CHAROTAR UNIVERSITY OF SCIENCE AND TECHNOLOGY (CHARUSAT)
DEVANG PATEL INSTITUE OF ADVANCE TECHNOLOGY AND RESEARCH
Subject: Machine Learning (CE473)
Question Bank

1. Explain the concept learning in brief. List down the application areas where concept learning works better than other approaches.
2. What is Machine Learning? What is the need of it? Briefly explain the Machine Learning characteristics.
3. What is Classification? Which are the phases of Classification process? Differentiate between Binary Classification and Multiclass Classification
4. How to measure the performance of a Classifier? How to compare the performance of one Classifier with the other? What is a Confusion Matrix? What is a Contingency Table?
5. Explain the Regression along with an equation. How to measure performance of Regression? Describe error calculation in Regression.
6. What is over-fitting and under-fitting? Which are the reasons for over-fitting and under-fitting?
7. Explain the k-NN classification along with example. Which are the effects of k on result of k-NN classifier?
8. What is Perceptron? What is Multi-Layer Feed Forward Networks? What is SVM? How does SVM differ from Artificial Neural Networks (ANN)?
9. Give three computer applications for which Machine Learning approaches seem appropriate and three applications for which they seem inappropriate.
10. Explain in brief Descriptive and Predictive Tasks of Machine Learning.
11. Suppose $\theta 1$ is at local optimum of $J(\theta 1)$ such as shown in the figure. What will one step of gradient descent will do? Justify your answer.
12. Give Decision Tree representations for following Boolean Functions.
    a. A $\vee$ (B $\wedge$ C)
    b. (A $\wedge$ B) $\vee$ (C $\wedge$ D)
13. Which of the following will be true about k in k-NN in terms of variance? Why?
    a. When you increase the k, the variance will increase
    b. When you decrease the k, the variance will increase
    c. Cannot say
    d. None of the above
14. Describe K-nearest neighbor algorithm. Why is it known as instance-based Learning?
15. What is Information Gain? Write down and discuss Inductive Decision Tree (ID3) algorithm.
16. Differentiate: Supervised Learning V/S Unsupervised Learning.
17. Apply K- means algorithm on given data for k=3. Use $C_1(2)$, $C_2(16)$ and $C_3(38)$ as initial cluster centres.

    Data: 2,4,6,3,31,12,15,16,38,35,14,21,23,25,30

18. Explain following:
    a. Manhattan Distance
    b. Over-fitting
    c. Euclidean distance
19. Divide the given sample data in two clusters using K- means algorithm. (Use Euclidean Distance)

HEIGHT	WEIGHT
185	72
170	56
168	60
179	68
182	72
188	77
180	71
180	70
183	84
180	88
180	67
177	76

20. Plot a dendogram using complete linkage (agglomerative clustering) for the following dataset.

ITEM	A	B	C	D	E
A	0	9	3	6	11
B	9	0	7	5	10
C	3	7	0	9	2
D	6	5	9	0	8
E	11	10	2	8	0

21. Perform agglomerative algorithm on the following data and plot a dendogram using single link approach. The given data indicates the distance between elements.

ITEM	E	A	C	B	D
E	0	1	2	2	3
A	1	0	2	5	3
C	2	2	0	1	6
B	2	5	1	0	3
D	3	3	6	3	0

22. If cluster centers are P3, P4 and P9 then What will be the clusters after two iterations? Use Manhattan Distance?

	X
P1	12
P2	15
P3	23
P4	4
P5	7
P6	28
P7	30
P8	35
P9	45
P10	51