

Total No. of printed pages = 8

CS 1318 E 033

BINA CHOWDHURY CENTRAL LIBRARY
(GIMT & GIPS)

Azara, Hatkhowapara
Guwahati - 781017

Roll No. of candidate

--	--	--	--	--	--	--	--	--	--

2019

B.Tech. 8th Semester End-Term Examination
INTRODUCTION TO MACHINE LEARNING -
(Elective III)

Full Marks – 100

Time – Three hours

The figures in the margin indicate full marks
for the questions.

Answer question No. 1 and any *six* from the rest.

1. Answer the following questions : (10 × 1 = 10)

- (i) The effectiveness of an SVM depends upon
- (a) Selection of Kernel
 - (b) Kernel Parameters
 - (c) Soft Margin Parameter
 - (d) All of the above
- (ii) Which of the following gives non-linearity to a neural network?
- (a) Stochastic Gradient Descent
 - (b) Rectified Linear Unit
 - (c) Convolution function
 - (d) None of the above

[Turn over

- (iii) Why is the XOR problem exceptionally interesting to neural network researchers?
- (a) Because it can be expressed in a way that allows you to use a neural network
 - (b) Because it is complex binary operation that cannot be solved using neural networks
 - (c) Because it can be solved by a single layer perceptron
 - (d) Because it is the simplest linearly inseparable problem that exists
- (iv) In an Unsupervised learning
- (a) Specific output values are given
 - (b) Specific output values are not given
 - (c) No specific Inputs are given
 - (d) Both inputs and outputs are given
- (v) Automated vehicle is an example of
- (a) Supervised learning
 - (b) Unsupervised learning
 - (c) Active learning
 - (d) Reinforcement learning
- (vi) A perceptron adds up all the weighted inputs it receives, and if it exceeds a certain value, it outputs a 1, otherwise it just outputs a 0.
- (a) True
 - (b) False
 - (c) Sometimes — it can also output intermediate values as well
 - (d) Can't say

(vii) The cost parameter in the SVM means

- (a) The number of cross-validations to be made
- (b) The kernel to be used
- (c) The tradeoff between misclassification and simplicity of the model
- (d) None of the above

(viii) Which of the following statement is true about k-MN algorithm?

- (1) k-NN performs much better if all of the data have the same scale
 - (2) k-NN works well with a small number of input variables (p), but struggles when the number of inputs is very large
 - (3) k-NN makes no assumptions about the functional form of the problem being solved
- (a) (1) and (2) (b) (1) and (3)
(c) Only (1) (d) All of the above

(ix) Which of the following machine learning algorithm can be used for imputing missing values of both categorical and continuous variables?

- (a) K-NN
- (b) Linear Regression
- (c) Logistic Regression

(x) k-MN algorithm does more computation on test time rather than train time.

- (a) True (b) False

Answer any six of the following questions :

(6 × 15 = 90)

2. (a) What is back propagation? Explain how the weights are propagated back in ANN.
 - (b) What are the popular activation functions? Write equation of any one of them.
 - (c) Prove that the multilayer neural networks are linear without activation function.
 - (d) Explain with example the firing rules of neural networks? (5 + 3 + 4 + 3)
3. (a) When should you use classification over regression?
 - (b) What's the "kernel trick" and how is it useful?
 - (c) What is margin in context to support vector machine?
 - (d) Why linear SVM works well for high dimensional problems? (3 + 4 + 3 + 5)
4. (a) Which is more important to you— model accuracy. or model performance? Justify.
 - (b) Why is naive bayes naive?
 - (c) Write the algorithm of K-Nearest Neighbor for regression.

- (d) Using K-Nearest Neighbor, predict the class that $X = \langle 3, 8 \rangle$ will take from the following table :

X1 :	6	5	4	10	9	10
X2 :	1	5	5	7	9	7
Y :	1	1	1	0	0	0

5. (a) Given are the points $A = (1, 2)$, $B = (2, 2)$, $C = (2, 1)$, $D = (-1, 4)$, $E = (-2, -1)$, $F = (-1, -1)$. Starting from initial clusters Cluster 1 = {A} which contains only the point A and Cluster 2 = {D} which contains only the point D, run the K-means clustering algorithm and report the final clusters. Use L1 distance as the distance between points which is given by

$$d((x_1, y_1), (x_2, y_2)) = |x_1 - x_2| + |y_1 - y_2|$$

- (b) What are the different types of hierarchical clustering algorithms? Explain with example.
- (c) State the advantages and disadvantages of hierarchical clustering algorithms. (8 + 4 + 3)
6. (a) What is linear regression? Write the simplest form of the regression equation with one dependent and one independent variable.
- (b) One end A of an elastic string was attached to a horizontal bar and a mass, m grams, was attached to the other end B. The mass was suspended freely and allowed to settle vertically below A. The length AB, l mm, was recorded, for various masses as follows :

m :	100	200	300	400	500
l :	228	236	256	278	285

- (i) Calculate the least squares line of regression of l on m and plot this line on graph along with the above points.
- (ii) Estimate the length of the string when a mass of 360 grams is attached at B.
- (c) Write a function in any language to calculate the Root Mean Squared Error. (4 + 8 + 3)
7. (a) How do you handle missing or corrupted data in a dataset? Give an example using any language of your choice.
- (b) State the different properties of distance.
- (c) What is object-by-object representation of data? Give an example.
- (d) The following is information about three patients. Find the pair most similar.

Name	fever	cough	test-1	test-2	test-3	test-4
a	Y	N	P	N	N	N
b	Y	N	P	N	P	N
c	Y	Y	N	N	N	N

(4 + 2 + 2 + 7)

8. (a) What is Laplacian correction? Give an example.
- (b) How is information gain (or entropy) of a dataset calculated?

- (c) Using Naïve Bayesian Classification find the class that x takes from the following dataset :

Gender	Age	BP	Drug
M	young	Normal	A
F	senior	Normal	B
M	midAge	High	A
M	midAge	Low	B
F	midAge	High	A
M	young	Normal	A
F	senior	Normal	B
M	midAge	Low	B
M	senior	Normal	B
F	midAge	Normal	A
F	young	Low	B
M	senior	High	A

$x = \langle M, \text{young}, \text{High} \rangle$

(3 + 4 + 8)

9. (a) Define the following terms :

- (i) \mathcal{E} -neighborhood
- (ii) Core object
- (iii) Directly density-reachable

- (b) Give the advantages and disadvantages of DBSCAN algorithm.
- (c) Given the following matrix of distances, cluster the cities using single-linkage hierarchical clustering.

	B	F	M	N	R	T
B	0	662	877	255	412	996
F	662	0	295	468	268	400
M	877	295	0	754	564	138
N	255	468	754	0	219	869
R	412	268	564	219	0	669
T	996	400	138	869	669	0

(6 + 4 + 5)