BCS602 TIE SIMP Questions

Set by CSE -TIE review team- Based on 50+ IA question papers, MQP, Textbook

Module 2

1. Differentiate between bivariate and multivariate data. Give examples and explain the importance of multivariate analysis in real-world datasets.
2. What are the common visualization techniques for multivariate data? Explain pair plots, parallel coordinates, heatmaps, and scatter matrix.
3. Explain the following multivariate statistical measures -Covariance, Correlation Matrix, Mahalanobis Distance, and their role in identifying relationships among features.
4. Solve the following

(i) Solve the following set of equations using Gaussian Elimination method.

$$2x_1 + 4x_2 = 6$$
$$4x_1 + 3x_2 = 7$$

(ii) Find *LU* decomposition of the given matrix:

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

5. Explain.
(i) conditional probability (ii) Bayes theorem, (iii) probability distributions (Normal, Bernoulli, Binomial).
6. Write short notes on gradient descent, partial derivatives, and cost minimization.
7. What is Feature Engineering? Explain feature selection vs extraction, importance of domain knowledge, and techniques like one-hot encoding, binning, and scaling – 12M
8. Explain Dimensionality Reduction technique and solve
(i) Let the data points be $\begin{pmatrix} 2 \\ 6 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ 7 \end{pmatrix}$. Apply PCA and find the transformed data.

(ii) **Find SVD of the matrix:**

$$A = \begin{pmatrix} 1 & 2 \\ 4 & 9 \end{pmatrix}$$

9. Explain (i)Concept Learning with a task example(Elephant learning task)
   (ii)Explain Hypotheses Space search by Find S Algorithm, Explain its limitations
   (iii) Explain and apply candidate elimination algorithm for the given dataset

Sky	Temperature	Humid	Wind	Water	Forest	Output
sunny	warm	normal	strong	warm	same	yes
sunny	warm	high	strong	warm	same	yes
rainy	cold	high	strong	warm	change	no
sunny	warm	high	strong	cool	change	yes

Module 1

1. Define Machine Learning. Explain its relationship with Artificial Intelligence, Data Science, Statistics, and Pattern Recognition.
2. Describe the types of Machine Learning with suitable examples and diagrams.
3. Define data. Explain the 6Vs of Big Data with examples..
4. Differentiate between structured, semi-structured, and unstructured data. Give examples..
5. Explain the Machine Learning process (CRISP-DM model) in detail.
   (Cover the six steps: Business understanding → Data understanding → Data preparation → Modeling → Evaluation → Deployment.)
6. List and explain five real-world applications of machine learning.
7. Differentiate between univariate, bivariate, and multivariate data with examples.
8. What are the common data preprocessing techniques used in Machine Learning?.
9. Define mean, median, mode, variance, standard deviation, skewness, and kurtosis. Explain with examples.

Module 3

1. Explain the K-Nearest Neighbor (KNN) algorithm.

Example 4.1: Consider the student performance training dataset of 8 data instances shown in Table 4.2 which describes the performance of individual students in a course and their CGPA obtained in the previous semesters. The independent attributes are CGPA, Assessment and Project. The target variable is 'Result' which is a discrete valued variable that takes two values 'Pass' or 'Fail'. Based on the performance of a student, classify whether a student will pass or fail in that course.

Table 4.2: Training Dataset *T*

S.No.	CGPA	Assessment	Project Submitted	Result
1.	9.2	85	8	Pass
2.	8	80	7	Pass
3.	8.5	81	8	Pass

(*Continued*)

S.No.	CGPA	Assessment	Project Submitted	Result
4.	6	45	5	Fail
5.	6.5	50	4	Fail
6.	8.2	72	7	Pass
7.	5.8	38	5	Fail
8.	8.9	91	9	Pass

(b) Apply weighted KNN to classify a test data point for the above example

2. Differentiate between Nearest Centroid Classifier and KNN. Predict the class using NNN for the below chart, the target classes are A or B

X	Y	Class
3	1	A
5	2	A
4	3	A
7	6	B
6	7	B
8	5	B

4. What is linear regression? Derive the equation for line of best fit using least squares. Explain slope (β1), intercept (β0), and residual error.
5. Explain Logistic Regression with sigmoid function. Derive decision boundary and show how probabilities are mapped using the sigmoid curve.
6. Explain Regressional analysis for all types of regression-10M
7. Describe the ID3 decision tree algorithm. Explain entropy, information gain, feature selection, and stopping conditions.
8. Construct the root node of a decision tree using ID3.

Example 6.3: Assess a student's performance during his course of study and predict whether a student will get a job offer or not in his final year of the course. The training dataset *T* consists of 10 data instances with attributes such as 'CGPA', 'Interactiveness', 'Practical Knowledge' and 'Communication Skills' as shown in Table 6.3. The target class attribute is the 'Job Offer'.

Table 6.3: Training Dataset *T*

S.No.	CGPA	Interactiveness	Practical Knowledge	Communication Skills	Job Offer
1.	≥9	Yes	Very good	Good	Yes
2.	≥8	No	Good	Moderate	Yes
3.	≥9	No	Average	Poor	No
4.	<8	No	Average	Good	No
5.	≥8	Yes	Good	Moderate	Yes
6.	≥9	Yes	Good	Moderate	Yes
7.	<8	Yes	Good	Poor	No
8.	≥9	No	Very good	Good	Yes
9.	≥8	Yes	Good	Good	Yes
10.	≥8	Yes	Average	Good	Yes

9. Compare decision trees with regression models. Explain interpretability, non-linearity, overfitting, and training time.

Module 4

1. What is probability-based learning? Explain its need in ML, advantages in uncertain domains, and the Bayesian approach to modeling.
2. State and explain Bayes' Theorem.S
3. Describe the Naïve Bayes classifier. State assumptions, algorithm steps, and use of conditional probability in classification.
4. Use Naïve Bayes classifier for a classification task.

Example 8.2: Assess a student's performance using Naïve Bayes algorithm with the dataset provided in Table 8.1. Predict whether a student gets a job offer or not in his final year of the course.

Table 8.1: Training Dataset

S.No.	CGPA	Interactiveness	Practical Knowledge	Communication Skills	Job Offer
1.	≥9	Yes	Very good	Good	Yes
2.	≥8	No	Good	Moderate	Yes
3.	≥9	No	Average	Poor	No
4.	<8	No	Average	Good	No
5.	≥8	Yes	Good	Moderate	Yes
6.	≥9	Yes	Good	Moderate	Yes
7.	<8	Yes	Good	Poor	No
8.	≥9	No	Very good	Good	Yes
9.	≥8	Yes	Good	Good	Yes
10.	≥8	Yes	Average	Good	Yes

5. Explain Gibbs Algorithm and Brute force Bayess algorithm

Example 8.3: Given the hypothesis space with 4 hypothesis h_1, h_2, h_3 and h_4. Determine if the patient is diagnosed as COVID positive or COVID negative using Bayes Optimal classifier.

Table 8.12: Posterior Probability Values

$P(h_i \mid T)$	P(COVID Positive $\mid h_i$)	P(COVID Negative $\mid h_i$)
0.3	0	1
0.1	1	0
0.2	1	0
0.1	1	0

Example 8.4: Assess a student's performance using Naïve Bayes algorithm for the continuous attribute. Predict whether a student gets a job offer or not in his final year of the course. The training dataset *T* consists of 10 data instances with attributes such as 'CGPA' and 'Interactiveness' as shown in Table 8.13. The target variable is Job Offer which is classified as Yes or No for a candidate student.

Table 8.13: Training Dataset with Continuous Attribute

S.No.	CGPA	Interactiveness	Job Offer
1.	9.5	Yes	Yes
2.	8.2	No	Yes
3.	9.3	No	No
4.	7.6	No	No
5.	8.4	Yes	Yes
6.	9.1	Yes	Yes
7.	7.5	Yes	No
8.	9.6	No	Yes
9.	8.6	Yes	Yes
10.	8.3	Yes	Yes

6.

7. Compare a biological neuron with an artificial neuron. Explain with diagrams and explain activation functions.

8. Explain the structure and working of a single-layer perceptron. Derive the learning rule, explain weights, activation threshold, and error correction.

Example 10.1: Consider a perceptron to represent the Boolean function AND with the initial weights $w_1 = 0.3$, $w_2 = -0.2$, learning rate $\propto = 0.2$ and bias $\theta = 0.4$ as shown in Figure 10.6. The activation function used here is the Step function $f(x)$ which gives the output value as binary, i.e., 0 or 1. If value of $f(x)$ is greater than or equal to 0, it outputs 1 or else it outputs 0. Design a perceptron that performs the Boolean function AND and update the weights until the Boolean function gives the desired output.

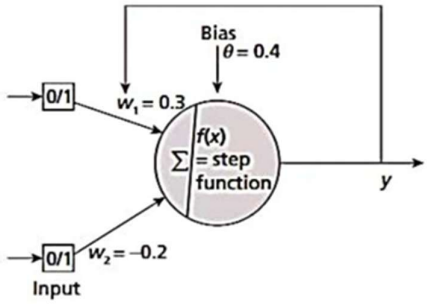


Figure 10.6: Perceptron for Boolean Function AND

9. Differentiate between feedforward and feedback neural networks.
10. List and explain types of Artificial Neural Networks.
11. Discuss the advantages, disadvantages, and challenges of training ANNs.

Module - 05

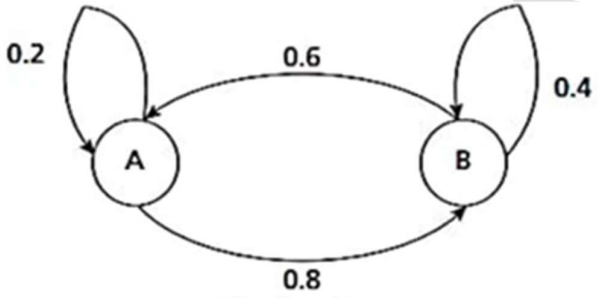
1. Explain the difference between Clustering and classification? Explain its applications, and challenges of clustering algorithms
2. What are proximity measures? Define and compare Euclidean, Manhattan, Cosine similarity, and Jaccard distance with example usage in clustering- 10M

3. Explain the working of K-Means clustering with an example.
   Given a dataset, with initial value of objects2 and 5 considered as initial seeds.

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

4. Describe the DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm.
5. Compare grid-based, density-based, and hierarchical clustering approaches.
   What is Reinforcement Learning (RL)?.
6. Explain the Markov Decision Process (MDP) in the context of RL.

Example 14.2: Assume the universities' hold on a town is initially 60% and 40%. So, what would be the prediction after one month, that is, the hold of university after an year. After two years, will the hold remain same or change? Assume the rest of the probability transition information from



7. What is the Multi-Arm Bandit Problem? Define the exploration-exploitation tradeoff, ε-greedy policy, and real-world applications.
8. Compare Model-Based and Model-Free RL. Explain Q-Learning and SARSA algorithms.