

[This question paper contains 12 printed pages.]

Your Roll No.....

Sr. No. of Question Paper : 3143

H

Unique Paper Code : 32347607

Name of the Paper : Machine Learning

Name of the Course : B.Sc. (H) Computer Science

**ADMISSIONS OF 2019,
2020 & 2021**

Semester : VI

Duration : 3 Hours

Maximum Marks : 75

Instructions for Candidates

1. Write your Roll No. on the top immediately on receipt of this question paper.
2. **Section A** is compulsory.
3. Attempt any **4** questions from **Section B**.
4. Use of scientific calculator is allowed.

कालिन्दी महाविद्यालय पुस्तकालय

KALINDI COLLEGE LIBRARY

P.T.O.

Section A
(Compulsory)

1. (a) Consider a scenario where 6000 patients are tested for Covid positive. Out of which 5000 are actually Covid negative and 1000 are actually Covid positive. For covid positive patients the test however gave positive indication for 700 only and for covid negative patients, the test gave positive indication for 200 patients. Construct a confusion matrix for above scenario and find the values of True Positive Rate (TPR), False Positive Rate (FPR), Specificity, Sensitivity metrics. (5)

- (b) Answer the following : (5)

(i) What is the impact of small dataset with respect to large number of features?

(ii) For the given values $\theta_0=0.2$, $\theta_1=0.1$, and $\theta_2=0.1$; predict values of dependent variable y for all

instances of independent variables x_1 and x_2 as given in following data table using linear regression. Also predict mean squared error.

x_1	x_2	y
1	2	3
2	4	5
3	8	9
2	1	1.5

- (c) Cluster the following set of data objects in two clusters by applying one iteration of k-means algorithm. Treat objects 2 and 5 as initial cluster centres. Use Euclidean distance as the distance metric. Determine updated cluster centre coordinates. (5)

Object Number	X-coordinate	Y-coordinate
1	2	4
2	4	6
3	6	8
4	10	4
5	12	4

- (d) Differentiate between linear regression and polynomial regression. Derive the gradient descent algorithm to find the unknown parameters in multivariate linear regression. (5)
- (e) How PCA (Principal Component Analysis) algorithm helps in dimension reduction in machine learning? Write the steps of PCA algorithm. (5)
- (f) What is regularization? Write equations of cost function for regularized linear and regularized logistic regression. What will be the effect on model when the regularization parameter is set to zero? (5)
- (g) Consider the following dataset with 8 training instances. Use k-NN algorithm (for $k=3$) to determine the 'Result' status for a new test instance with values CGPA = 7.6, Assessment = 60 and Project Points = 7. (5)

S.No.	CGPA	Assessment	Project Points	Result
1	9.2	85	8	Pass
2	8	80	7	Pass
3	8.5	81	8	Pass
4	6	45	5	Fail
5	6.5	50	4	Fail
6	8.2	72	7	Pass
7	5.8	38	5	Fail
8	8.9	91	9	Pass

Section – B

2. (a) Consider two features in a dataset and their possible values as shown below : (4)

- Income: values (medium, low, high, very high)
- Status: values (SO, AO, Clerk)

Answer the following questions :

- (i) Using Cartesian product on above feature set, construct a new feature and generate its possible values list.

कालिन्दी महाविद्यालय पुस्तकालय
KALINDI COLLEGE LIBRARY

P.T.O.

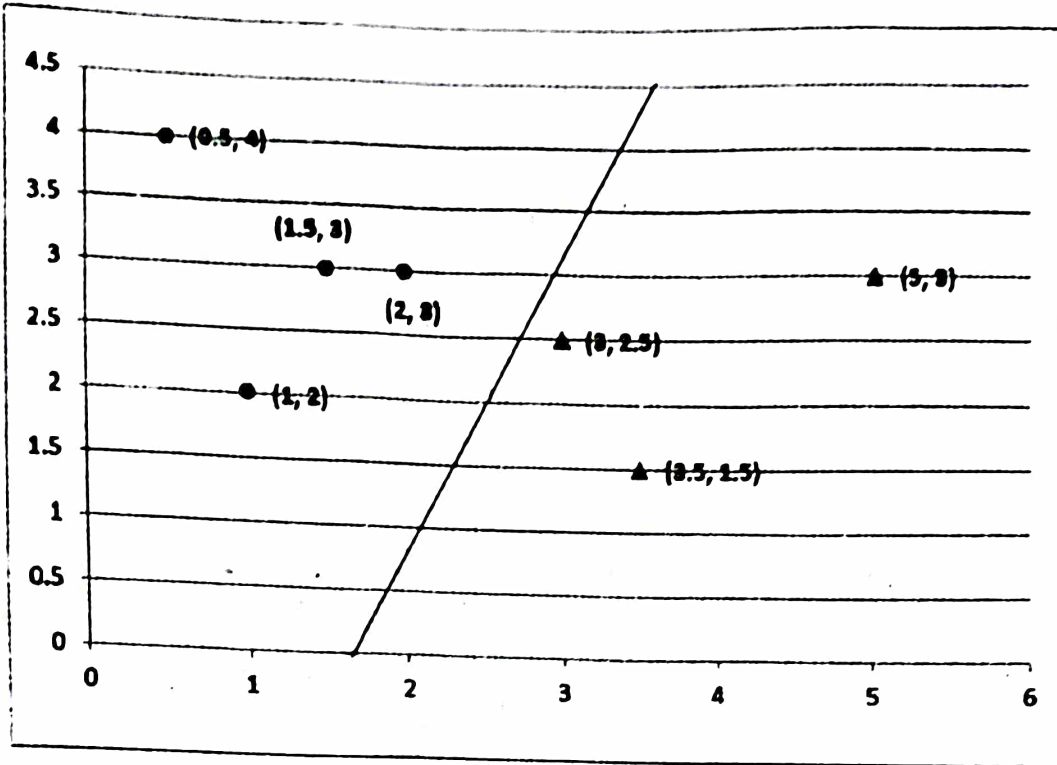
(ii) State one advantage and one disadvantage of above approach for feature construction.

(b) For the given set of points, identify clusters using complete linkage in agglomerative clustering. Use Euclidean distance to calculate the distance between two points. (6)

Points	X coordinate	Y coordinate
P1	1	1
P2	1.5	1.5
P3	5	5
P4	3	4

3. (a) Consider the following two dimensional space with some data points such that circle points represent positive class points and triangular points represent negative class points separated by a decision boundary as shown. (5)

कालिन्दी महाविद्यालय पुस्तकालय
KALINDI COLLEGE LIBRARY



Answer the following questions :

- (i) Identify support vectors, (with respect to SVM classifier applied on above data)
 - (ii) Draw marginal planes, (with respect to SVM classifier applied on above data)
 - (iii) Define Marginal Distance in SVM algorithm.
- (b) Construct neural network for a two input NOR gate using truth table. Show diagram for your generated neural network model with weights.

(5)

कालिन्दी महाविद्यालय पुस्तकालय

KALINDI COLLEGE LIBRARY P.T.O.

4. (a) Apply Naive Bayesian Classifier to Predict whether a car is stolen or not with features {Color:RED, Origin:Domestic, Typer:SUV} based on given dataset. (5)

Color	Type	Origin	Stolen
RED	SPORTS	DOMESTIC	YES
RED	SPORTS	DOMESTIC	NO
RED	SPORTS	DOMESTIC	YES
YELLOW	SPORTS	DOMESTIC	NO
YELLOW	SPORTS	IMPORTED	YES
YELLOW	SUV	IMPORTED	NO
YELLOW	SUV	IMPORTED	YES
YELLOW	SUV	DOMESTIC	NO
RED	SUV	IMPORTED	NO
RED	SPORTS	IMPORTED	YES

- (b) Differentiate between hold out method, leave one out method and k-fold method for cross-validation. Which of the above methods has low bias and high variance. Justify. (5)

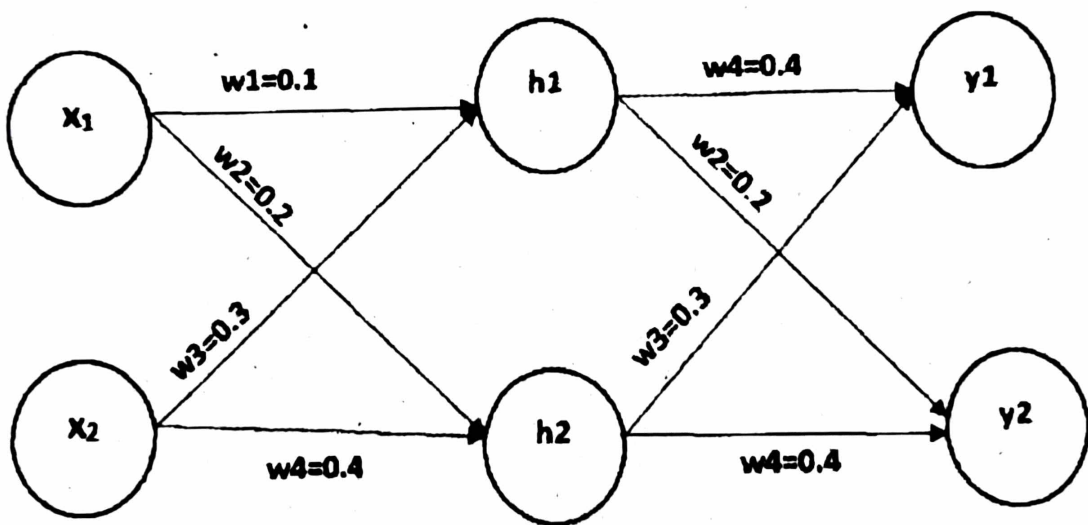
5. (a) Using the data given below, build a logistic regression model to predict whether a student is pass or fail based on exam score using gradient descent algorithm. Assume initial values for model parameters (thetas) as 0 and learning rate as 0.3. Use one iterations of gradient descent algorithm to update the model parameters. (6)

Exam Score (x)	Pass/Fail (y)
50	0
55	0
60	0
65	1
70	1
75	1
80	1
85	1
90	1
95	1

- (b) Using least squares method, learn the regression coefficients for the data given below. Also predict the value of y for $x=12$ using your learned coefficients. (4)

X	Y
2	21
4	27
6	29
8	64
10	86

6. (a)



For given input values of x_1 and x_2 as 0.3 and 0.5 respectively, determine the values of output nodes y_1 and y_2 . Use bias $b_1=0.5$ and $b_2=0.5$. Use sigmoid as the activation function for hidden as well as output layer.

(7)

(b) Explain the effect of following factors in achieving model convergence with respect to gradient descent algorithm.

- Learning rate is too small.

- Learning rate is too large. (3)

7. (a) Consider following training data for 5 persons. For binary classification of a person as sick or not sick create a decision tree model. Show all the steps. (8)

Person No	A1	A2	A3	Class
1	Yes	Yes	Yes	Not Sick
2	Yes	No	Yes	Sick
3	No	No	Yes	Sick
4	No	Yes	Yes	Not Sick
5	No	Yes	No	Sick

(b) Consider the expected and predicted outcomes of a machine learning classifier on a data set containing 7 observations. Calculate the

कालिन्दा महाविद्यालय पुस्तकालय

KALINDI COLLEGE LIBRARY

P.T.O.

3143

12

performance of the classifier using Jaccard Index
metric. (2)

Y expected	0	0	0	0	1	1	1
Y predicted	1	0	0	1	0	1	0