

**BIRLA INSTITUTE OF TECHNOLOGY, MESRA, RANCHI
(MID SEMESTER EXAMINATION)**

**CLASS: BTech
BRANCH: Food Engineering and Technology**

**SEMESTER : V
SESSION : MO/2025**

SUBJECT: FE325 STATISTICAL MACHINE LEARNING II

TIME: 02 Hours

FULL MARKS: 25

INSTRUCTIONS:

1. The question paper contains 5 questions each of 5 marks and total 25 marks.
 2. Attempt all questions.
 3. The missing data, if any, may be assumed suitably.
 4. Tables/Data handbook/Graph paper etc., if applicable, will be supplied to the candidates
-

- | | | CO | BL | | | | | | | | | | | | | | | | | | | | | | | | |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------|-------|----------|----------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|--|--|--|
| Q.1 Given two objects represented by the tuples (22, 1, 42, 10) and (20, 0, 36, 8): | [5] | 1 | 1,2 | | | | | | | | | | | | | | | | | | | | | | | | |
| i. Compute the Euclidean distance between the two objects. | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| ii. Compute the Manhattan distance between the two objects. | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| iii. Compute the Minkowski distance between the two objects, using $h = 3$. | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| iv. Compute the supremum distance between the two objects. | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Q.2 Let $X_1 = (1, -1, -1)$, $y_1 = -1$, $X_2 = (-3, 1, 1)$, $y_2 = 1$, $X_3 = (-3, 1, -1)$, $y_3 = -1$, $X_4 = (1, 2, 1)$, $y_4 = -1$, and $X_5 = (-1, -1, 2)$, $y_5 = 1$, be five binary labeled training examples. These points are linearly separable. Derive the optimum margin classifier (support vectors, weights and threshold value) and the margin. | [5] | 1 | 3,4 | | | | | | | | | | | | | | | | | | | | | | | | |
| Q.3 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 C_1 , C_2 , and C_3 are as follows:
$C_1: \{(2,2), (4,4), (6,6)\}$
$C_2: \{(0,4), (4,0)\}$
$C_3: \{(5,5), (9,9)\}$
Now K-means is run for a single iteration; what are the new clusters and what are their centroids? | [5] | 1 | 2,3 | | | | | | | | | | | | | | | | | | | | | | | | |
| Q.4(a) After winning a race, an Olympic runner is tested for the presence of steroids. The test comes up positive, and the athlete is accused of doping. Suppose, it is known that 5% of all victorious Olympic runners do use performance enhancing drugs. For this particular test, the probability of a positive finding given that drugs are used is 95%. The probability of a false positive is 2%. What is the probability that the athlete did in fact use steroids, given the positive outcome of the test. | [2] | 1 | 2,3 | | | | | | | | | | | | | | | | | | | | | | | | |
| Q.4(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. | [3] | 1 | 2,3 | | | | | | | | | | | | | | | | | | | | | | | | |
| <table style="margin-left: auto; margin-right: auto; border-collapse: collapse;"> <thead> <tr> <th style="border-bottom: 1px solid black;">F_1</th> <th style="border-bottom: 1px solid black;">F_2</th> <th style="border-bottom: 1px solid black;">F_3</th> <th style="border-bottom: 1px solid black;">Category</th> </tr> </thead> <tbody> <tr><td style="text-align: center;">a</td><td style="text-align: center;">c</td><td style="text-align: center;">a</td><td style="text-align: center;">+</td></tr> <tr><td style="text-align: center;">c</td><td style="text-align: center;">a</td><td style="text-align: center;">c</td><td style="text-align: center;">+</td></tr> <tr><td style="text-align: center;">a</td><td style="text-align: center;">a</td><td style="text-align: center;">c</td><td style="text-align: center;">-</td></tr> <tr><td style="text-align: center;">b</td><td style="text-align: center;">c</td><td style="text-align: center;">a</td><td style="text-align: center;">-</td></tr> <tr><td style="text-align: center;">c</td><td style="text-align: center;">c</td><td style="text-align: center;">b</td><td style="text-align: center;">-</td></tr> </tbody> </table> | F_1 | F_2 | F_3 | Category | a | c | a | + | c | a | c | + | a | a | c | - | b | c | a | - | c | c | b | - | | | |
| F_1 | F_2 | F_3 | Category | | | | | | | | | | | | | | | | | | | | | | | | |
| a | c | a | + | | | | | | | | | | | | | | | | | | | | | | | | |
| c | a | c | + | | | | | | | | | | | | | | | | | | | | | | | | |
| a | a | c | - | | | | | | | | | | | | | | | | | | | | | | | | |
| b | c | a | - | | | | | | | | | | | | | | | | | | | | | | | | |
| c | c | b | - | | | | | | | | | | | | | | | | | | | | | | | | |
| a) How would a Naive Bayes system classify the following test example?
$F_1 = a$ $F_2 = c$ $F_3 = b$ | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Q.5(a) Discuss the type of clustering algorithms with appropriate examples. | [2] | 2 | 1,2 | | | | | | | | | | | | | | | | | | | | | | | | |
| Q.5(b) Given the dataset $\{A, B, C, D, E\}$ and the distance matrix in Table. | [3] | 2 | 2,3 | | | | | | | | | | | | | | | | | | | | | | | | |

0	1	2	2	3
1	0	2	4	3
2	2	0	1	5
2	4	1	0	3
3	3	5	3	0

Apply divisive analysis clustering algorithm (DIANA) to form clusters.