MANIPAL INSTITUTE OF TECHNOLOGY
MANIPAL
A Constituent Institution of Manipal University
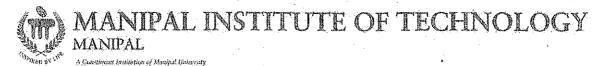
VI SEMESTER B.TECH.

END SEMESTER EXAMINATIONS, APRIL/MAY 2017

SUBJECT: OPEN ELECTIVE-II MACHINE LEARNING [ICT 3285]
REVISED CREDIT SYSTEM
(03/05/2017)

Time: 3 Hours   MAX. MARKS: 50

Instructions to Candidates:

❖ Answer ALL questions.

❖ Missing data, if any, may be suitable assumed.

1A	Describe the following: (i) n-fold cross validation. (ii) Logistic Regression. (iii) Descriptive and Predictive Modeling.	5
1B	A store has been selling 200 DVD burners a week at $350 each. A market survey indicates that for each $10 rebate offered to buyers, the number of units sold will increase by 20 a week. Find the demand function and the revenue function. How large a rebate should the store offer to maximize its revenue?	3
1C	Describe in brief any four major steps of data preprocessing for a supervised learning task.	2

2A	Describe the Backpropagation Algorithm for supervised learning.	5
2B	Consider the data set in Table Q.2B.	3

Table Q.2B

	A1	A2
x1	1.5	1.7
x2	2	1.9
x3	1.6	1.8
x4	1.2	1.5
x5	1.5	1.0

Given a new data point, $x = (1.4, 1.6)$ as a query, rank the database points based on similarity with the query using Euclidean distance, and cosine similarity measures.

2C	In each of the following sets of variables, identify which of the variables can be regarded as a response variable and which can be used as predictors? (i) Number of cylinders and gasoline consumption of cars. (ii) SAT scores, grade point average, and college admission.	2

3A	Solve the given problem using AdaBoost Algorithm: There are five data points on a X-Y plane: (2, 2), (2, 6), (4, 4), (6, 2) and (6, 6).	5

Assume that we have four classifiers: x<2; x<6; x>2; x>4. Also assume that all the points lying to the left of the axis lines belong to positive class '+' and all the points lying to the right of the axis lines belong to negative class '-'. Determine the ensemble of classifiers that classify the given data points.

3B   Give any three real time examples of machine learning in the domain of Data Mining.   3

3C   How does the splitting criterion aid in generating decision trees?   2

4A   Describe the following concepts in reference to reinforcement learning: 5
    (i) Markov Decision Process.
    (ii) Agent – Environment Interface.
    (iii) Goals and Rewards.

4B   Data from a sample of 10 pharmacies are used to examine the relation between prescription sales volume and the percentage of prescription ingredients purchased directly from the supplier. The sample data is given in Table Q.4B.   3

<p align="center">Table Q.4B</p>

Pharmacy	1	2	3	4	5	6	7	8	9	10
Sales volume, y (in $1,000)	25	55	50	75	110	138	90	60	10	100
% of ingredients purchased directly, x	10	18	25	40	50	63	42	30	5	55

(i) Find the least squares estimates for the regression line:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 * x$$

(ii) Predict the sales volume for a pharmacy that purchases 15% of its prescription ingredients directly from the supplier.

4C   Explain the two types of clustering techniques mentioning at least one algorithm used in each category.   2

5A   Suppose that the data mining task is to cluster points (with (x, y) representing location) into three clusters, where the points are   5
    A1(2,10), A2(2,5), A3(8,4), B1(5,8), B2(7,5), B3(6,4), C1(1,2), C2(4,9)

The distance function to be used is Squared Euclidean distance. Suppose initially we assign A1, B1, and C1 as the center of each cluster, respectively. Use the k-means algorithm to show only:
    (i) The three cluster centers after the first round of execution.
    (ii) The final three clusters.

5B   A small manufacturing company made 23 radios and four of them were found to be defective. Two were selected at random.   3
    (i) If random sampling was done without replacement, find the probability of the second one being defective given the first was defective.
    (ii) If random sampling was done with replacement, find the probability of the second one being defective given the first was defective.

5C   Describe in brief Agglomerative and Divisive hierarchical clustering methods.   2