Charotar University of science & Technology (CHARUSAT)
Devang Patel Institute of Advance Technology and Research
Unit test 1

Subject Name: Machine Learning (CS344)
Date: 20/01/2020
Time: 9:45 to 10:45
Total Marks: 30

SET 2

1. In which algorithm, computation time complexity is more to test unseen sample? **(1 Mark)**

    a) Decision Tree

    b) K-Nearest Neighbors

    c) SVM

    d) neighbourhood

2. Cluster quality depends on _____ intra-class distance and _____ inter-class distance. **(1 Mark)**

    a) minimum, maximum

    b) average, minimum

    c) maximum, minimum

    d) minimum, average

3. Following is/are library/libraries of Machine Learning algorithm. **(1 Mark)**

    a) TensorFlow

    b) scikit-learn

    c) keras

    d) All of the above

4. Gini Index would be _____ if dataset is perfectly classified. **(1 Mark)**

    a) 0

    b) 1

    c) 1/2

    d) 1/3

5. If the samples are an equally divided by target classes, it has entropy of _____. **(1 Mark)**

    a) 0

    b) 1

    c) 1/2

    d) 1/3

6. In machine learning, most of the applied features need to be identified by an expert before feeding to an algorithm compared to deep learning. **(1 Mark)**

    a) True

    b) False

7. How do you handle missing or corrupted data in a dataset? **(1 Mark)**

    a) Drop missing rows or columns

    b) Replace missing values with mean/median/mode

    c) Assign a unique category to missing values

    d) All of the above

8. Which of the following is finally produced by Hierarchical Clustering? **(4 Mark)**
    a. final estimate of cluster centroids
    b. tree showing how close things are to each other
    c. assignment of each point to clusters
    d. all of the mentioned

9.

Price of LED bulb in dollars (X)	Number of LED sold (y)
2	4
3	5
5	7
7	10
9	15

Ronak is the owner of a Electrical shop. Given table indicates the price of different LED bulb vs number of LED bulb sold at his shop over a period of one week. Use LSM (Least Squares Method) algorithm for liner regression and predict the number of bulb will be sold if price of led bulb is 8.3 dollars. **(1 Mark)**

a. 12.5
b. 13
c. 14
d. 13.5

10. For above data, where "Diabetic" is the target variable, using Gini Index if a decision tree is made. Which attribute has the highest Gini Index? **(6 Mark)**

WEIGHT	FOOD INTAKE	Exercising	DIABETIC
< 80	Low	Never	No
>= 80	Medium	Regularly	No
< 80	High	Never	Yes
>= 80	High	Occasionally	No
< 80	Medium	Never	No
>= 80	Low	Never	Yes
< 80	Low	Occasionally	No
>= 80	High	Never	Yes
< 80	Low	Regularly	No

a. Weight
b. Food Intake
c. Exercising
d. Can't Say

11. For K=3, and centers initialized as C1 = A4, C2 = A5 and C3 = A6, what will be the cluster center after the first iteration of K-means clustering algorithm. Use Euclidian Distance **(6 Mark)**

	X	Y
A1	1	2
A2	3	5
A3	6	0
A4	2	7
A5	4	1
A6	5	5
A7	7	3
A8	9	4

a. (2, 7), (4, 1), (5, 5)
b. (2, 7), (3.66, 1), (6, 4.75)
c. (2.33, 6.66), (3.66, 1.33), (6,4.25)
d. (2, 7), (3.66, 1), (6, 4.25)

12. Using K- Nearest Neighbour, what will be the value marked as " ? ". K = 5. Use Manhattan Distance ($X_i - X_j$) + ($Y_i - Y_j$). Values for ID no. 8 and ID no. 9 are: **(6 Mark)**

ID	LENGTH	WEIGHT	SPEED	ANIMAL
1	10	40	80	Tiger
2	12	45	110	Leopard
3	11	42	115	Leopard
4	13	38	120	Leopard
5	09	39	90	Tiger
6	14	43	95	Leopard
7	15	41	100	Tiger
8	16	46	105	?
9	10	40	100	?

a. Leopard, Leopard
b. Leopard, Tiger
c. Tiger, Leopard
d. Tiger, Tiger