Q. Code: 838841

M.E. / M.TECH. DEGREE EXAMINATIONS, DEC 2020 (Held during April, 2021)

First Semester

CP18105-MACHINE LEARNING TECHNIQUES

(Common to CP & NW)

(Regulation 2018)

Time: Three hours Maximum : 80 Marks

Answer **ALL** questions

PART A - (8 X 2 = 16 marks)

1. Concept learning inferred a ______ valued function from training examples of its input and output.

   A) Decimal B) Hexadecimal C) Boolean D) All of the above

2. In k-NN algorithm, given a set of training examples and the value of k < size of training set (n), the algorithm predicts the class of a test example to be the

   A) Most frequent class among the classes of k closest training examples.

   B) Least frequent class among the classes of k closest training examples.

   C) Class of the closest point.

   D) Most frequent class among the classes of the k farthest training examples.

3. For which of the following cases feature selection may be used?

   A) If large number of completely irrelevant features are present in the data.

   B) For better interpretability of data.

   C) Both A and B

   D) None of A or B

4. The probability density function of a Markov process is

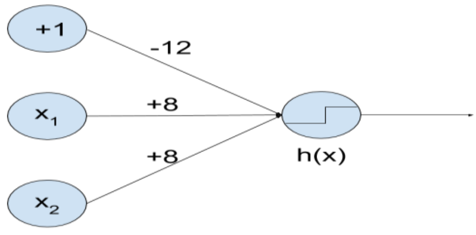
   A) p(x1,x2,x3.......xn) = p(x1)p(x2/x1)p(x3/x2).......p(xn/xn-1)

   B) p(x1,x2,x3.......xn) = p(x1)p(x1/x2)p(x2/x3).......p(xn-1/xn)

   C) p(x1,x2,x3......xn) = p(x1)p(x2)p(x3).......p(xn)

   D) p(x1,x2,x3......xn) = p(x1)p(x2 *x1)p(x3*x2)........p(xn*xn-1)

5. You are given the following neural networks which take two binary valued inputs x1,x2∈{0,1} and the activation function is the threshold function(h(x) = 1 if x > 0; 0 otherwise). What logical function does it compute?



6. Find the derivative of sigmoid function.

7. Given a list of 14 examples including 9 positive and 5 negative examples. Find the entropy of the dataset with respect to this classification.

8. Comment on the limitation of principle component analysis.

PART B - (4 X16 = 64 marks)

09. (a) (i) Consider the following set of 2-dimensional vectors from two linearly separable classes **(16)**

C_1		**C_2**	
x_1	**x_2**	**x_1**	**x_2**
0	1	0	-1
1	2	-2	-1
2	0	-1	-2
3	2	-2	-3

where $\mathbf{y}=[c_1\ c_2]^t$. Assume a learning rate , $\eta = 0.5$. Determine the weights after two iterations using perceptron learning by considering $\mathbf{w}=[-1,-1]^t$ and $w_0 = -2$ as the initial weights.

(OR)

(b) (i) How Candidate-Elimination algorithm is used to compute a concept description that is consistent with all the positive examples and none of the negative examples. Learn the concept of "Japanese Economy Car" using the same. The features are Country of Origin, Manufacturer, Color, Decade, Type **(16)**

Origin	**Manufacturer**	**Color**	**Decade**	**Type**	**Example Type**
Japan	Honda	Blue	1980	Economy	Positive
Japan	Toyota	Green	1970	Sports	Negative
Japan	Toyota	Blue	1990	Economy	Positive
USA	Chrysler	Red	1980	Economy	Negative
Japan	Honda	White	1980	Economy	Positive

10. (a) (i) Consider a 3-layer feed-forward neural network with 5 input neurons, 4 hidden neurons and 2 output neurons contains how many weights? (Include biases.) Show the architecture and the notations used for the back propagation algorithm derivation. Compute the gradient from hidden to output layer. Assume sigmoid activation function is used in hidden and output layers. Linear function is used in input layer. **(16)**

(OR)

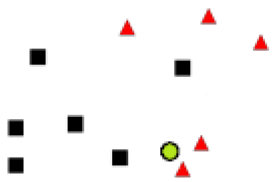
(b) (i) Derive the primal and dual forms of the Lagrangian objective function for the constrained optimization problem in constructing the maximal-margin hyper-plane for linearly non-separable data using slack variables. Derive the constraints on the Lagrange coefficients. **(16)**

11. (a) (i) It is necessary to build a Gaussian mixture model (GMM) for a univariate multi-modal distribution that represents the data of a class. Define the responsibility term, and derive the reestimation formulae for the mixture coefficients, means and variances of the M components in the GMM. Suggest a method for initialization of the parameters of the GMM. **(16)**

(OR)

(b) (i) Show how the class conditional densities are approximated by the K-nearest neighbor classifier. **(4)**

(ii) Consider the following training set consisting of 6 points in class 1 (the squares) and 5 points in class 2 (the triangles). Each point has two features (x, y) corresponding to its 2D coordinates. A new point (the circle) is to be classified using k − NN and Euclidean distance. **(12)**



(a) When k = 3, how should the new point be classified? Which points are its closest neighbours?

(b) When k = 5, how should the new point be classified? Which points are its closest neighbours?

(c) When k = 1, draw the approximate decision boundary that separates the two classes.

12. (a) (i) Consider the HMM representing a 2-coin tossing experiment. The model parameters are given below: **(16)**

$\pi 1 = \pi 2 = 1/2$

$a_{ij} = 1/2$, $1 \leq i,j \leq 2$

$b_1(H) = 1/2$ and $b_2(H)=3/4$.For the observation sequence O=HHT, compute $P(O \mid \lambda)$ using the forward method.

(OR)

(b) (ii) Explain the Naïve Bayesian classifier in detail. Consider the given dataset in the table below: **(16)**

Cook	**Mood**	**Cuisine**	**Tasty**
Sita	Bad	Indian	Yes
Sita	Good	Continental	Yes
Asha	Bad	Indian	No
Asha	Good	Indian	Yes
Usha	Bad	Indian	Yes
Usha	Bad	Continental	No
Asha	Bad	Continental	No
Asha	Good	Continental	Yes
Usha	Good	Indian	Yes
Usha	Good	Continental	No

Consider a test pattern X= (Cook='Usha', Mood='Bad', Cuisine='Continental')).

Find the class label Tasty(yes/no) using Naïve Bayesian Classifier