HALDIA INSTITUTE OF TECHNOLOGY

(AN AUTONOMOUS INSTITUTION UNDER MAULANA ABUL KALAM AZAD UNIVERSITY OF TECHNOLOGY, WEST BENGAL)

Paper Code: PCC-CS 603

Paper Name: Machine Learning

Time Allotted: 2 Hours *Full Marks*: 70

The figures in the margin indicate full marks
Candidates are required to give their answers in their own words as far as practicable

Group – A
(Multiple Choice Type Questions)

1. Choose the correct alternatives from the followings: 30 x 1 = 30

(i) Linear Regression is a machine learning algorithm based on _____. [CO1]

a) Unsupervised learning b) Supervised learning c) Reinforcement learning

d) None of these

(ii) The problem of finding hidden structure in unlabeled data is called… [CO1]
a) Supervised learning b) Unsupervised learning c) Reinforcement learning
d) None of the above

(iii) Real-time decisions, Game AI, Learning tasks, Skill acquisition, and Robot Navigation are application of which of the following [CO1]

a) Reinforcement Learning

b) Supervised Learning

c) Unsupervised Learning

d) None of the above

(iv)Decision tree is the most powerful for _____ [CO1]

a) Classification b) Prediction c) Both (a) and (b) d) None of these

(v) Support vector machines is _____ [CO1]

a) Classification learning

b) Unsupervised Machine Learning

c) Supervised Machine Learning

d) Reinforcement learning

(vi) Linear regression assumes that the data follows a linear function, Logistic regression models the data using the_____ [CO1]

a) Continuous function b) Linear function c) Sigmoid function d) Sample function

(vii) What is 'Overfitting' in Machine learning? [CO1]
a) When a statistical model describes random error or noise instead of underlying relationship 'overfitting' occurs.
b) Robots are programed so that they can perform the task based on data they gather from sensors.
c) While involving the process of learning 'overfitting' occurs.

d) A set of data is used to discover the potentially predictive relationship

(viii) Which of the following statement is true about outliers in linear regression? [CO1]

a) Linear regression is sensitive to outliers

b) Linear regression is not sensitive to outliers

c) Can't say

d) None of these

(ix) Which of the following statements is true for K-NN classifiers? [CO1]

a) The classification accuracy is better with larger values of k

b) The decision boundary is smoother with smaller values of k

c) The decision boundary is linear

d) K-NN does not require an explicit training step

(x) What's the objective of the support vector machine algorithm? [CO1]

a) To find an optimal hyperplane in an N-dimensional space that distinctly classifies the data points where N is the number of features.

b) To find an optimal hyperplane in an N-dimensional space that distinctly classifies the data points where N is the number of samples.

c) To find an optimal hyperplane in an N-dimensional space that distinctly classifies the data points where N is the number of target variables.

d) None of these

(xi) How do we perform Bayesian classification when some features are missing? [CO4]

a) We assuming the missing values as the mean of all values.

b) We ignore the missing features.

c) We integrate the posteriors probabilities over the missing features.

d) Drop the features completely.

(xii) Decision tree is a flowchart like_____ [CO1]

a) Leaf structure b) Tree structure c) Steam d) None of these

(xiii) What is 'Test set'? [CO1]
a) Test set is used to test the accuracy of the hypotheses generated by the learner.

b) It is a set of data is used to discover the potentially predictive relationship.

c) Both a & b

d) None of above

(xiv) In SVM, non linear problem can be solved by transforming data from _____ dimensional space into _____ dimensional space. [CO1]

a) High, Low b) Low, High c) Low, Medium d) Medium, High

(xv) Naive Bayes classifiers are a collection----------of algorithms. [CO1]

a) Classification b) Clustering c) Regression d) all

(xvi) Which of the following is true about Residuals? [CO3]
a) Lower is better

b) Higher is better

c) (a) or (b) depend on the situation

d) None of these

(xvii) In SVM, the dimension of the hyperplane depends upon which one? [CO5]

a) The number of features b) The number of samples

c) The number of target variables d) All of the above

(xviii) Decision trees can handle_____ [CO2]

a) High dimensional data b) Low dimensional data

c) Medium dimensional data d) None of these

(xix) Which of the following is not an example of ensemble method? [CO1]

a) AdaBoost b) Decision tree c) Random Forest d) Bootstrapping

(xx) The goal of clustering a set of data is to [CO1]

a) Divide them into groups of data that are near to each other

b) Choose the best data from the set

c) Predict the class of data

d) Determine the nearest neighbors of each of the data

(xxi) Which one of the following statements is TRUE for a Decision Tree? [CO4]

a) Decision tree is only suitable for the classification problem statement.

b) In a decision tree, the entropy of a node decreases as we go down a decision tree.

c) In a decision tree, entropy determines purity.

d) Decision tree can only be used for only numeric valued and continuous attributes.

(xxii) Slack variables ε, can be added to allow misclassification of difficult or noisy examples, resulting margin is called? [CO1]

a) Soft Margin b) Null Margin c) High Margin d) Low Margin

(xxiii) _____is the measure of uncertainty of a random variable, it characterizes the impurity of an arbitrary collection of examples. [CO1]

a) Information Gain b) Gini Index c) Entropy d) none of these

(xxiv) In SVM, if the number of input features is 3, then the hyperplane is a _____. [CO1]

a) line b) circle c) plane d) None of these

(xxv) A nearest neighbor approach is best used [CO3]

a) With large-sized datasets.

b) When irrelevant attributes have been removed from the data.

c) When a generalized model of the data is desirable.

d) When an explanation of what has been found is of primary importance.

(xxvi) If there is only a discrete number of possible outcomes (called categories), the process becomes a_____. [CO1]

a)Regression b) Classification c) Model free d) Categories

(xxvii) Which of the following statement is False in the case of the KNN Algorithm? [CO2]

a) For a very large value of K, points from other classes may be included in the neighborhood.

b) For the very small value of K, the algorithm is very sensitive to noise.

c) KNN is used only for classification problem statements.

d) KNN is a lazy learner.

(xxviii) How do you choose the right node while constructing a decision tree? [CO3]

a) An attribute having high entropy

b) An attribute having high entropy and information gain

c) An attribute having the lowest information gain.

d) An attribute having the highest information gain.

(xxix) What is the minimum no. of variables/ features required to perform clustering? [CO1]

a) 0 b) 1 c) 2 d) 3

(xxx) A linear regression model assumes "a linear relationship between the input variables and the single output variable." What's the meaning of this assumption? [CO1]

a) The output variable can't be calculated from a linear combination of the input variables

b) The output variable can be calculated from a linear combination of the input variables

c) Input variables can be calculated from a linear combination of the output variables

d) Output variable = sum of the input variables

Group – B

(Multiple Choice Type Questions)

2. Choose the correct alternatives from the followings: 10 x 2 = 20

(i) A machine learning problem involves four attributes plus a class. The attributes have 3, 2, 2, and 2 possible values each. The class has 3 possible values. How many maximum possible different examples are there? [CO1,CO5]

a) 12 b) 24 c) 48 d) 72

(ii) Designing a machine learning approach involves:- [CO1]

a) Choosing the type of training experience

b) Choosing the target function to be learned and representation for the target function

c) Choosing a representation for the target function

d) All of the above

(iii) For the dataset given in Table 1 below to learn a decision tree, find the approximate entropy H(Passed). This decision tree predicts whether students pass or not (Y for yes or N for no), based on their past CGPA scores (H for high, A for average, and L for Low) and whether they prepared or not (Y or N). [CO5, CO6]

CGPA	Prepared	Passed
H	F	T
H	T	T
M	F	F
M	T	T
L	F	F
L	T	T

a) 0.92 b) 0.66 c) 1.92 d) 1.32

(iv) Which of the following statements is true about the learning rate alpha in gradient descent? [CO3]

a) If alpha is very small, gradient descent will be fast to converge. If alpha is too large, gradient descent will overshoot

b) If alpha is very small, gradient descent can be slow to converge. If alpha is too large, gradient descent will overshoot

c) If alpha is very small, gradient descent can be slow to converge. If alpha is too large, gradient descent can be slow too

d) If alpha is very small, gradient descent will be fast to converge. If alpha is too large, gradient descent will be slow

(v) In the mathematical Equation of Linear Regression $Y = \beta1 + \beta2X + \epsilon$, ($\beta1$, $\beta2$) refers to _____ [CO1]

a) (X-intercept, Slope) b) (Slope, X-Intercept)

c) (Y-Intercept, Slope) d) (slope, Y-Intercept)

(vi) If TP=9 FP=6 FN=26 TN=70 then Error rate will be [CO2]

a) 45 percentage b) 99 percentage c) 28 percentage

d) 20 percentage

(vii) In an election, N candidates are competing against each other and people are voting for either of the candidates. Voters don't communicate with each other while casting their votes. Which of the following ensemble method works similar to above-discussed election procedure? [CO4]

Hint: Persons are like base models of ensemble method.

a) Bagging b) Boosting c) (a) or (b)

d) None of these

(viii) 8 observations are clustered into 3 clusters using K-Means clustering algorithm. After first iteration clusters,
C1, C2, C3 has following observations:
C1: {(2,2), (4,4), (6,6)}
C2: {(0,4), (4,0),(2,5)}
C3: {(5,5), (9,9)}
What will be the cluster centroids if you want to proceed for second iteration? [CO5,CO6]

a) C1: (4,4), c2: (2,2), c3: (7,7) b) C1: (6,6), c2: (4,4), c3: (9,9)

c) C1: (2,2), c2: (0,0), c3: (5,5) d) C1: (4,4), c2: (2,3), c3: (7,7)

(ix) In Naive Bayes equation P(C / X)= (P(X / C) *P(C)) / P(X) which part considers "likelihood"? [CO1]

a) p(x/c) b) p(c/x) c) p(c) d) p(x)

(x) The selling price of a house depends on many factors. For example, it depends on the number of bedrooms, number of kitchen, number of bathrooms, the year the house was built, and the square footage of the lot. Given these factors, predicting the selling price of the house is an example of _____ task. [CO4]
a) Binary classification b) Multi label classification
c) Simple linear regression d) Multiple linear regression

Group – C

(Short Answer Type Questions)

3. Attempt any four from the followings: 4 x 5 = 20

i) a) Why does logistic regression come under the category of classification problems?[CO1]

b) How is logistic regression different from linear regression? [CO1] 3+2

ii) Apply k-means algorithm in given data for k=2. Use $C_1(80)$ and $C_2(250)$ as initial cluster centers. Data:234,123,456,23,34,56,78,90,150,116,117,118,199.[CO5,CO6] +5

iii) a) Explain in brief the working of SVM.

b) What is margin in SVM and how optimal hyperplane is drawn with the help of margins? [CO1] 3+2

iv) a) Define Bayes Theorem. [CO1]

b) How does a naïve Bayes' classifier help in classifying the output class? [CO1]

v) a) How does the K-NN algorithm make the predictions on the unseen dataset? [CO2]

b) Why K-NN algorithm known as the lazy learner? [CO3] 4+1

vi) a) What is ensemble learning? [CO1]

b) Differentiate between bagging and boosting? [CO3, CO4] 2+3

mohuya sasmal ray

Signature of the Paper Setter/Moderator

Answer Key:

Group – A

1. (i) b	(ii) b	(iii) a	(iv) c	(v) c	(vi) c
(vii) a	(viii) a	(ix) d	(x) a	(xi) c	(xii) b
(xiii) a	(xiv) b	(xv) a	(xvi) a	(xvii) a	(xviii) a
(xix) b	(xx) a	(xxi) b	(xxii) a	(xxiii) c	(xxiv) c
(xxv) b	(xxvi) b	(xxvii) c	(xxviii) d	(xxix) b	(xxx) b

Group – B

2. (i) d	(ii) d	(iii) a	(iv) b	(v) c	(vi) c
(vii) a	(viii) d	(ix) a	(x) d		

Group – C

3. (i)	(ii)	(iii)	(iv)	(v)	(vi)