University of Mumbai

Examinations Summer 2022

Time: 2 hour 30 minutes Max. Marks: 80

==========

Q1.	**Choose the correct option for following questions. All the Questions are compulsory and carry equal marks**
1.	Machine learning is a branch of..
Option A:	Artificial intelligence
Option B:	speech processing
Option C:	Language processing
Option D:	java
2.	What does K stand for in K mean algorithm?
Option A:	Number of Clusters
Option B:	Number of Data
Option C:	Number of Attributes
Option D:	Number of Iterations
3.	Feature selection tries to eliminate features that are
Option A:	Rich
Option B:	important
Option C:	Irrelevant
Option D:	Relevant
4.	During the treatement of cancer patients , the doctor needs to be very careful about which patients need to be given chemotherapy. Which metric should we use in order to decide the patients who should given chemotherapy?
Option A:	precision
Option B:	recall
Option C:	call
Option D:	score
5.	Targetted marketing, Recommended Systems, and Customer Segmentation are applications in which of the following
Option A:	Supervised Learning: Classification
Option B:	Unsupervised Learning: Clustering
Option C:	Unsupervised Learning: Regression
Option D:	Reinforcement Learning
6.	CART stands for…
Option A:	classification and regression tree
Option B:	choosing a regression task
Option C:	classification and regression task
Option D:	classification along regression task
7.	Naïve Bayes Algorithm is a ……. learning algorithm.
Option A:	Supervised
Option B:	Reinforcement
Option C:	Semi supervised
Option D:	Unsupervised

8.	Which of the following can only be used when training data are linearly separable?
Option A:	linear hard-margin svm
Option B:	linear logistic regression
Option C:	linear soft margin svm
Option D:	the centroid method
9.	Impact of high variance on the training set ?
Option A:	depents upon the dataset
Option B:	underfitting
Option C:	both underfitting & overfitting
Option D:	overfitting
10.	What do you mean by a hard margin?
Option A:	The SVM allows very low error in classification
Option B:	The SVM allows very high error in classification
Option C:	The SVM allows no error in classification
Option D:	The SVM does not allow error in classification

Q2. (20 Marks Each)	**Solve any Two Questions out of Three 10 marks each**
A	Explain the steps of developing Machine Learning applications in detail.
B	Explain regression line, scatter plot, error in prediction ; best fitting line.
C	Cluster the following eight points (with (x, y) representing locations) into three clusters: A1(2, 10), A2(2, 5), A3(8, 4), A4(5, 8), A5(7, 5), A6(6, 4), A7(1, 2), A8(4, 9) Initial cluster centers are: A1(2, 10), A4(5, 8) and A7(1, 2). The distance function between two points a = (x1, y1) and b = (x2, y2) is defined as- d(a, b) = |x2 – x1| + |y2 – y1| Use K-Means Algorithm to find the three cluster centers after the one iteration
Q3. (20 Marks Each)	**Solve any Two Questions out of Three 10 marks each**
A	Compare and contrast Linear and Logistic regressions with respect to their mechanisms of prediction.
B	Explain in detail PCA for dimension reduction.

C

Find complete linkage method of hierarchical clustering to find clusters of 5 data points with following distance matrix.

	1	2	3	4	5
1	*0*	*9*	*3*	*6*	*11*
2	*9*	*0*	*7*	*5*	*10*
3	*3*	*7*	*0*	*9*	*2*
4	*6*	*5*	*9*	*0*	*8*
5	*11*	*10*	*2*	*8*	*0*

Q4. (20 Marks Each)

Solve any Two Questions out of Three **10 marks each**

A

Explain K-mean clustering algorithm giving suitable example. Also, explain how K-mean clustering differs from hierarchical clustering.

B

What is support vector machine? What do you mean by support vectors, hyper plane and margin,support vectors?What will be the boundary for one dimensional data, two dimensional data and three dimensional data. Explain with suitable examples.

What is SVM? Explain the following terms: hyperplane, separating hyperplane, margin and support vectors with suitable example.

C

Create a decision tree using gini index to classify following dataset:-

Sr. No.	**Income**	**Age**	**Own house**
1	Very high	Young	Yes
2	High	Medium	Yes
3	Low	Young	No
4	High	Medium	Yes
5	Very high	Medium	Yes
6	Medium	Young	Yes
7	High	Old	Yes
8	Medium	Medium	No
9	Low	Medium	No
10	Low	Old	No
11	High	Young	Yes
12	Medium	Old	No