Mid-Semester Examination
School of Computer Engineering
KIIT University, Bhubaneswar-24

Time: 2hrs   *Full Mark: 25*

(Group A is compulsory, Answer any 4 from Group B)

GROUP --- A

[1X5]

1. (a) Define data generalization and data summarization.
   (b) Discuss concepts of support and confidence.
   (c) Justify classification is a supervised learning or unsupervised learning.
   (d) What is discrete and continuous data in the data mining world.
   (e) What is the difference between data matrix and dissimilarity matrix.

GROUP --- B

[5X4]

2. Discuss data mining as a step in the process of knowledge discovery, with suitable figure.
3. What is meant by machine learning?. Discuss different types of learning techniques.
4.

Name	Gender	Feaver	Cough	Test-1	Test-2	Test - 3	Test- 4
Jack	M	Y	N	P	N	N	N
Jim	M	Y	Y	N	N	N	N
Mary	F	Y	N	P	N	P	N

   By considering the above information, calculate the distance measure for symmetric and asymmetric binary variables between each pair of the three patients.

5. Given two objects represented by the tuples (22, 1, 42, 10) and (20, 0, 36, 8):
   (a) Compute the Euclidean distance between two objects.
   (b) Compute Manhattan distance.
   (c) Compute Minkowski distance between two objects, using q = 3.
   (d) Compute the supremum distance between two objects.

6. Find out the frequent item set using Apriori algorithm, by considering the following transactional table (consider min_Sup = 2)

TID	Items
T100	M, O, N, K, E, Y
T200	D, O, N, K, E, Y
T300	M, A, K, E
T400	M, U, C, K, Y
T500	C, O, O, K, I, E

---GOOD LUCK --