USN | | | | | | | | | | |

Internal Assessment Test II – May 2025

Sub:	**Machine Learning**			Sub Code:	**BCS602**	Branch:	**AInDS**
Date:		Duration:	**90 minutes**	Max Marks:	**50**	Sem	**VI**	**OBE**

		Answer any FIVE Questions	**MARKS**	**CO**	**RBT**
1	a	(see below)	[7]	C03	
	b	**Explain the Types of Regression Methods** Regression methods are statistical techniques used to model and predict relationships between variables. (1 Mark)	[3]	C03	

1 a

Consider the following training dataset of 10 data instances shown in Table 4.12 which describes the award performance of individual students based on GPA and No. of projects done. The target variable is 'Award' which is a discrete valued variable that takes 2 values 'Yes' or 'No'.

Table 4.12: Training Dataset

S.No.	GPA	No. of Projects Done	Award
1.	9.5	5	Yes
2.	8.0	4	Yes
3.	7.2	1	No
4.	6.5	5	Yes
5.	9.5	4	Yes
6.	3.2	1	No
7.	6.6	1	No
8.	5.4	1	No
9.	8.9	3	Yes
10.	7.2	4	Yes

Given a test instance (GPA - 7.8, No. of projects done - 4), use the training set to classify the test instance. Choose $k = 3$.

- k-Nearest Neighbor classifier

Test → (GPA - 7.8, No. of projects done - 4)

$k = 3$

sl.no	GPA	No. of projects done	Award	Euclidian distance
1	9.5	5	Yes	$\sqrt{(9.5-7.8)^2+(5-4)^2}$
2	8.0	4	Yes	$\sqrt{(8.0-7.8)^2+(4-4)^2}$
3	7.2	1	No	$\sqrt{(7.2-7.8)^2+(1-4)^2}$
4	6.5	5	Yes	$\sqrt{(6.5-7.8)^2+(5-4)^2}$
5	9.5	4	Yes	$\sqrt{(9.5-7.8)^2+(4-4)^2}$
6	3.2	1	No	$\sqrt{(3.2-7.8)^2+(1-4)^2}$
7	6.6	1	No	$\sqrt{(6.6-7.8)^2+(1-4)^2}$
8	5.4	1	No	$\sqrt{(5.4-7.8)^2+(1-4)^2}$
9	8.9	3	Yes	$\sqrt{(8.9-7.8)^2+(3-4)^2}$
10	7.2	4	Yes	$\sqrt{(7.2-7.8)^2+(4-4)^2}$

1) 1.9723
2) 0.2
3) 3.0594
4) 1.6401
5) 1.7
6) 5.4918
7) 3.2310
8) 3.8418
9) 1.4866
10) 0.6

Ascending order & select first 3.

0.2, 0.6, 1.4866
(2) (10) (9)
yes yes yes.

→ Yes

Simple Linear Regression: Predicts a single dependent variable based on one independent variable. (1 Mark)

Polynomial Regression: Models non-linear relationships by adding polynomial terms to the independent variables. (1 Mark)

2

Explain the Structure, Types, Advantages and Disadvantages of ANN. List the Activations Used in ANN.

A neural network is structured as interconnected layers of artificial neurons, with input, hidden, and output layers. These layers are connected by weights and biases, which determine the strength and influence of connections between neurons. (2 Marks) Diagram (1 Mark)

Types include feedforward : where information flows in one direction from input to output., recurrent : Designed for sequential data, such as time series or natural language, and convolutional neural networks : Specialized for image processing and pattern recognition.Use convolutional layers to learn hierarchical features from input data.(3 Marks)

Advantages : Artificial Neural Networks offer numerous advantages, including their ability to learn complex relationships from data, process information in parallel, and handle noisy or incomplete data. They can also automatically extract features, adapt to new information, and generalize well to unseen data. (1 Mark)

DisAdvantages : including difficulty in interpretability (the "black box" nature), high computational costs, potential for overfitting, and the need for large amounts of data (1 Mark)

Activations Function :

- Linear activation.
- Sigmoid activation.
- Tanh (hyperbolic tangent) activation.
- ReLU (rectified linear unit) activation.
- Softmax activation. (2 Marks)

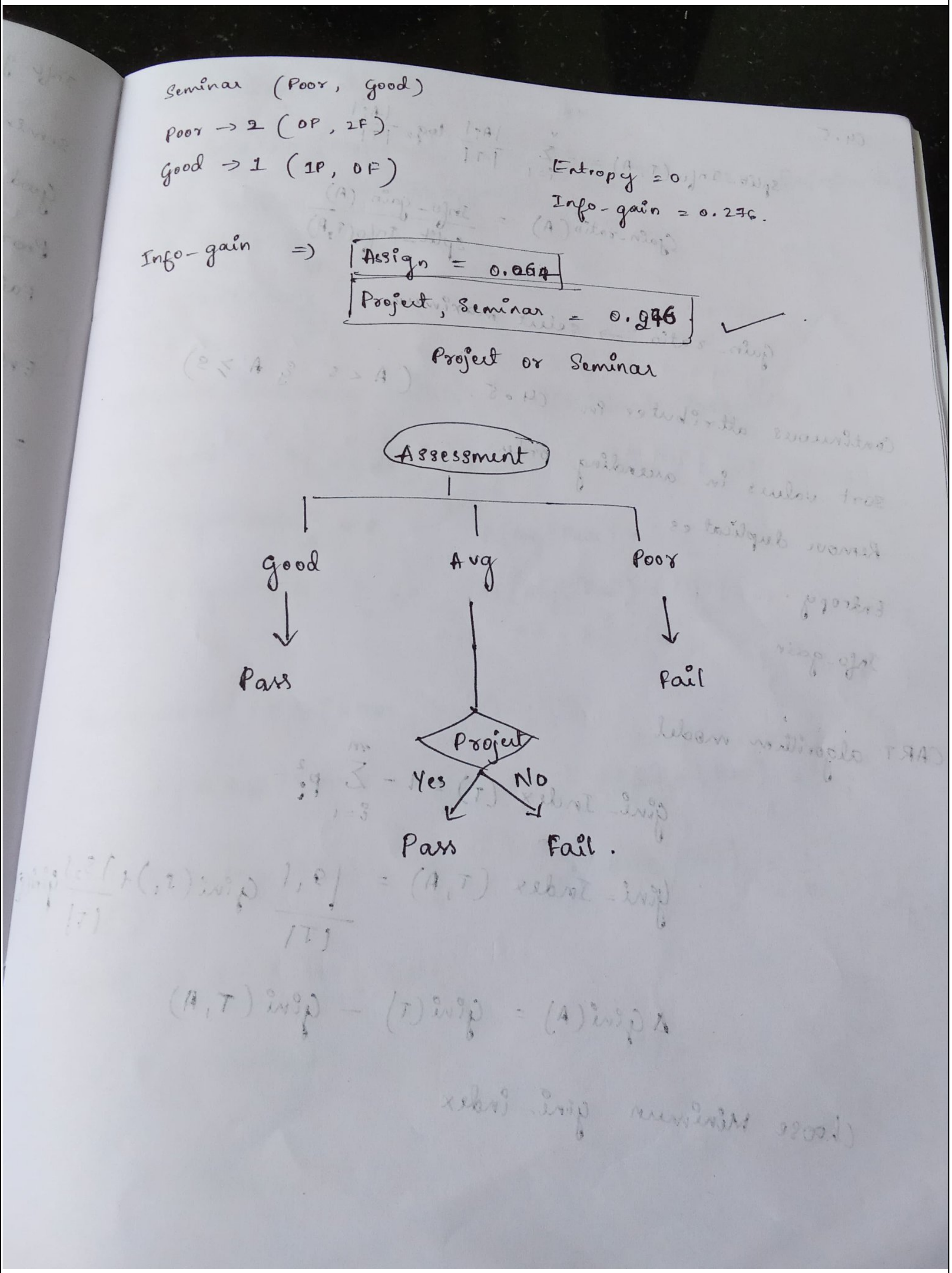
[8+2] C04

3

S.No.	Assessment	Assignment	Project	Seminar	Result
1.	Good	Yes	Yes	Good	Pass
2.	Average	Yes	No	Poor	Fail
3.	Good	No	Yes	Good	Pass
4.	Poor	No	No	Poor	Fail
5.	Good	Yes	Yes	Good	Pass
6.	Average	No	Yes	Good	Pass
7.	Good	No	No	Fair	Pass
8.	Poor	Yes	Yes	Good	Fail
9.	Average	No	No	Poor	Fail
10.	Good	Yes	Yes	Fair	Pass

Consider the Training Dataset and Construct Decision Tree using ID3.

[10] C03

					
4	a	**Explain about Proximity Measures** Proximity measures, also known as similarity or dissimilarity measures, quantify how alike or different data points are Euclidean Distance: $d(\mathbf{p},\mathbf{q}) = \sqrt{\sum_{i=1}^{n}(q_i - p_i)^2}$ Manhattan Distance: d = |x_1 - x_2| + |y_1 - y_2|) Jaccard Index: J = d / b+c+d Hamming Distance:D(x, y) = $\sum (x_i \neq y_i)$	[5]	C05	

<table>
<tr><td></td><td></td><td>Minkowski Distance: D(x, y) = (∑|xi - yi|^p)^(1/p)</td><td></td><td></td><td></td></tr>
<tr><td></td><td>b</td><td>Consider the following dataset in Table 5.11 where the week and number of working hours per week spent by a research scholar in a library are tabulated. Based on the dataset, predict the number of hours that will be spent by the research scholar in the 7th and 9th week. Apply linear regression model.

Table 5.11: Sample Data
<table>
<tr><td>x, (Week)</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr>
<tr><td>y, (Hours Spent)</td><td>12</td><td>18</td><td>22</td><td>28</td><td>35</td></tr>
</table>
Linear Regression model
$y = a_0 + a_1 x$
$a_0 = (\bar{y}) - a_1 \bar{x}$
$a_1 = \dfrac{(\overline{xy}) - (\bar{x})(\bar{y})}{(\overline{x_i^2}) - (\bar{x})^2}$
<table>
<tr><td>x_i (week)</td><td>y_i (Hour spent)</td><td>x_i^2</td><td>$x_i \times y_i$</td></tr>
<tr><td>1</td><td>12</td><td>1</td><td>12</td></tr>
<tr><td>2</td><td>18</td><td>4</td><td>36</td></tr>
<tr><td>3</td><td>22</td><td>9</td><td>66</td></tr>
<tr><td>4</td><td>28</td><td>16</td><td>112</td></tr>
<tr><td>5</td><td>35</td><td>25</td><td>175.</td></tr>
<tr><td>Sum = 15</td><td>Sum = 115</td><td>Sum = 55</td><td>Sum = 401</td></tr>
<tr><td>$Avg(\bar{x}_i) = \frac{15}{5}$ ⇒ 3</td><td>$Avg(\bar{y}_i) = \frac{115}{5}$ ⇒ 23</td><td>$Avg(\overline{x_i^2}) = \frac{55}{5}$ ⇒ 11</td><td>$Avg(\overline{x_i y_i}) = \frac{401}{5}$ ⇒ 80.2</td></tr>
</table>
$a_1 = \dfrac{80.2 - \overbrace{(3)(23)}^{69}}{11 - 3^2} = \dfrac{28}{5} = 5.6.$
$a_0 = 23 - \underbrace{(5.6)(3)}_{16.8} = \dfrac{31}{5} = 6.2.$
$y = 6.2 + 5.6x \Longrightarrow y = 6.2 + 5.6(7) = 45.4$
$y = 6.2 + 5.6(9) = 56.6$</td><td>[5]</td><td>C03</td><td></td></tr>
<tr><td>5</td><td></td><td>Take a real-time example of predicting the result of a student using Naïve Bayes algorithm. The training dataset T consists of 8 data instances with attributes such as 'Assessment', 'Assignment', 'Project' and 'Seminar' as shown in Table 8.17. The target variable is Result which is classified as Pass or Fail for a candidate student. Given a test data to be (Assessment = Average, Assignment = Yes, Project = No and Seminar = Good), predict the result of the student. Apply Laplace Correction if Zero probability problem occurs.</td><td>[10]</td><td>C04</td><td></td></tr>
</table>

S.No.	Assessment	Assignment	Project	Seminar	Result
1.	Good	Yes	Yes	Good	Pass
2.	Average	Yes	No	Poor	Fail
3.	Good	No	Yes	Good	Pass
4.	Average	No	No	Poor	Fail
5.	Average	No	Yes	Good	Pass
6.	Good	No	No	Poor	Pass
7.	Average	Yes	Yes	Good	Fail
8.	Good	Yes	Yes	Poor	Pass

Ans (12)

S.N	Assessment	Assignment	Proj	Seminar	Result
1	Good	Yes	Yes	Good	pass.
2	Avg	Yes	No	Poor	fail
3	Good	No	Yes	Good	pass
4	Avg	No	No	Poor	fail
5	Avg	No	Yes	Good	pass
6	Good	No	No	poor	pass
7	Avg	Yes	Yes	Good	fail
8	Good	Yes	No	poor	pass

1.97 Step1 target class – Result

Result → pass (5), fail (3)

prob = $\frac{5}{8}$, $\frac{3}{8}$

$(Result = p) \Rightarrow \frac{1}{5} \times \frac{2}{5} \times \frac{2}{5} \times \frac{3}{5} \times \frac{5}{8}$

$\Rightarrow \frac{3}{250} = 0.012$

$(Result = F) = \frac{3}{3} \times \frac{2}{3} \times \frac{2}{3} \times \frac{1}{3} \times \frac{3}{8} \Rightarrow \frac{1}{18}$

$\Rightarrow 0.055$

frequency Matrix			likelihood prob	
assessment	Pass	fail	prob(Pass)	Prob (fail)
Good	4	0	4/5	0/3
Avg	1	3	1/5	3/3
total	5	3		

freq mat			likelihood P	
Assign	Pass	fail	Pass	fail
Yes	2	2	2/5	2/3
No	3	1	3/5	1/3

freq M			likeli Prob	
Proj	Pass	fail	Pass	fail
Yes	3	1	3/5	1/3
No	2	2	2/5	2/3

freq Mat			likelihood prob	
Seminar	Pass	fail	Pass	fail
Good	3	1	3/5	1/3
poor	2	2	2/5	2/3

es award.

6		Consider the following data shown in Table 13.15. Use k-means algorithm with $k = 2$ and show the result.	[10]	C05	

Table 13.15: Sample Data

S.No.	X	Y
1.	3	5
2.	7	8
3.	12	5
4.	16	9

		18)			

Sr no	x	y
1	3	5
2	7	8
3	12	5
4	16	9

K = 2

Take any 2 values as centroid

cluster 1	cluster 2
(3,5)	(16,9)
centroid (3,5)	centroid (16,9)

Now for remaining points find Euclidean dist wrt each cluster and whichever has lesser value add into it i.e

point (7,8) from c_1 = $\sqrt{(7-3)^2+(8-5)^2} = \sqrt{4^2+3^2} = 5$

from c_2 = $\sqrt{(16-7)^2+(8-9)^2} = \sqrt{9^2+1^2} = 9.055$

∴ point (7,8) goes to cluster 1

point (12,5) from c_1 = $\sqrt{(12-3)^2+(5-5)^2} = \sqrt{9^2} = 9$

from c_2 = $\sqrt{(12-16)^2+(5-9)^2} = \sqrt{4^2+4^2} = 5.65$

∴ point (12,5) goes to cluster 2

New table

cluster 1	cluster 2
(3,5)	(12,5)
(7,8)	(16,9)
centroid (3,5)	centroid (16,9)

Now calculate new centroid

cluster 1 : (3,5) (7,8)

mean x = (3+7)/2 = 5

mean y (5+8)/2 = 6.5

New centroid 1 :- (5,6.5)

cluster 2 : (12,5) (16,9)

mean x = 14

mean y = 7

New centroid 2 : (14,7)

point	centroid 1 (5,6.5)	centroid 2 (14,7)	cluster
(3,5)	$\sqrt{2^2+1.5^2} = 2.5$	11.18	1
(7,8)	2.5	7.07	1
(12,5)	7.16	2.83	2
(16,9)	11.26	2.83	2

cluster 1	cluster 2
(3,5)	(12,5)
(7,8)	16,9
centroid (5,6.5)	centroid (14,7)

CI **CCI** **HOD**