Nirma University

Institute of Technology

Semester End Examination (IR/RPR), December - 2023

B. Tech. in Computer Science and Engineering, Semester-V

2CS501-O Machine Learning

Roll / Exam No.		Supervisor's initial with date	

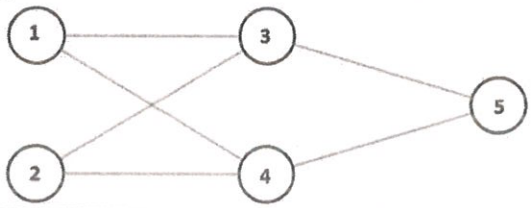
Time: 3 Hours | Max. Marks: 100

Instructions:
1. Attempt all questions.
2. Figures to the right indicate full marks.
3. Assume suitable assumptions and specify them.
4. Section-wise separate answer book to be used.

Section – I

Q-1. Do as directed. **[16]**

A CO3,BL4 Consider the following feed-forward neural network. [08]



Input		Weight					
X1	X2	W13	W14	W23	W24	W35	W45
0.35	0.9	0.1	0.4	0.8	0.6	0.3	0.9

Assume that the neurons have a sigmoid activation function, perform a forward pass and a backward pass on the network. Assume that the actual output y is 0.5 and the learning rate is 1.

B CO3,BL1 What is the need of pre-processing of the given data? How it improves accuracy, efficiency and scalability of classification/Regression. Give a suitable example. [08]

OR

B CO3,BL4 Answer the following question based on a self-organising map neural network. Let the training set T consist of six 3-dimensional vectors, i1 to i6. T={i1=(1.1,1.7,1.8),i2=(0,0,0),i3=(0,0.5,1.5),i4=(1,0,0),i5=(0.5,0.5,0.5),i6=(1,1,1). There are three Neurons in the output layer: A, B, C. The Connection strengths (weights) are initially chosen randomly, and are given by the weight matrix [08]

$$W(0) = \begin{pmatrix} w1: & 0.2 & 0.7 & 0.3 \\ w2: & 0.1 & 0.1 & 0.9 \\ w3: & 1 & 1 & 1 \end{pmatrix}$$

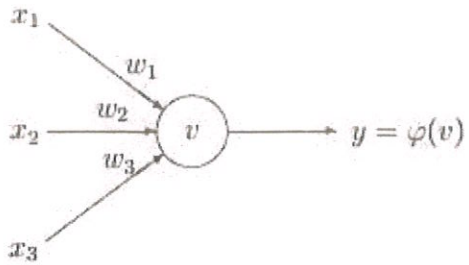
For D(t)=1, learning rate η = 0.5. For each sample presented show the winner node till t=6. Also show the weight matrix for each input. (Show all the calculations using squared Euclidean distance)

Q-2. Answer the following. **[14]**

A CO2,BL1 What is Normal Equation? Where will it be useful? [05]

OR

A CO2,BL4 Below is a diagram of a single artificial neuron (unit): [05]



The node has three inputs x = ($x1$, $x2$, $x3$) that receive only binary signals (0 or 1). How many different input patterns can this node receive? What if the node had four inputs? Give a formula that computes the number of binary input patterns for a given number of inputs.

B CO2,BL4 Suppose we are given the dataset in the table below for a classification problem. There are two binary inputs attributes X1 and X2, and three classes Y belongs to {1,2,3}. Is decision tree regressor applicable? If yes, do the following: [09]

1. Calculate the information gain for both X1, X2
2. Which attribute is used as first spilt at the root node of the tree? (Depth of the tree =2)
3. Classify at new tuple {1,0}

X1	X2	Y
0	0	1
0	0	1
0	0	2
0	1	3
1	1	2
1	1	3

Q-3. Do as directed. **[20]**

A CO3,BL4 Consider the data given in the following table. [20]

X Values	Y Values
60	3.1
61	3.6
62	3.8
63	4
65	4.1

Fit a first-order regression model to the data (y = $\beta 0$ + $\beta 1$x). Estimate parameters of the model ($\beta 0$ and $\beta 1$) through one epoch (one complete pass through the training set) of batch gradient descent and stochastic gradient descent assuming learning rate = 0.01. Assume the initial value of $\beta 0$ and $\beta 1$ to be 1 and 4, respectively. The loss function is (1/2)*(mean squared error)

Section – II

Q-4. Do as directed. **[18]**

A CO1,BL1 The model you have trained has a low bias and high variance. How would you deal with it? Which cross-validation technique would you suggest for a time-series dataset and why? [06]

OR

A CO1,BL4 Supervised learning is a learning algorithm paradigm where the model is given a labelled data and the outcome is a learned model. A user is training NB and [06]

ANN using the same training data. The training data is wine data with 150 instances and three class and hence first 50 instances belongs to class-1, followed by 50 instances of Class-II and finally 50 instances of Class-III. The user trains both the model with shuffling and without shuffling the sequences of instances. What will be the difference in the outcome in both the models and with both the options? Explain with proper reasons.

B CO1,BL3 Consider the following positive data points {(2,0), (3,1), (3,-1), (4,0)} and negative data points {(2,0), (0,1), (0,-1), (-2,0)}. Use linear SVM function to find hyperplane equation. [06]

C CO1,BL2 What is Feature Importance in Machine Learning, and how do you determine it? Is it true that we need to scale our feature values when they vary greatly? [06]

Q-5. Answer the following. **[16]**

A CO3,BL4 Use the k-means algorithm and Euclidean distance to cluster the following 8 examples into 3 clusters: [10]

A1=(2,10), A2=(2,5), A3=(8,4), A4=(5,8), A5=(7,5), A6=(6,4), A7=(1,2), A8=(4,9).

Suppose that the initial seeds (centers of each cluster) are A1, A4 and A7. Run the k-means algorithm for 1 epoch only. At the end of this epoch show:

1. The new clusters (i.e. the examples belonging to each cluster)
2. The centers of the new clusters
3. Draw a 10 by 10 space with all the 8 points and show the clusters after the first epoch and the new centroids.
4. How many more iterations are needed to converge? Draw the result for each epoch.

B CO1,BL1 In semi-supervised learning, learning from labelled and unlabelled data which techniques are used. Explain them in detail. [06]

OR

B CO1,BL2 Differentiate between supervised, unsupervised, semi-supervised and reinforcement learning in detail. [06]

Q-6. Do as directed. **[16]**

A CO2,BL4 Consider the training data shown in the below table. Predict the class of the following tuple X using naïve-Bayes classification using Gaussian Distribution. Show all the computation steps. [08]

X = {Height=6, Weight=130, Foot Size=?}.

Gender	Height (feet)	Weight (lbs)	Foot Size (inches)
Male	6	180	12
Male	5.92	190	11
Male	5.58	170	12
Male	5.92	165	10
Female	5	100	6
Female	5.5	150	8
Female	5.42	130	7
Female	5.75	150	9

B CO2,BL4 Assume 7 one-dimensional input patterns {0.0, 0.17, 0.33, 0.51, 0.67, 0.83, 1.0}. Assume the first three patterns belong to class-0 (with desired output 0) and the remaining patterns belong to class-1 (with desired output 1). Design a perceptron to classify these patterns. Use the perceptron learning rule. Assume learning rate=0.1 and initial weight and bias to be (-0.4) and (-0.1), respectively. Show computation for one epoch. [08]