



# VIT

Vellore Institute of Technology  
(Deemed to be University under section 3 of UOE Act, 1986)

REG.NO.:

## SCHOOL OF COMPUTER SCIENCE AND ENGINEERING CONTINUOUS ASSESSMENT TEST - II FALL SEMESTER 2025-2026

SLOT: F1+TF1

Programme Name & Branch : B.Tech- CSE  
 Course Code and Course Name : BCSE334L- Predictive Analytics  
 Faculty Name(s) : Dr. VIJAYASHERLY V, Dr. SAYAN SIKDER , Dr K SANTHI,  
 Dr. THIRUNAVUKKARASAN M , Dr . ANURADHA J, Dr. UMA PRIYA D  
 Class Number(s) : VL2025260101653 , VL2025260101658, VL2025260101650 ,  
 VL2025260101656, VL2025260101648 , VL2025260101661  
 Date of Examination : 10-Oct-2025  
 Exam Duration : 90 minutes Maximum Marks: 50

**General instruction(s):** Answer All Questions

CO – Course Outcome; BL – Blooms Taxonomy Level (1 – Remember, 2 – Understand, 3 – Apply, 4 – Analyse, 5 – Evaluate, 6 – Create)

Course Outcomes : CO3:Apply regression and classification model on applications for decision making and evaluate the performance.CO4:Analyze the impact of class imbalance on performance measure for model predictions and models that can mitigate the issue during training.

Q. No	Question	M	CO	BL																																				
1.	<p>Consider the performance of eight basketball players</p> <table border="1"> <thead> <tr> <th>Player ID</th> <th>Points Per Game (y)</th> <th>Hours Trained Weekly (X<sub>1</sub>)</th> <th>Age (X<sub>2</sub>)</th> </tr> </thead> <tbody> <tr><td>1</td><td>140</td><td>60</td><td>22</td></tr> <tr><td>2</td><td>155</td><td>62</td><td>25</td></tr> <tr><td>3</td><td>159</td><td>67</td><td>23</td></tr> <tr><td>4</td><td>179</td><td>70</td><td>20</td></tr> <tr><td>5</td><td>192</td><td>71</td><td>15</td></tr> <tr><td>6</td><td>200</td><td>72</td><td>14</td></tr> <tr><td>7</td><td>212</td><td>75</td><td>14</td></tr> <tr><td>8</td><td>215</td><td>78</td><td>11</td></tr> </tbody> </table> <p>Using the multiple linear regression model a) Calculate the regression coefficients: <math>\beta_0</math>, <math>\beta_1</math>, and <math>\beta_2</math>. b) Interpret the meaning of the coefficients for "Hours Trained" and "Age".c) Predict the "Points Per Game" for a player who trains 65 hours a week and is 19 years old.</p>	Player ID	Points Per Game (y)	Hours Trained Weekly (X <sub>1</sub> )	Age (X <sub>2</sub> )	1	140	60	22	2	155	62	25	3	159	67	23	4	179	70	20	5	192	71	15	6	200	72	14	7	212	75	14	8	215	78	11	10	CO3	3
Player ID	Points Per Game (y)	Hours Trained Weekly (X <sub>1</sub> )	Age (X <sub>2</sub> )																																					
1	140	60	22																																					
2	155	62	25																																					
3	159	67	23																																					
4	179	70	20																																					
5	192	71	15																																					
6	200	72	14																																					
7	212	75	14																																					
8	215	78	11																																					
2.	<p>Analyze the following student data using a pre-trained SVR model with Hyperplane: Predicted Score = <math>1.5 * (\text{Hours Studied}) + 75</math> and Margin (<math>\epsilon</math>): 4 points</p> <table border="1"> <thead> <tr> <th>Student</th> <th>Hours Studied (X)</th> <th>Actual Score (Y)</th> </tr> </thead> <tbody> <tr><td>1</td><td>5</td><td>70</td></tr> <tr><td>2</td><td>8</td><td>78</td></tr> <tr><td>3</td><td>10</td><td>85</td></tr> <tr><td>4</td><td>2</td><td>60</td></tr> <tr><td>5</td><td>20</td><td>95</td></tr> <tr><td>6</td><td>15</td><td>98</td></tr> </tbody> </table> <p>Identify the support vectors. You must show your calculations for the predicted score and error for each student to justify your answer.</p>	Student	Hours Studied (X)	Actual Score (Y)	1	5	70	2	8	78	3	10	85	4	2	60	5	20	95	6	15	98	10	CO3	3															
Student	Hours Studied (X)	Actual Score (Y)																																						
1	5	70																																						
2	8	78																																						
3	10	85																																						
4	2	60																																						
5	20	95																																						
6	15	98																																						
3.	<p>Apply the k-Nearest Neighbors algorithm (with <math>k=3</math>) to determine the primary use case for a new Generative AI model with the features {Primary Modality: Image, Architecture: Diffusion, Training Data: Web &amp; Images}. Your classification must be justified by identifying the model's nearest neighbors from the provided dataset.</p>	10	CO3	3																																				



VIT<sup>®</sup>

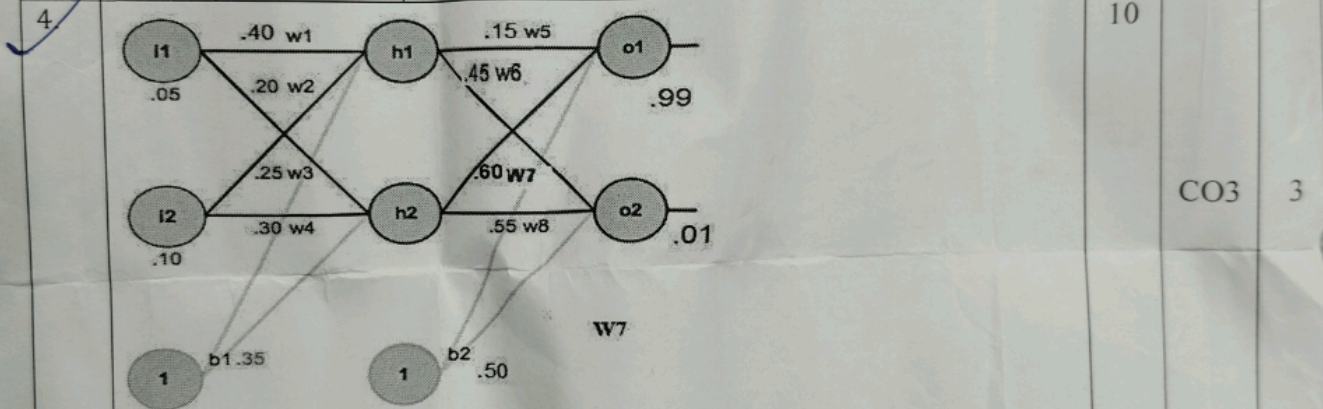
Vellore Institute of Technology  
(Deemed to be University under section 3 of UGC Act, 1956)

REG.NO.:

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING  
CONTINUOUS ASSESSMENT TEST - II  
FALL SEMESTER 2025-2026

SLOT: F1+TF1

Model ID	Primary Modality	Architecture	Training Data	Primary Use Case
1	Text	Transformer	Books & Web	Text Generation
2	Image ✓	Diffusion ✓	Images ✓	Image Generation
3	Text	Transformer	Code	Code Generation
4	Image ✓	GAN	Faces	Image Generation
5	Text	Transformer	Web & Code	Text Generation
6	Image ✓	Diffusion ✓	Art & Images ✓	Image Generation
7	Text	LSTM	Scientific Papers	Text Generation
8	Audio	Transformer	Speech	Audio Generation
9	Image ✓	GAN	Medical Scans ✓	Image Generation
10	Text	Transformer	Code & Web	Code Generation



Using the Sigmoid activation function and a learning rate of 0.5, perform one training step on the given neural network. Calculate the new values for all weights and bias units after backpropagating the error from the target outputs.

5. a) ✓	A hospital is choosing between two AI models, "MediScan" and "PathoFind," to analyze 5,000 patient scans for a rare cancer. In the test data, 100 scans are truly cancerous. MediScan correctly identified 88 of these cancerous scans but also incorrectly flagged 196 healthy scans. In contrast, PathoFind correctly identified 96 cancerous scans but incorrectly flagged 392 healthy ones. Analyze both models by calculating their respective precision and recall. Based on your calculations, write a concise recommendation for the hospital. You must justify which model is the better choice by explaining the clinical trade-off between minimizing false positives versus minimizing false negatives in this specific context.	6	CO4	4
b)	Describe the significance of the metric "coefficient of determination" with an example visualization of a two-dimensional dataset appropriately.	4		

\*\*\*\*\*