

Roll No:

(To be filled in by the candidate)

PSG COLLEGE OF TECHNOLOGY, COIMBATORE - 641 004

CONTINUOUS ASSESSMENT TEST -1

MSc Cyber Security Semester: 5

23XC53-MACHINE LEARNING

Time : 1.5 Hours

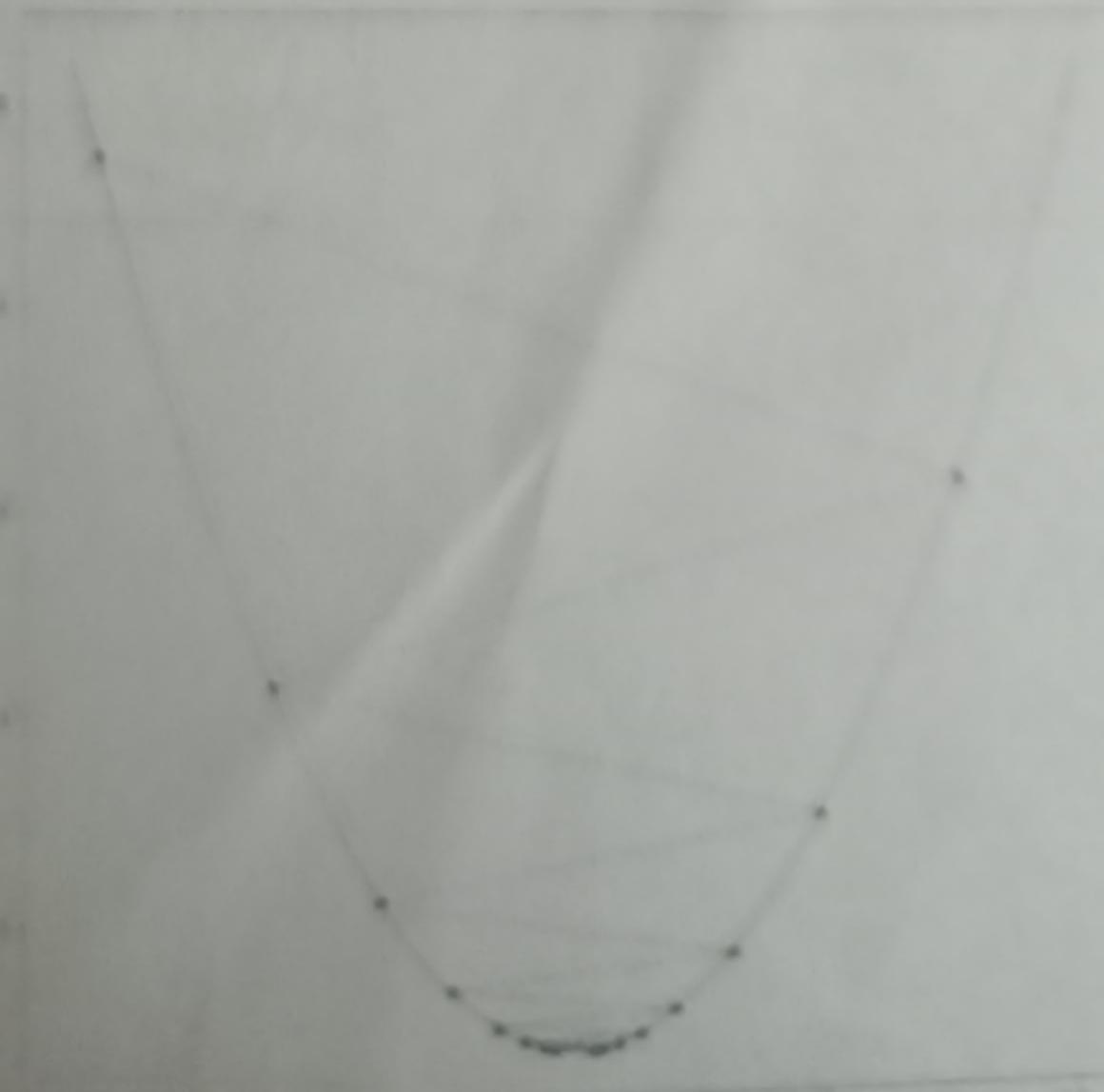
Maximum Marks : 40

INSTRUCTIONS:

1. Answer ALL questions. Each question carries 20 Marks.
2. Subdivision (a) carries 2+2 marks each, subdivision (b) carries 6 marks each and subdivision (c) carries 10 marks each.
3. Course Outcome Table

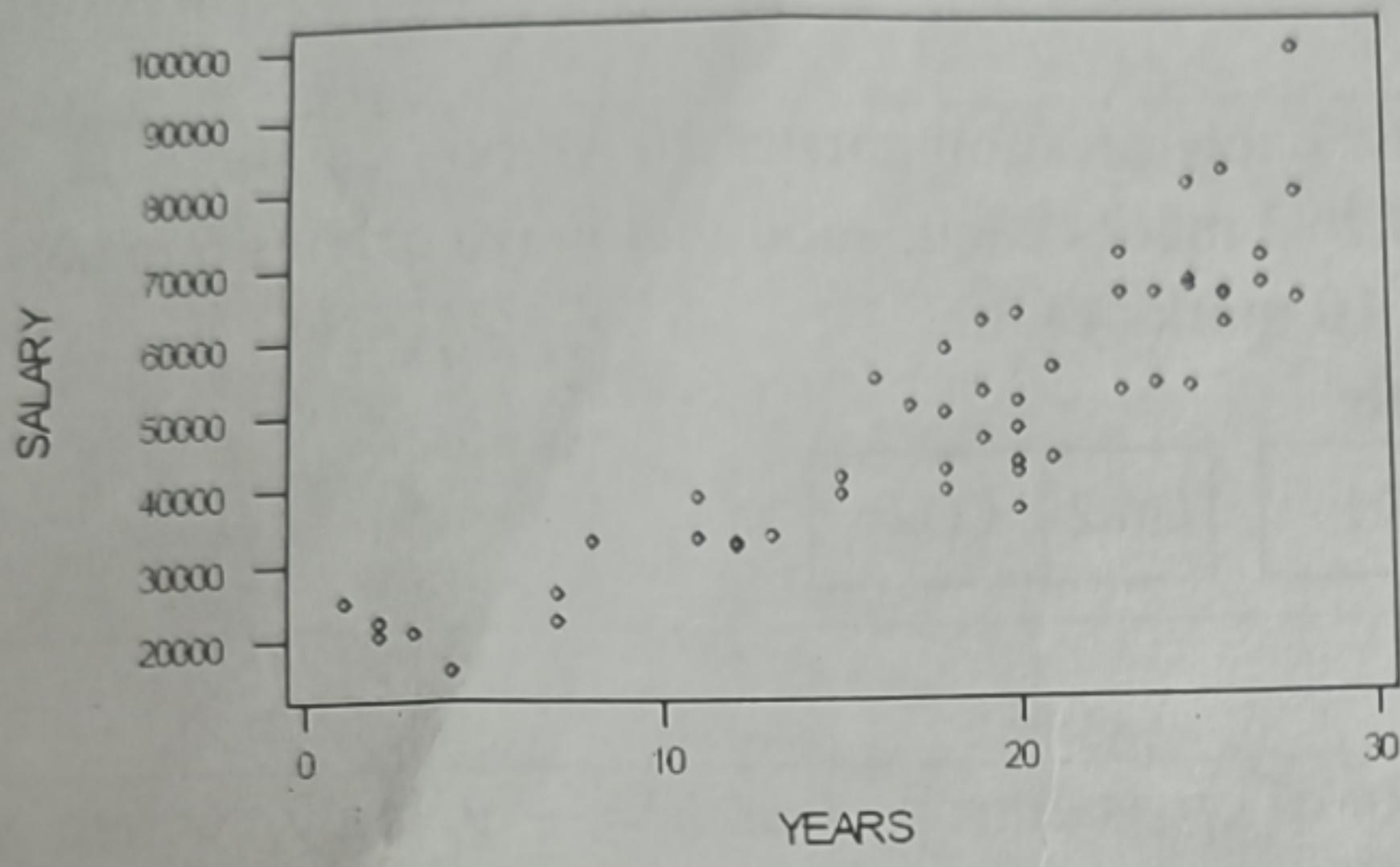
Qn.1	CO1	Qn.2	CO2
------	-----	------	-----

I.	a.i.	Check the convexity of cross entropy loss $L = -y_i \log \hat{y}_i$ where $\hat{y}_i = W^T X_i$. (L5)
	ii.	Comment on "K-Nearest Neighbor is a parametric algorithm". Justify your answer with proper reasons. (L3)
	b.	Explain different outlier detection techniques. (L2)
	c.	Write a gradient descent algorithm to find the parameter of Simple linear regression. What are the possible criteria for convergence in Gradient descent (GD) method? Justify the convergence mathematically. (L4) Given a cost function $f(\theta) = \theta^2 - \theta + 1$, what is the weight updation equation for the gradient descent technique at iteration t+1? Consider eta = 0.2 and initial $\theta = 0.2$, find θ which minimizes $f(\theta)$ at the end of third iteration. Consider the figure given below. Name the scenario which occurs in gradient descent method. Why does it happen?



2. a.i. The scatter plot shows the employees' years of experience vs their salary for a sample of 50 managers. The regression model obtained from data is $\text{salary} = 11369 + 2141 * \text{years}$. Find the following: (L3)

- sum of residuals of the model?
- Strength of association or relationship between years and salary



- ii. Given a multiple regression model $y = a + bx + cz$ where x, y, and z are independent variables with SSE, SST and SSR answer the following: (L3)
- Given a data point X, how do you check whether X is an outlier?
 - What is the impact of R^2 in inferring the model? How does it overcome?

- b. What is multi collinearity? What is the impact of multi-collinearity in linear regression? Explain different methods to remove multi-collinearity from a data set. (L4)

- c. Given a data set with an independent variable abdominal perimeter and a dependant variable foetal weight, construct a linear regression model using closed form solution (Matrix based). Analyze the relationship between dependant and independent variables from the parameters. Predict foetal weight when perimeter is 30 mm. (L6)

Abdominal Perimeter (in mm)	28	31	33	34	32
Foetal weight (in kgs)	2	2.5	3	3.5	4