

Question Bank

Module-1

1. What is machine learning? Discuss the issues in machine learning.
2. Differentiate between supervised, unsupervised, and reinforcement learning.
3. List and explain the challenges and applications of Machine Learning.
4. Explain the knowledge pyramid and its relation with machine learning.
5. Briefly explain the relation of Machine Learning with AI and Deep Learning.
6. “ Data science is an Umbrella term that encompasses many fields”. Justify the statement.
7. Differentiate between Classification and Regression.
8. Briefly discuss the Machine Learning Process with a neat diagram.
9. Justify how the six Vs forms the Big Data.
10. Differentiate between Structured and Unstructured Data.
11. Big Data Analysis Framework consists of four layered architecture. Explain
12. State and explain the different ways to handle the missing data.
13. What is Binning? Consider the following set $S=\{12,14,19,22,24,26,28,31,34\}$. Apply various binning techniques and show the result.
14. What are the different ways of data transformation?
15. Consider the set $S=\{88,90,92,94\}$. Apply min-max procedure and map the marks to a new range {0,1}
16. Differentiate between Qualitative and Quantitative data.
17. How is the data classified based on the number of variables?
18. Illustrate types of categorical and numerical data types with examples.
19. Explain the importance of visualizing data before preparing it for a machine learning model.
20. Explain the different measures of central tendency with suitable examples.
21. Are classification and clustering the same or different? Justify
22. Considering the following table

For the given Tables 2.6 and 2.7, perform the descriptive analysis of data:

Table 2.6: Sample Data

Age	Weight
1	4.2
2	4.5
3	4.7
4	5.2
5	6
6	6.2
7	7
8	7.2
9	7.5
10	8.5

Table 2.7: Students Marks Table

Sid	English	Hindi	Maths	Science
1	45	70.5	90	40
2	60	72.5	80	45
3	60	80	90	50
4	80	80	90	80
5	85	72	70	60

For univariate attribute weight,english and maths find the following

- i. Mean, median ,mode
 - ii. variance and Standard deviation
 - iii. five point summary
 - iv. skewness and kurtosis
 - v. covariance between english and hindi marks
23. Explain the primary data cleaning process for handling missing Data.
24. Discuss the different ways to store and organise the data.
25. “A computer program is said to learn from experience E, with respect to task T and some performance measure P, if its performance on T measured by P improves with experience E.” Justify

Question Bank
Module-2

1. Explain the role of the Scatter Plot for displaying the relationship for bivariate data.
2. Find the Covariance and correlation for the data $X=\{1,2,3,4,5\}$ and $Y=\{1,4,9,16,25\}$.
3. Explain the role of Heat Map and Pair Plot in visualising the Multivariate data.
4. Explain Gaussian Elimination Method and apply it on :

$$2x_1 + 5x_2 = 7$$

$$6x_1 + 12x_2 = 18$$

5. Explain the importance of Probability Distribution for machine learning.
6. Differentiate between continuous and discrete probability distribution.
7. Explain Parametric Density estimation.
8. Discuss the importance of Expectation-Maximization model.
9. Explain Parzen window
10. What is Feature Engineering? Explain its importance in the field of Machine Learning.
11. Discuss the problems related to machine learning.
12. Discuss the aspects based on which features can be removed from the dataset.
13. Differentiate between Stepwise Forward Selection and Stepwise Backward Selection.
14. Explain the role of Principal Component Analysis(PCA) in dimensionality reduction.
15. What do you mean by a well –posed learning problem? Explain the important features that are required to well –define a learning problem.
16. What are the steps to be followed for designing a learning system.
17. Explain the importance of training experience in designing a learning system.
18. What is concept learning and what are its requirements?
19. Differentiate between Hypotheses Space and Version Space.
20. What is a consistent hypothesis? Explain with example.
21. Illustrate the process of Specialization and Generalization with suitable example.
22. Explain Find-S algorithm and explain its working for the given dataset:

Example	Sky	AirTemp	Humidity	Wind	Water	Forecast	Enjoy Sport
1	Sunny	Warm	Normal	Strong	Warm	Same	Yes
2	Sunny	Warm	High	Strong	Warm	Same	Yes
3	Rainy	Cold	High	Strong	Warm	Change	No
4	Sunny	Warm	High	Strong	Cool	Change	Yes

23. Consider the below given dataset and generate the Version Space using Candidate-Elimination algorithm:

S.No.	Horns	Tail	Tusks	Paws	Fur	Color	Hooves	Size	Elephant
1.	No	Short	Yes	No	No	Black	No	Big	Yes
2.	No	Short	No	No	No	Brown	No	Medium	Yes
3.	Yes	Short	No	No	No	Brown	Yes	Medium	No
4.	No	Short	Yes	No	No	Black	No	Medium	Yes
5.	No	Long	No	Yes	Yes	White	No	Medium	No
6.	No	Short	Yes	Yes	Yes	Black	No	Big	Yes

24. Consider the “Japanese Economy Car” concept and instance given in Table 1., Illustrate the hypothesis using Candidate Elimination Learning algorithm.

CGPA	Interactivity	Practical Knowledge	Communication Skills	Logical Thinking	Interest	Job Offer
≥9	Yes	Excellent	Good	Fast	Yes	Yes
≥9	Yes	Good	Good	Fast	Yes	Yes
≥8	No	Good	Good	Fast	No	No
≥9	Yes	Good	Good	Slow	No	Yes

25. What are the different approaches used for selecting a machine learning model.
 26. Compare k-fold cross validation and LOOCV Re-sampling methods.
 27. Differentiate between k-fold cross validation and stratified k-fold cross validation.
 28. What is confusion matrix? Explain the below terms with respect to the confusion matrix:

- a) Sensitivity
- b) Specificity
- c) Accuracy
- d) Precision
- e) Recall

29. Discuss the importance of ROC Curve in visualising the model performance.
 30. Apply LU decomposition for the given matrix:

$$A = \begin{pmatrix} 1 & 2 & 4 \\ 3 & 3 & 2 \\ 3 & 4 & 2 \end{pmatrix}$$

MACHINE LEARNING (BCS602)
Question Bank
Module-3

1. Compare and contrast Instance based learning and Model based Learning.
2. Why instance based learners are called lazy learners?
3. k-NN method is referred to as memory based method. Justify.
4. What is meant by locally-weighted regression.
5. The values of independent variable x and dependent value y are given as:

X	Y
1	2
2	5
3	3
4	4
5	5

Apply Linear Regression and evaluate the value of y when x=6 and x=9

6. What is the role of a regression model in exploratory data analysis?
7. Distinguish between Logistic regression and linear regression.
8. Briefly explain Locally Weighted Regression.
9. How does polynomial regression model works?
10. How does the structure of a decision tree help in classifying a data instance.
11. What are the different metrics used in deciding the splitting attribute.
12. Relate Entropy and Information Gain.
13. How are continuous attributes discretized? Consider the training dataset and discretize the attribute “Percentage”:

S.No.	Percentage	Award
1.	95	Yes
2.	80	Yes
3.	72	No
4.	65	Yes
5.	95	Yes
6.	32	No
7.	66	No
8.	54	No
9.	89	Yes
10.	72	Yes

14. Write the algorithm for constructing a decision tree using CART
15. Explain how ID3 and C4.5 are different for constructing decision tree.

Module-4

1. Design a **perceptron network** to implement **NAND gate**. Choose w_1 & w_2 as 0.2 & 0.2 respectively and bias as -0.2. Use a **learning rate $\alpha = 0.4$** . Train the model using the truth table of the NAND gate and show the updated weights after each epoch. Use Step function as activation function.
2. Consider a perceptron to represent the boolean function AND with the initial weights $w_1=0.3$ and $w_2=-0.2$, learning rate $\alpha = 0.2$ and bias $\Theta = 0.4$. The activation function used is step function $f(x)$, which gives the output as binary. If the value of $f(x)$ is greater than or equal to 0 then the output is 1 else it is 0. Design a perceptron that performs the boolean function AND and update the weights until the boolean function gives the desired output.
3. Explain the role of activation functions in neural network.
4. Differentiate between prior, posterior and likelihood probability.
5. Explain MAP hypothesis using bayes theorem?
6. Define perceptron? explain the representational power of perceptrons.
7. Given a test data (Assessment Marks=75, Assignment Marks=6, Seminar Done=Poor), predict the result of the student using Gaussian Naive Bayes Classifier

Outlook	Temperature	Humidity	Windy	PlayTennis
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes

8. Apply Naïve Bayes Algorithm on the given dataset:

Test data: True, False, False, True

S.No.	Toothed	Hair	Breathes	Legs	Species
1	True	True	True	True	Mammal
2	True	True	True	True	Mammal
3	True	False	True	False	Reptile
4	False	True	True	True	Mammal
5	True	True	True	True	Mammal
6	True	True	True	True	Mammal
7	True	False	True	False	Reptile
8	True	False	True	False	Reptile
9	True	True	True	True	Mammal
10	False	False	True	True	Reptile

9. Explain zero probability error with an example.
10. Define Artificial neural network? Explain biological learning models

Module-5

1. Illustrate CLIQUE algorithm. Also provide its advantages and disadvantages.
2. Compare Reinforcement Learning and Supervised Learning.
3. Describe Monte-Carlo method with algorithm.
4. what are the advantages and disadvantages of clustering algorithm?
5. For the below table apply single and complete linkage algorithm and draw the dendrogram for the same .

Objects	X	Y
0	1	4
1	2	8
2	5	10
3	12	18
4	14	28

6. Illustrate the DBSCAN algorithm and discuss its benefits.
7. Explain the types of rewards. How are rewards evaluated?
8. Describe Markov Decision Process of Reinforcement learning with necessary diagram.
9. Differentiate clustering and classification?
10. What are the components of RL? explain with a diagram?
11. What is Policy? What are its types?
12. Explain the different types of agents in Reinforcement Learning.
13. Briefly explain Q-Learning.
14. Consider the following data shown in Table. Apply the k-means algorithm with k=2 with seeds (3,5) (7,8) and show the result.

S.No.	X	Y
1.	3	5
2.	7	8
3.	12	5
4.	16	9

15. Apply k-means clustering algorithm for the following dataset with initial value of object 2 and 5 with coordinates values (4,6) and (12,4) as initial seed

Objects	X-coordinates	Y-coordinates
1	2	4
2	4	6
3	6	8
4	10	4
5	12	4