



# VIT

Vellore Institute of Technology

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

CAT 1 Examination – January 2025

Programme	: B.Tech CSE	Semester	: Winter 24-25
Course Title	: Machine Learning	Code	: BCSE209L
Faculty (s)	: Dr.Priyadarshini.J Dr. Bhargavi R Dr. Jothi R	Slot	: D1+TD1
		Class Nbr	: CH2024250501910 CH2024250501914 CH2024250501905
Time	: 1 1/2 Hours	Max. Marks	: 50

Answer all the Questions

1. Table 1 shows the Body Mass and Running Speed of different animals.

Table 1

Animal	Body Mass (Kg)	Running Speed (Km/hr)
Rhino	1400	45
Horse	400	70
Pronghorn	50	100
Giraffe	1000	60
wildebeest	300	90
Cheetah	60	110

15

- a) Identify the type of relation that exists between the body mass and the running speed of the animals. You may roughly plot the values to check the relationship. (2 marks)
- b) Analyze the data and train a suitable machine learning approach for predicting the running speed of animals. (11 marks)
- c) Now, predict the running speed for the rabbit, if the body mass is 50 kg. (2 marks)

2. Consider the truth table for XOR gate given below and apply the activation function  $f$  as given for the input variables  $X_1$ ,  $X_2$  and Target/Output variable  $Z_1$ . Assume the learning rate = 1.5, threshold ( $\theta$ ) = 1 and the initial weights with bias = 0. Show how the perceptron model is trained on the data given for 3 epochs.

15

$X_1$	$X_2$	$Z_1$
0	0	0



0	1	0
1	0	1
1	1	0

The activation function

$$f(y_{in}) = \begin{cases} 1 & \text{if } y_{in} \geq \theta \\ 0 & \text{if } y_{in} < \theta \end{cases}$$

3.

Consider each of the following scenarios :

1. We collect a set of data on the top 500 firms in India. For each firm we record profit, number of employees, industry and the CEO salary. We are interested in understanding which factors affect CEO salary.
2. An Insurance company is considering launching a new product and wish to know whether it will be a success or a failure. The Insurance company collects data on 50 similar products that were previously launched. For each product they have recorded whether it was a success or failure, price charged for the product, marketing budget, competition price, and ten other variables.
3. We are interested in predicting the % change in the Tumor in relation to the weekly changes in different Medicines. Hence we collect weekly data for the complete year. For each week we record the % change in the Tumor, the % change in the Medicine dosage, the % change in the Medicine 2 dosage, and the % change in the Medicine3 dosage.

- a) Explain whether each scenario given is a classification or regression problem. (3 marks)
- b) Indicate whether we are most interested in inference or prediction. Also, mention one ML model for the task and how the task is achieved. (6 Marks)

4.

You have a movie dataset. Dataset has various features like average\_review\_audience, number\_of\_theaters, release\_date, boxOffice\_success etc. collected from different sources for different movies along with the movie\_category. You are required to develop a machine learning model that predicts the movie\_category using all other features. There are 5 categories to which a movie can belong to. You are required to train a linear classifier that predicts movies category, but the input data has non-linear relationship with the movie\_category. Assume that your ML library has only a binary classifier that can work fine on the linearly separable data.

- a. How do you address this problem of non-linear relation? Explain. (2 marks)
- b. How can you use the binary classifier for predicting 5 classes? Explain. (4 marks)

5

You are training a machine learning model using gradient descent.

- a. What is the role of the learning rate in gradient descent? What happens if it is too large or too small? (2 marks)
- b. Suppose the loss function is not decreasing during training. List three potential reasons and their solutions. (3 marks)

