Roll No.: _

Amrita Vishwa Vidyapeetham

Amrita School of Computing, Coimbatore

B.Tech. Degree Examinations – February 2023

Fifth Semester

Computer Science and Engineering

19CSE305 Machine Learning

Duration: Three hours | Maximum: 100 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. A. Illustrate the steps involved in K-Means clustering algorithm with a neat flow diagram. Also, elaborate on how the optimal value of K is determined. [5][CO4][BTL2]

   B. Apply K-Means clustering manually with K = 2 on a small example of six observations with two features given below. Use the Euclidean measure for distance calculation. Suppose that we initially take P1 and P2 as centroids of cluster 1 & 2, respectively. Report the cluster centroids and cluster assignments after each iteration. [10][CO5][BTL3]

Data Point	x	y
P1	1	4
P2	1	3
P3	6	2
P4	5	1
P5	0	4
P6	4	0

2. A. Write down the steps involved in dimensionality reduction using Principal Component Analysis (PCA). [5][CO2][BTL1]

   B. Perform PCA on the dataset listed below and list down the explained variance corresponding to the principal components. Derive the vector which the projection of the data matrix onto 1-D using the first principal component. [10][CO2][BTL3]

X	Y
-1.98	0.75
-0.45	-0.12
-0.33	-0.77
-0.82	-0.43
1.18	0.34
0.17	-0.77
0.01	0.68
-0.52	-0.27
-1.37	-0.68
-0.87	-0.05

3. Use Decision tree, to obtain the tree for classifying the given 14 samples. Identify the drug to be administered for the given test case (Patient ID: 15). [15][CO4][BTL3]

Patient ID	Age	Sex	BP	Cholesterol	Drug
p1	Young	F	High	Normal	Drug A
p2	Young	F	High	High	Drug A
p3	Middle-age	F	Hiigh	Normal	Drug B
p4	Senior	F	Normal	Normal	Drug B
p5	Senior	M	Low	Normal	Drug B
p6	Senior	M	Low	High	Drug A
p7	Middle-age	M	Low	High	Drug B
p8	Young	F	Normal	Normal	Drug A
p9	Young	M	Low	Normal	Drug B
p10	Senior	M	Normal	Normal	Drug B
p11	Young	M	Normal	High	Drug B
p12	Middle-age	F	Normal	High	Drug B
p13	Middle-age	M	High	Normal	Drug B
p14	Senior	F	Normal	High	Drug A
p15	Middle-age	F	Low	Normal	?

4. For the weight of the car engine the corresponding Carbon Dioxide (CO2) emissions are given below. Using linear regression predict the CO2 emission. Calculate the regression coefficients and write the equation. Show the values predicted by your model and sum of squared error of the predicted values. An engine manufacturer uses your model to predict CO2 emission for his engine with weight of 1235, what will be the value predicted? If the Government imposes a rule on Maximum CO2 emission to be 106, will the manufacturer's engine adhere to the rule? If CO2 emission increases by 50 % what will be the percentage of increase or decrease in weight? [15][CO2][BTL3]

Weight	929	1109	1150	1365	1119	1415	1490	1725	1705	1390
CO2	95	90	99	99	104	99	104	114	114	108

5. Verify whether the following Boolean function can be represented as a single logistic threshold unit? (i.e., a single unit from a neural network)? If yes, show the weights. If not, explain why not? [5][CO5][BTL2]

A	B	f(A,B)
1	1	0
0	0	0
1	0	1
0	1	0

6. A. Write a short note on Convolutional Neural Networks with suitable diagram. [3][CO3][BTL2]
   B. If you decide to use a neural network to solve problem to detect a particular disease. We have two choices: either to train a separate neural network for each of the diseases or to train a single neural network with one output neuron for each disease, but with a shared hidden layer. Which method do you prefer? Justify your answer [2][CO3][BTL2]

7. A. Briefly write how Singular Value Decomposition identifies the most information-carrying components in a data matrix. [2.5][CO3][BTL2]
   B. Given a dimension 'k' less than the original dimension of the data matrix 'X' (of dimension m x n), illustrate how can the dimensionality reduction be done to get the reduced matrix 'X_reduced', with a diagram illustrating the choice of rows and columns of the decomposition along with the dimensions. [2.5][CO3][BTL4]

8. Apply Hierarchical Divisive clustering over the dissimilarity matrix given below to create singleton clusters that are corresponding to the data points present in the dataset. Also, graphically illustrate the whole process of how the initial cluster is getting divided into singleton clusters in different steps. [5][CO5][BTL3]

	a	b	c	d	e	f
a	0	2	6	10	9	7
b	2	0	5	9	8	6
c	6	5	0	4	5	3
d	10	9	4	0	3	2
e	9	8	5	3	0	1
f	7	6	3	2	1	0

9. The count of fruits for 3 binary features is summarized in the following table. Calculate the probability P(long, sweet, yellow). Identify the fruit for the features (long, sweet, yellow) according to Naïve bayes Theorem. Calculate the class priors, likelihood in Naïve bayes formula. [5][CO4][BTL3]

Type	Long	Not Long	Sweet	Not Sweet	Yellow	Not Yellow	Total
Banana	400	100	350	150	450	50	500
Orange	0	300	150	150	300	0	300
Other	100	100	150	50	50	150	200
Total	500	500	650	350	800	200	1000

10. A child decides to buy candy based on the cost. Find the equation for applying logistic regression for the given tabulated data. [5][CO2][BTL3]

Cost of candy	Decision
8.00	Yes
10.00	No
12.00	No
9.50	Yes
15.00	No
9.00	Yes
6.00	Yes

11. A. What you mean by ensemble learning? [2.5][CO3][BTL2]
    B. Differentiate the Bagging and Boosting ensemble learning methods. [2.5][CO3][BTL2]

12. Given the confusion matrix. Calculate (i) TP, TN. FP, & FN, (ii) Accuracy, (iii) Precision, (iv) Recall, and (v) F1- Score. [5][CO4][BTL1]

		Actual		
Predicted		Apple	Orange	Mango
	Apple	7	8	9
	Orange	1	2	3
	Mango	3	2	1

****

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

CO	Marks	BTL	Marks
CO01		BTL 1	10
CO02	35	BTL 2	22.5
CO03	15	BTL 3	65
CO04	30	BTL 4	2.5
CO05	20	BTL 5	
		BTL 6	