BIRLA INSTITUTE OF TECHNOLOGY, MESRA, RANCHI
(END SEMESTER EXAMINATION)

CLASS: I M.Sc. **SEMESTER : VI**
BRANCH: QEDS **SESSION : SP/2025**

SUBJECT: ED317 STATISTICAL MACHINE LEARNING I

TIME: 3 Hours **FULL MARKS: 50**

INSTRUCTIONS:
1. The question paper contains 5 questions each of 10 marks and total 50 marks.
2. Attempt all questions.
3. The missing data, if any, may be assumed suitably.
***4*. Before attempting the question paper, be sure that you have got the correct question paper.**
5. Tables/Data hand book/Graph paper etc. to be supplied to the candidates in the examination hall.

			CO	BL
Q.1(a)	You and a few of your friends are suffering from allergic reactions and visited BIT Mesra Dispensary. The doctor specified that allergy is a direct consequence of certain foods you'll eat at the place where it is consumed, time by day, day of week and the amount spent on food. The summary of all the students is given on the table below.	[5]	1	1,2

Number	Restaurant	Meal	Day	Cost	Reaction
1	Sam's	breakfast	Friday	cheap	yes
2	Hilton	lunch	Friday	expensive	no
3	Sam's	lunch	Saturday	cheap	yes
4	Denny's	breakfast	Sunday	cheap	no
5	Sam's	breakfast	Sunday	expensive	no

Describe the hypothesis that fits this data using Find-S algorithm. Show traces.

Q.1(b) Consider the following set of attributes: [5] 1 2,3

Study: *Intense, Moderate, None*, **Difficulty**: Easy, Hard, **Sleepy**: Very, Somewhat, **Attendance**: Frequent, Rare, **Hungry**: Yes, No, **Thirsty**: Yes, No, **PassTest**: Yes, No

Suppose we have the data set:

Example	Study	Difficulty	Sleepy	Attendance	Hungry	Thirsty	Earn-A
1	Intense	Normal	Extremely	Frequent	No	No	**Yes**
2	Intense	Normal	Slightly	Frequent	No	No	**Yes**
3	None	High	Slightly	Frequent	No	Yes	**No**
4	Intense	Normal	Slightly	Frequent	Yes	Yes	**Yes**

Consider the space *H* of conjunctive hypotheses, which, for each attribute, either:

- indicates by a "?" that any value is acceptable for this attribute,
- specifies a single required value (e.g., *Normal*) for this attribute, or
- indicates by a "▫" that no value is acceptable.

Let a version space (a subset of consistent hypotheses in *H*) be represented by an *S* set (specific boundary, at the top) and a *G* set (general boundary, at the bottom). Suppose the 4 training examples above are presented in order.
Draw a diagram showing the evolution of the version space for concept ***Earn-A*** given the training examples, by writing down *S1*, *G1*, *S2*, *G2*, *S3*, *G3*, *S4*, and *G4*. If the *G* set does not change given a new example, just write *Gi+1* = *Gi* next to the drawing of *Gi* (similarly for *S*).

PTO

Q.2(a) Consider the following training data set of (seven) points X's in a plane and their binary class label y's: [5] 2 2,3

X	*(1, 0)*	*(0, 1)*	*(0, -1)*	*(-1, 0)*	*(0, 2)*	*(0, -2)*	*(-2, 0)*
y	*-1*	*-1*	*-1*	*+1*	*+1*	*+1*	*+1*

We perform the following non-linear transform of the input vector X = (x_1, x_2) to obtain the transformed feature vector Z = (z1, z2) = (ϕ1(X), ϕ2(X)), with ϕ1(X) = x_2^2 -2x_1 +3, ϕ2(X) = x_1^2 -2x_2 -3. Write the equation of the optimal separating hyperplane in transformed space Z. Explain your answer.

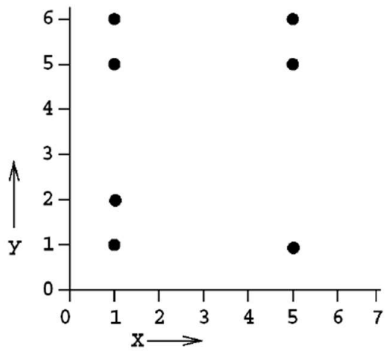
Q.2(b) Imagine that you are given the following set of training examples. Each feature can take on one of three nominal values: a, b, or c. [5] 2 3

F1	*F2*	*F3*	*Category*
a	c	a	+
c	a	c	+
a	a	c	-
b	c	a	-
c	c	b	-

i) Design a Naive Bayes system classifying the following test example? F1 = a F2 = c F3 = b

ii) Describe how a 3-nearest-neighbor algorithm would classify Part a's test example.

Q.3(a) The following picture shows a dataset with one real-valued input x and one real-valued output y. There are seven training points. [5] 3 4



Suppose you are training using kernel regression using some unspecified kernel function. The only thing you know about the kernel function is that it is a monotonically decreasing function of distance that decays to zero at a distance of 3 units (and is strictly greater than zero at a distance of less than 3 units). Evaluate the following with parameters specified.

a. What is the predicted value of y when x = 1?
b. What is the predicted value of y when x = 3?
c. What is the predicted value of y when x = 4?
d. What is the predicted value of y when x = 7?

Q.3(b) For the given data points, Class 1 has 5 samples c1= [(1,2), (2,3), (3,3), (4,5), (5,5)]; Class 2 has 6 samples c2= [(1,0), (2,1), (3,1), (3,2), (5,3), (6,5)]. Describe which algorithm will be best justified to create decision boundaries between these two classes. Justify your choice of the algorithm. [5] 3 3,4

Q.4(a) At the beginning of an exam, you try to predict whether each problem is easy or difficult (D = + if it is difficult and – if it is easy). Let us assume that you use two observable problem attributes (predicates): [5] 4 4,4, 6

· The text length L of the problem (L = 1 if it is long, 0 otherwise)

· The amount M of math in the text (M = 1 if there is a lot of math, 0 otherwise)

For training data, assume that you have examined 12 previous problems from the homeworks, and have collected the following data:

L	M	D	#
0	0	-	4
0	0	+	1
0	1	-	0
0	1	+	3
1	0	-	1
1	0	+	2
1	1	-	1
1	1	+	0

The first line of this table reads as follows: 4 problems for which L = 0 and M = 0 were not difficult (D = -). The second line says: 1 problem for which L = 0 and M = 0 was difficult (D = +). etc..... Note that you observed no problem for which L = 0 and M = 1, or L = 1 and M = 1.

Based on this training data, you want to compute a representation of a difficult problem (D) in the form of a decision tree using the two binary attributes L and M. Construct the best decision tree you can for the training data.

Q.4(b) Assume the following dataset is given: (2,2), (4,4), (5,5), (6,6),(9,9) (0,4), (4,0) . K-Means is run with k=3 to cluster the dataset. Moreover, Manhattan distance is used as the distance function to compute distances between centroids and objects in the dataset. Moreover, K-Mean's initial clusters C1, C2, and C3 are as follows: [5] 4 4

C1: {(2,2), (4,4), (6,6)}

C2: {(0,4), (4,0)}

C3: {(5,5), (9,9)}

Now K-means is run for a single iteration; what are the new clusters and what are their centroids?

Q.5 Consider a neural network with two inputs, two hidden layers and two outputs. The input received by I_1 and I_2 are 0.1 and 0.5 respectively. The weights w_1 from I_1 to H_1 is 0.1, w_2 from I_1 to H_2 is 0.2, w_3 from I_2 to H_1 is 0.3 and w_4 from I_2 to H_2 is 0.4. The weights w_5 from H_1 to O_1 is 0.5, w_6 from H_2 to O_1 is 0.6, w_7 from H_1 to O_2 is 0.7 and w_8 from H_2 to O_2 is 0.8. The bias from Input to Hidden layer is 0.25 and from Hidden to Output layer is 0.35. The expected outputs from O_1 and O_2 are 0.05 and 0.95 respectively. Sigmoid function (Logistic function) is the activation function. Perform the backpropagation algorithms for any of the weights of your choice up to six decimal points precision. The learning rate is 0.6 [10] 5 4,5, 6

:::::28/04/2024 M:::::