PSG COLLEGE OF TECHNOLOGY, COIMBATORE - 641 004

Department of Computer Science and Engineering

BE CSE - G2 & SEM 6

CONTINUOUS ASSESSMENT TEST 1 **Date: 20.1.2025**

19Z601 – Machine Learning

Time: 1 Hour 30 minutes. **Maximum Marks: 50**

INSTRUCTIONS:

1. Answer **ALL** questions. Each Question carries 25 Marks.
2. In each question, subdivision **a** contains 5 questions and the weightage of each question is one mark, subdivision **b(i)** and **b(ii)** carries 5 marks each and subdivision **c** carries 10 marks each.
3. Subdivisions (a) and (b) will be with no choice and Subdivision (c) may be with choice but not in more than 1 question.
4. Course Outcome Table

Qn. 1	CO1

Qn.2	CO2

1.a		
	(5x1mark=5marks)	**BTL**
i)	You are given a dataset containing customer information, including their age, income, purchase history, and geographic location. The goal is to segment customers into distinct groups based on their purchasing behaviour to tailor marketing strategies. Which type of machine learning technique is most suitable for this task? **A)** Supervised Learning **B)** Unsupervised Learning **C)** Reinforcement Learning **D)** Semi-Supervised Learning	**L2**
ii)	How does one-hot encoding handle categorical variables with multiple unique values, and what is a potential drawback of using this technique? **A)** It replaces each category with its frequency and can lead to biased models. **B)** It assigns a unique integer to each category, which may imply an unintended ordinal relationship. **C)** It creates a binary column for each category, which can increase dimensionality significantly. **D)** It combines similar categories into a single column, which may lose information.	**L2**

iii)	Consider the following statements about types of missing values in a dataset: 1. MCAR (Missing Completely at Random) means that the missing data is unrelated to both observed and unobserved data. 2. MNAR (Missing Not at Random) can lead to biased analysis if not handled appropriately, as the missing data is related to unobserved values. Which of the following is correct? **A)** Both statements are true. **B)** Both statements are false. **C)** Only statement 1 is true. **D)** Only statement 2 is true.	**L1**
iv)	The number of wickets taken in last seven games by a cricket player is specified as data points (vector) [2,3,1,4, 5, ____2____, 2]. Fill the missing data point with mode data imputation.	**L3**
v)	The first principal component accounts for the __high__ variance in the data, while subsequent components account for the remaining variance in desceding order.	**L2**
b.	**(2 x 5 marks = 10 marks)**	
i)	Consider the training dataset of 4 instances. It contains the details of the performance of students and their likelihood of getting a job offer or not in their final semester. Apply the Find-S Algorithm.	**L3**

CGPA	Interactiveness	Practical Knowledge	Job Offer
>=9	Yes	Excellent	Yes
>=9	Yes	Good	Yes
>=8	No	Good	No
>=9	Yes	Good	Yes

ii)	You are given a dataset containing the sepal width of Iris flower in centimeters: Sepal Width (cm): [3.5,3,3.2,3.6] Use the following data scaling techniques and provide the scaled values for the above dataset. (i) Min-Max Normalization (ii) Z-score Normalization	**L3**
c.	**(1 x 10 marks = 10 marks)**	
	Using the Candidate Elimination Method, determine the version space	**L4**

(S and G boundaries) after processing the given dataset. Assume that "EnjoySport" is the target concept.

Example	*Sky*	*AirTemp*	*Humidity*	*Wind*	*Water*	*Forecast*	*EnjoySport*
1	Sunny	Warm	Normal	Strong	Warm	Same	Yes
2	Sunny	Warm	High	Strong	Warm	Same	Yes
3	Rainy	Cold	High	Strong	Warm	Change	No
4	Sunny	Warm	High	Strong	Cool	Change	Yes

Tabulate inconsistent hypothesis.

Analyze the changes in the version space (S and G boundaries) after processing the dataset. Identify any inconsistencies in the data or concepts that might lead to conflicts within the version space

2.a	(5x1mark=5marks)	BTL
i)	Which of the following statements about linear regression models is **TRUE**? A) Linear regression can only be applied if the relationship between variables is non-linear. ✗ B) The slope coefficient in a simple linear regression represents the average change in the dependent variable for a one-unit increase in the independent variable. C) In multiple linear regression, adding more independent variables always improves the model's accuracy. D) Linear regression models do not assume any relationship between the independent and dependent variables. ✗	L2
ii)	Which of the following is an example of cultural bias in datasets? **A)** Missing values in the dataset. **B)** Overrepresentation of one demographic group leading to skewed predictions. **C)** Using a dataset that includes diverse geographic and cultural populations. **D)** Applying feature scaling to normalize data.	L3
iii)	The following are the assumptions of linear regression : Assumption 1 : Linearity Assumption 2: Heteroscedasticity Which of the following is correct? A) Both Assumptions are true. B) Both Assumptions are false. C) Only Assumption1 is true. D) Only Assumption2 is true.	L2

iv)	Loss/Error of a Machine Learning (ML) model is decomposed into Bias & variance	L1
v)	High variance and low bias in a machine learning model can lead to overfitting fitting of model.	L1
b.	**(2 x 5 marks = 10 marks)**	
i)	Mathematically derive the estimates of the unknown parameters in a linear regression model using the Ordinary Least Squares (OLS) method.	L5
ii)	Let us assume a binomial logistic regression problem where the classes are pass and fail. The student dataset has entrance mark based on the historic data of those who are selected or not selected. Based on the logistic regression, the values of the learnt parameters are $a_0 = 1$ and $a_1 = 8$. Assuming marks of $x = 60$, compute the resultant class.	L3
c.	**(1 x 10 marks = 10 marks)**	
i)	You are given a dataset containing information about the number of hours students spend studying and their corresponding scores on a test. Your task is to perform simple linear regression to predict test scores based on the number of hours studied using the following dataset.	L4

No.of Hours Studies	Test Scores
2	75
3	82
4	93
5	89
6	98

Analyze whether a linear model is the most suitable for this data. Suggest scenarios where the relationship might require a non-linear approach, and justify your reasoning.

$\beta_0 = 66.2$

$\beta_1 = 5.3$

$y = 66.2 + 5.3x$

PSG COLLEGE OF TECHNOLOGY, COIMBATORE - 641 004

Department of Computer Science and Engineering

BE CSE - G2 & SEM 6

CONTINUOUS ASSESSMENT TEST 2 Date: 24.2.2025

19Z601 – Machine Learning

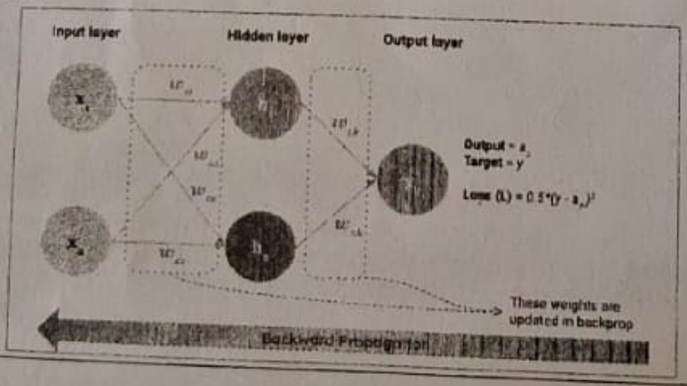
Time: 1 Hour 30 minutes. **Maximum Marks: 50**

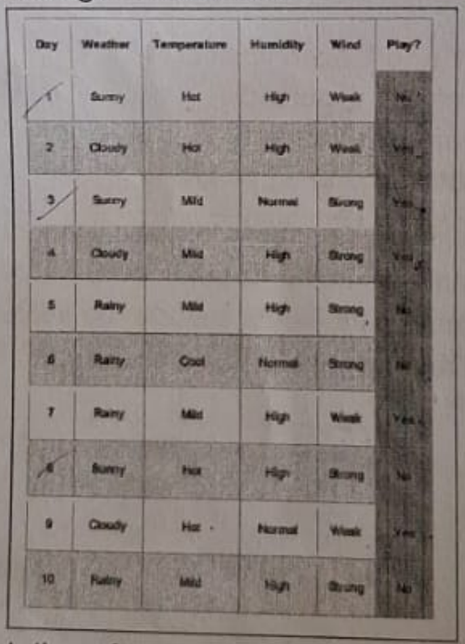
INSTRUCTIONS:

1. Answer **ALL** questions. Each Question carries 25 Marks.
2. In each question, subdivision **a** contains 5 questions and the weightage of each question is one mark, subdivision **b(i)** and **b(ii)** carries 5 marks each and subdivision **c** carries 10 marks each.
3. Subdivisions (a) and (b) will be with no choice and Subdivision (c) may be with choice but not in more than 1 question.
4. Course Outcome Table :

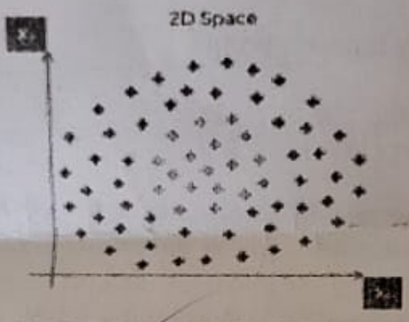
Qn. 1	CO3

Qn.2	CO4

1.a		(5x1mark=5marks)	BTL
i)	Which of the following statements about Decision tree is **TRUE**? S1: Statement 1: Decision tree is a non-parametric, supervised learning algorithm. S2: Statement 2: Entropy represents the randomness in data and information gain is obtained by removing randomness in data. A. S1 and S2 are correct B. Only S1 is correct C. Only S2 is correct D. Both S1 and S2 is wrong		L2
ii)	Neurons in the next layer that receive input from the current neuron are downstream neurons. In other words, Downstream (j) = the set of units whose immediate inputs include the output of unit j. For the Neural Network specified below , find the downstream neurons of X1, X2, H1, H2.  		L3

iii)	Which type of decision tree uses utilizes Gini impurity to identify the ideal attribute to split on. A) ID3 B) C4.5 C) CART D) None of the Above	L1
iv)	Perceptron rule is suitable for linearly separable data.	L2
v)	The weight update rule in Perceptron training rule method is mathematically specified as $w \leftarrow w + \Delta w$ $\Delta w_i \leftarrow \eta(t-o)x_i$	L2
b.	**(2 x 5 marks = 10 marks)**	
i)	Using a perceptron model, design a two input perceptron that implements the Boolean functions OR, NAND, NOR, and XOR. a) Graphically represent on a 2D plane. b) Determine if each function is linearly separable. c) For linearly separable functions, derive appropriate weights and bias values for a perceptron that implements them. d) Explain why XOR cannot be implemented using a single-layer perceptron.	L3
ii)	Derive the backpropagation rule using stochastic gradient descent algorithm by considering the output unit case alone. (Squashing activation function can be used).	L4
c.	**(1 x 10 marks = 10 marks)**	
	Construct a Decision Tree using the ID3 algorithm to classify whether a child will enjoy sport based on the given attributes. Calculate the Information Gain for each attribute at the root node and choose the best splitting attribute. Draw the resulting Decision Tree.  What are the interpretations from the final decision tree obtained ?	L4

Day	Weather	Temperature	Humidity	Wind	Play?
1	Sunny	Hot	High	Weak	No
2	Cloudy	Hot	High	Weak	Yes
3	Sunny	Mild	Normal	Strong	Yes
4	Cloudy	Mild	High	Strong	Yes
5	Rainy	Mild	High	Strong	No
6	Rainy	Cool	Normal	Strong	No
7	Rainy	Mild	High	Weak	Yes
8	Sunny	Hot	High	Strong	No
9	Cloudy	Hot	Normal	Weak	Yes
10	Rainy	Mild	High	Strong	No

2.a	(5x1mark=5marks)	BTL
i)	Which are the statements true with respect to Support Vector Machine (SVM)? S1: Statement 1 : SVM uses the kernel trick that takes as low dimensional space as input and transforms to higher dimensional space. S2 : Support vectors are datapoints near to the hyperplane (decision boundary) A) Both S1 and S2 are true B) Only S1 is true C) Only S2 is true D) Both S1 and S2 are not true	L2
ii)	The Naïve Bayes algorithm mathematically specified as ----------	L3
iii)	For the data points presented in the graph below, what kind of kernel function is applicable to project it into higher dimensional space ?  A) Radial Basis Function B) Polynomial Function C) Sigmoid Function D) Linear Kernel function	L2
iv)	The Naïve part of Naïve Bayes algorithm is ----------	L1
v)	The hyperplane of SVM is represented mathematically as ----------	L1
b.	**(2 x 5 marks = 10 marks)**	
i)	Apply the Naïve Bayes with Laplace smoothing algorithm to the following email example , predict this email is Spam or Ham : You Buy Valium!	L3

Email	Label
Buy Valium!	Spam
You good?	Ham
Valium help you.	Spam
Good Valium help.	Spam
I need Valium for my health condition.	Ham

ii)	A hospital wants to classify whether a patient has a disease (Yes or No) based on medical test results using Support Vector Machine (SVM) and Naïve Bayes (NB). **Compare SVM and Naïve Bayes in terms of:** • Assumptions • Handling of numerical medical data • Performance on small vs. large datasets • Computational complexity • Sensitivity to missing or noisy data • Architectures (Mathematical Intuition)	**L5**
c.	**(1 x 10 marks = 10 marks)**	
i)	Compute the suitable SVM that accurately discriminates the following two classes. Positively labeled points $\binom{3}{1}, \binom{3}{-1}, \binom{6}{1}, \binom{6}{-1}$ Negatively labeled points $\binom{1}{0}, \binom{0}{1}, \binom{0}{-1}, \binom{-1}{0}$ Identify the support vectors, show the SVM architecture, slope and the decision hyperplane. Show all necessary diagrams. Consider the point $\binom{3}{1}$. Predict the label. What are the interpretations from the obtained final hyperplane ?	**L4**

PSG COLLEGE OF TECHNOLOGY, COIMBATORE - 641 004
Department of Computer Science and Engineering
III BE CSE – G2, 6th Semester
CONTINUOUS ASSESSMENT TEST 3
19Z601 - Machine Learning
Date: 15.04.2025

Time: 1 Hour 15 Minutes **Maximum Marks: 35**

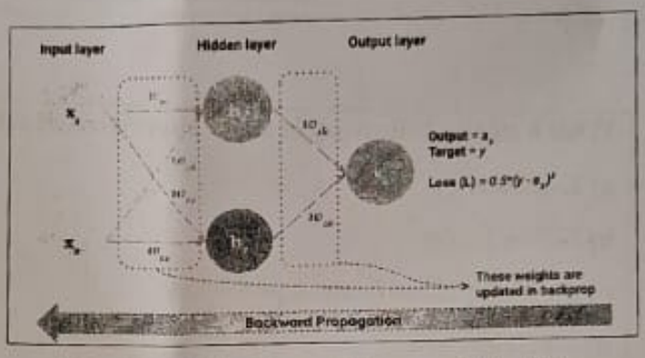
INSTRUCTIONS:

1. Answer **ALL** questions.
2. Question No. **1** carries 8 marks and question No. **2** carries 27 marks
3. In question No. **1**, subdivision **a** carries total of 8 marks (one mark for each question).
4. In question No. 2, subdivision **a** carries total of 7 marks (one mark for each question), subdivisions **b(i)**and **b(ii)** carries 5 marks each and subdivision **c** carries 10 marks.
5. Course Outcome Table :

Qn. 1	CO 1 to 4
Qn.2	CO 5

1. a **(8 x 1 mark = 8 marks)**

i.	**Which of the following is not a supervised machine learning algorithm?** a) K-means b) Naïve Bayes c) SVM for classification problems d) Decision tree **[CO1 – Remember]**
ii.	**_____ model is a generative model used in ML** a) Naive Bayes b) Linear Regression c) Logistic Regression d) Support vector machines **[CO2 – Understand]**
iii.	How is the model's performance affected by the bias-variance trade-off? a) High bias can lead to underfitting, and high variance can lead to overfitting b) Increasing bias improves model accuracy c) Low bias and low variance lead to better ML models d) Increasing variance improves model generalization **[CO2 – Understand]**

iv.	The first principal component accounts for the _____ variance in the data, w subsequent components account for the remaining variance in descending order. **[CO1 – Remember]**	
v.	The Naïve part of Naïve Bayes algorithm is ---------- **[CO4 – Understand]**	
vi.	Neurons in the next layer that receive input from the current neuron are downstream neurons. In other words, Downstream (j) = the set of units whose immediate inputs include the output of unit j. For the Neural Network specified below , find the downstream neurons of X1, X2, H1, H2.  Input Layer Neurons : X1, X2; Hidden Layer Neurons : H1, H2; Output Layer Neurons : Yk **[CO3– Apply]**	
vii.	Which type of decision tree uses utilizes Gini impurity to identify the ideal attribute to split on. A) ID3 B) C4.5 C) CART D) None of the Above **[CO3 – Remember]**	
viii.	Which are the statements true with respect to Support Vector Machine (SVM)? S1: Statement 1 : SVM uses the kernel trick that takes as low dimensional space as input and transforms to higher dimensional space. S2 : Support vectors are datapoints near to the hyperplane (decision boundary) A) Both S1 and S2 are true B) Only S1 is true C) Only S2 is true D) Both S1 and S2 are not true **[CO4– Remember]**	
2.a.		**(7 x 1 mark = 7 marks)**
i)	Given the two vectors A = [1, 2] and B = [2, 3], what is the cosine distance between them?	

	A) 0.07 B) 0.22 C) 0.33 D) 0.45 **[CO5– apply]**
ii)	What is the Levenshtein (edit) distance between the words "form" and "fork"? A) 0 B) 1 C) 2 D) 3 **[CO5– Apply]**
iii)	In Q-learning, what does the Q-value represent? A) The action taken by the agent in a given state. B) The expected future reward for a given state-action pair. C) The discount factor used in the learning process. D) The learning rate used in updating the Q-values. **[CO5– Remember]**
iv)	Which of the following statements is true about k-means clustering? A) K-means clustering is a supervised learning algorithm. B) K-means clustering requires the number of clusters (k) to be known beforehand. C) K-means clustering can only be used for numerical data. D) K-means clustering always converges to the global minimum. **[CO5– Remember]**
v)	Which of the following is not a type of clustering algorithm? A) K-means clustering B) Hierarchical clustering C) DBSCAN (Density-Based Spatial Clustering of Applications with Noise) D) K-Nearest Neighbors (KNN) **[CO5– Remember]**
vi)	L1 distance measure is otherwise known as ---------- **[CO5– Remember]**
vii)	L∞ distance measure is represented mathematically as ---------- **[CO5– Understand]**
b.	**(2 x 5 marks = 10 marks)**
i)	Calculate the SMC and Jaccard Similarity Co-efficient for the following two binary vectors. $\mathbf{x} = (1,0,0,0,0,0,0,0,0,0)$ $\mathbf{y} = (0,0,0,0,0,0,1,0,0,1)$ $f_{01} = 2$ the number of attributes where **x** was 0 and **y** was 1 $f_{10} = 1$ the number of attributes where **x** was 1 and **y** was 0 $f_{00} = 7$ the number of attributes where **x** was 0 and **y** was 0 $f_{11} = 0$ the number of attributes where **x** was 1 and **y** was 1 **[CO5– Apply]**

ii) What is Reinforcement Learning ? Explain the steps involved in Q Learning .

[CO5– Understand]

c. **(1 x 10 marks = 10 marks)**

i) For the given data, compute two clusters using K-means algorithm for clustering where initial cluster centers are (1.0,1.0) and (5.0,7.0). Execute for two iterations.

Record Number	A	B
R1	1.0	1.0
R2	1.5	2.0
R3	3.0	4.0
R4	5.0	7.0
R5	3.5	5.0
R6	4.5	5.0
R7	3.5	4.5

[CO5– Apply]

OR

ii) Use the distance matrix in table given below to perform single link and complete link hierarchical clustering. Show your results by drawing a dendrogram. The dendrogram should clearly show the order in which the points are merged.

	P1	P2	P3	P4	P5
P1	0.00	0.10	0.41	0.55	0.35
P2	0.10	0.00	0.64	0.47	0.98
P3	0.41	0.64	0.00	0.44	0.85
P4	0.55	0.47	0.44	0.00	0.76
P5	0.35	0.98	0.85	0.76	0.00

[CO5– Apply]