Roll No.: _

Amrita Vishwa Vidyapeetham

Amrita School of Computing, Coimbatore

B.Tech. First Assessment Examinations – October 2022

Fifth Semester

Computer Science and Engineering

19CSE305 Machine Learning

Duration: Two hours Maximum: 50 Marks

CO#	Course Outcomes
CO01	Understand issues and challenges of machine learning: data, model selection, model complexity
CO02	Design and implement various machine learning algorithms in a range of real-world applications
CO03	Understand strengths and weaknesses of many popular machine learning approaches
CO04	Analyze the underlying mathematical relationships within and across Machine Learning algorithms
CO05	Apply the paradigms of supervised and un-supervised learning

Answer all questions

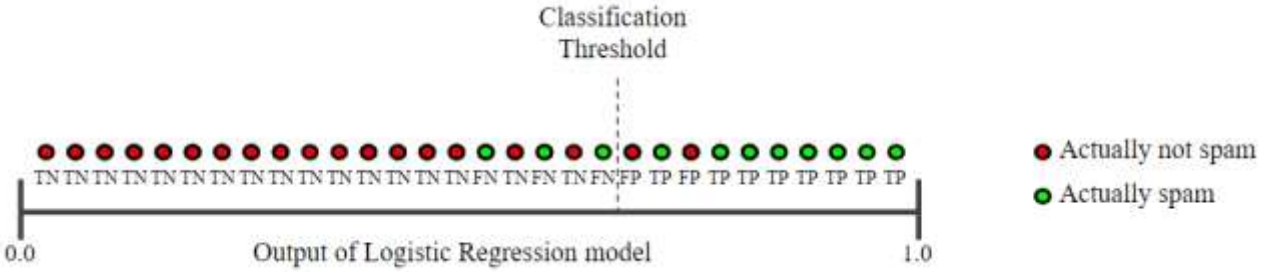
1) Match the following attributes to their best-matching data type and type. Justify your choices with examples. [3] [CO1] [BTL2]

Attribute	Datatype	Type
Age	Floating point number	Label - Nominal
Salary	Integers	Numeric - Continuous
Movie Rating	Int	Numeric - Discrete
Fruit	String	Interval

2) When would accuracy be a bad metric when compared to F1-score? Explain your answer in terms of: a) True vs False predictions, b) Balance among classes in the dataset. [3] [CO1] [BTL2]

3) Explore precision and recall by looking at the following figure, which shows 30 predictions made by an email classification model. Those to the right of the classification threshold are classified as “spam", while those to the left are classified as "not spam." Calculate the precision and recall.

[3] [CO1] [BTL1]



4) A. The given distribution shows the number of runs scored by some top batsmen of the world in one-day international cricket matches. Find the mode. [1.5] [CO1] [BTL2]

Run Scored	Number of Batsman
3000-4000	4
4000-5000	18
5000-6000	9
6000-7000	7
7000-8000	6
8000-9000	3
9000-10000	1
10000-11000	1

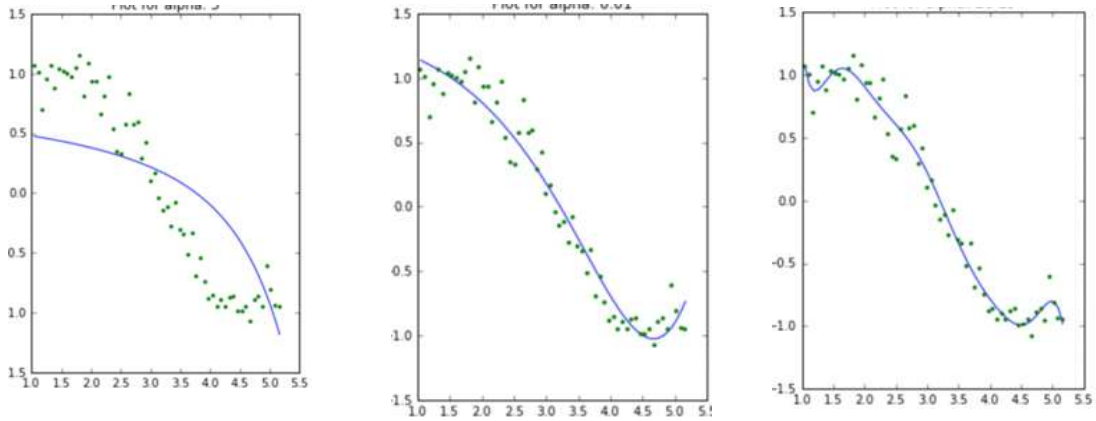
B. What is the median of the following data set? [1.5] [CO1] [BTL2]
32, 6, 21, 10, 8, 11, 12, 36, 17, 16, 15, 18, 40, 24, 21, 23, 24, 24, 29, 16, 32, 31, 10, 30, 35, 32, 18, 39, 12, 20

5) How is the error calculated in a Linear Regression model? Explain with a suitable diagram. [3] [CO4] [BTL1]

6) Explain the trade-off between bias and variance with a suitable diagram. [3] [CO1] [BTL4]

7) A. Identify the type of learning adopted in each of the following algorithm K Nearest Neighbour, K – Means clustering algorithm, Decision tree [1.5] [CO1] [BTL2]

B. Identify the underfit, overfit, and balanced fit in the following models [1.5] [CO1] [BTL2]



(i) (ii) (iii)

8) How does Regression differ from Classification? Explain this in detail with some good representative examples for each. [3] [CO4] [BTL2]

9) Illustrate with a diagram how 5-fold cross-validation will be done with a dataset that consists of 40 instances in it. Also, how is the model's skill reported? [3] [CO1] [BTL2]

10) Compare and contrast supervised and unsupervised learning. [3] [CO1] [BTL2]

11) For the confusion matrix below, compute sensitivity, specificity, precision, negative predictive value, false positive rate, false negative rate, accuracy, and F1-score. [5] [CO4] [BTL3]

	Ground Truth True	Ground Truth False
Predicted True	80	5
Predicted False	5	10

12) For the given set of data, use linear regression to fit a curve. Also find whether the equation is a proper fit for the given data using relevant metrics. [5] [CO4] [BTL3]

Sl.no	**Size (in square feet)**	**Prize (in lakhs)**
1.	2324	437
2.	2500	460
3.	2978	500
4.	3126	540
5.	3540	678

13) Estimate the slopes and intercept for the dataset provided below using Ordinary Least Square method and write down the equation used for making the predictions. [5] [CO4] [BTL3]

X1	60	62	67	70	71	72
X2	22	25	24	20	15	14
Y	140	155	159	179	192	200

14) Using gradient descent algorithm with least square error as cost function, calculate the parameters for linear regression for 2 successive iterations. Assume the initial values of m=1.15 and c = -15 and learning rate as 0.1 [5] [CO4] [BTL3]

Sl.no	**Mothers' height**	**Daughters' height**
1.	58	60
2.	62	60
3.	60	58
4.	64	60
5.	67	70
6.	70	72

****

Course Outcome /Bloom's Taxonomy Level (BTL) Mark Distribution Table

CO	**Marks**	**BTL**	**Marks**
CO01	24	BTL 1	6
CO02		BTL 2	21
CO03		BTL 3	20
CO04	26	BTL 4	3
CO05		BTL 5	
		BTL 6	