> <tr> <th>Type</th> <th>Long</th> <th>Not long</th> <th>Sweet</th> <th>Not sweet</th> <th>Yellow</th> <th>Not yellow</th> <th>Total</th> </tr> </thead> <tbody> <tr> <td>Banana</td> <td>400</td> <td>100</td> <td>350</td> <td>150</td> <td>450</td> <td>50</td> <td>500</td> </tr> <tr> <td>Orange</td> <td>0</td> <td>300</td> <td>150</td> <td>150</td> <td>300</td> <td>0</td> <td>300</td> </tr> <tr> <td>Other</td> <td>100</td> <td>100</td> <td>150</td> <td>50</td> <td>50</td> <td>150</td> <td>200</td> </tr> <tr> <td>Total</td> <td>500</td> <td>500</td> <td>650</td> <td>350</td> <td>800</td> <td>200</td> <td>1000</td> </tr> </tbody> </table>	Type	Long	Not long	Sweet	Not sweet	Yellow	Not yellow	Total	Banana	400	100	350	150	450	50	500	Orange	0	300	150	150	300	0	300	Other	100	100	150	50	50	150	200	Total	500	500	650	350	800	200	1000			
Type	Long	Not long	Sweet	Not sweet	Yellow	Not yellow	Total																																						
Banana	400	100	350	150	450	50	500																																						
Orange	0	300	150	150	300	0	300																																						
Other	100	100	150	50	50	150	200																																						
Total	500	500	650	350	800	200	1000																																						
	c)	Discuss the model of simple linear regression and explain using an example	<i>CO2</i>	<i>PO1</i>	06																																								
		UNIT - V																																											
9	a)	Suppose we only know a person's height and we want to predict whether that person is male or female. We can talk about the probability of being male or female, or we can talk about the odds of being male or female. Let's say that the probability of being male at a given height is 0.90. Then what would be the odds of being male and female?	<i>CO3</i>	<i>PO2</i>	08																																								
	b)	Consider the dataset given below, apply Logistic Regression to answer the following :	<i>CO3</i>	<i>PO2</i>	12																																								
		<table border="1"> <thead> <tr> <th>x1</th> <th>x2</th> <th>y (class)</th> </tr> </thead> <tbody> <tr> <td>2.78</td> <td>2.55</td> <td>0</td> </tr> <tr> <td>1.47</td> <td>2.36</td> <td>0</td> </tr> <tr> <td>3.39</td> <td>4.40</td> <td>0</td> </tr> <tr> <td>1.38</td> <td>1.85</td> <td>0</td> </tr> <tr> <td>3.06</td> <td>3.00</td> <td>0</td> </tr> <tr> <td>7.63</td> <td>2.76</td> <td>1</td> </tr> <tr> <td>5.33</td> <td>2.09</td> <td>1</td> </tr> <tr> <td>6.92</td> <td>1.77</td> <td>1</td> </tr> <tr> <td>8.67</td> <td>-0.24</td> <td>1</td> </tr> <tr> <td>7.67</td> <td>3.51</td> <td>1</td> </tr> </tbody> </table>	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										
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																																											

		<p>i. Calculate the probability/prediction of the first training instance that belongs to class 0</p> <p>ii. Update the coefficients for Epoch 1. Assume learning rate=0.3</p> <p>iii. Suppose the coefficients after 10 Epochs are $a=-0.41$, $b_1=0.85$ and $b_2=-1.10$ find the predictions and predicted class for above data set.</p>																																																				
		OR																																																				
10	a)	<p>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.</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <tr> <td>Class 1</td><td>(1,1), (2,1), (1,-1), (2,-1)</td></tr> <tr> <td>Class 2</td><td>(4,0), (5,1), (5,-1), (6,0)</td></tr> </table> <p>Apply SVM algorithm to answer the following.</p> <ol style="list-style-type: none"> 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. 	Class 1	(1,1), (2,1), (1,-1), (2,-1)	Class 2	(4,0), (5,1), (5,-1), (6,0)	CO3	PO2	10																																													
Class 1	(1,1), (2,1), (1,-1), (2,-1)																																																					
Class 2	(4,0), (5,1), (5,-1), (6,0)																																																					
	b)	<p>Plot a dendrogram using Agglomerative clustering for the following data elements :</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th>Item</th><th>A</th><th>B</th><th>C</th><th>D</th><th>E</th><th>F</th></tr> </thead> <tbody> <tr> <td>A</td><td>0</td><td></td><td></td><td></td><td></td><td></td></tr> <tr> <td>B</td><td>0.71</td><td>0</td><td></td><td></td><td></td><td></td></tr> <tr> <td>C</td><td>5.66</td><td>4.95</td><td>0</td><td></td><td></td><td></td></tr> <tr> <td>D</td><td>3.61</td><td>2.92</td><td>2.24</td><td>0</td><td></td><td></td></tr> <tr> <td>E</td><td>4.24</td><td>3.54</td><td>1.41</td><td>1.00</td><td>0</td><td></td></tr> <tr> <td>F</td><td>3.20</td><td>2.50</td><td>2.50</td><td>0.50</td><td>1.12</td><td>0</td></tr> </tbody> </table>	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	CO3	PO2	10
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																																																
