

**FUNDAMENTALS OF AIML
(INF3102)**

Time Allotted : 2½ hrs

Full Marks : 60

Figures out of the right margin indicate full marks.

*Candidates are required to answer Group A and
any 4 (four) from Group B to E, taking one from each group.*

Candidates are required to give answer in their own words as far as practicable.

Group – A

1. Answer any twelve:

12 × 1 = 12

Choose the correct alternative for the following

- (i) Iterative Deepening DFS combines:
 - (a) BFS and DFS
 - (b) BFS and UCS
 - (c) DFS and UCS
 - (d) UCS and Bidirectional Search
- (ii) Implicit state space:
 - (a) Stores all states
 - (b) Generates states on-the-fly
 - (c) Cannot be searched
 - (d) Uses BFS only
- (iii) Bayes rule provides posterior probability of the system under experiment. The posterior probability is obtained as
 - (a) Likelihood probability only
 - (b) Product of likelihood probability and prior probability
 - (c) Prior probability only
 - (d) Likelihood probability prior probability
- (iv) Bias-Variance Trade-off aims to balance:
 - (a) Underfitting and overfitting
 - (b) Training and test data
 - (c) Features and labels
 - (d) Learning rate and loss
- (v) The main goal of Gradient Descent is to:
 - (a) Maximize error
 - (b) Minimize loss function
 - (c) Standardize data
 - (d) Increase learning rate
- (vi) The process of removing branches that contribute little to classification accuracy is called
 - (a) Splitting
 - (b) Pruning
 - (c) Partitioning
 - (d) Branching
- (vii) The "curse of dimensionality" affects KNN because:
 - (a) Distance measures become less meaningful in high dimensions
 - (b) Training time decreases
 - (c) Accuracy increases with dimensions
 - (d) It ignores irrelevant features

- (viii) A confusion matrix compares:
 (a) True values vs. Predicted values (b) Training vs. Test data
 (c) Features vs. Labels (d) Noise vs. Error
- (ix) In Fuzzy C-Means clustering, what does the membership value represent.
 (a) The distance from a data point to the cluster center.
 (b) The likelihood of a data point belonging to a specific cluster.
 (c) The number of clusters the data point belongs to.
 (d) The number of iterations the algorithm needs to converge.
- (x) Which of the following is an advantage of DBSCAN over K-Means.
 (a) DBSCAN requires a pre-defined number of clusters.
 (b) DBSCAN can find clusters of arbitrary shapes.
 (c) DBSCAN is less sensitive to noise and outliers than K-Means.
 (d) Both (b) and (c) are advantages.

Fill in the blanks with the correct word

- (xi) In Uniform Cost Search, the node with the _____ is expanded first.
- (xii) KMeans poorly cluster _____ shaped Dataset
- (xiii) The Naive Bayes classifier is based on the _____ assumption, which assumes that the features are conditionally independent given the class label.
- (xiv) The Perceptron algorithm can only solve problems where the data is _____ separable.
- (xv) A common activation function used in neural networks for binary classification is the _____ function, which outputs values between 0 and 1.

Group - B

2. (a) Define an agent and explain the agent–environment relationship with a diagram.
 (b) Explain the difference between a rational agent and a simple reflex agent with examples.
 (c) List and briefly describe the properties of an environment. [[CO1](Understand/LOCQ)]
- 4 + 4 + 4 = 12**
3. (a) Define state-space search and its components (state, action, transition model, goal test, path cost).
 (b) Formulate the 8-puzzle problem as a state-space search problem. [[CO2](Analyse/IOCQ)]

5 + 7 = 12

Group - C

4. (a) How does the curse of dimensionality affect the performance of k-NN, and what are some techniques to mitigate its impact? [[CO5](Evaluate/HOCQ)]
 (b) You are given the following training dataset with two predictor attributes Age and Income and a class label indicating whether the person has taken loan or not:

Age	Annual Income	Taken Loan
25	40K	No
30	60K	Yes
35	55K	Yes
20	30K	No
40	80K	Yes
45	60K	No
50	50K	No

Predict using 3-Nearest Neighbour Classifier whether a person having Age = 32, Income = 58K has taken loan or not.

[[CO5](Evaluate/HOCQ)]

5 + 7 = 12

5. (a) Briefly explain what is coefficient of determination (R^2) and what is the significance of this measure in the context of linear regression.
- (b) Suppose using linear regression on the following dataset we fitted a line with following equation. Determine Coefficient of Determination (R^2)
 $y = 0.9x + 1.3$

Training Dataset:

x (Independent Variable)	y (Dependent Variable)
1	2
2	3
3	5
4	4
5	6

[[CO4](Evaluate/HOCQ)]

5 + 7 = 12

Group - D

6. (a) Find the Class(X) using Naïve Bayes on the following Dataset, where X = (Age=30; Income=high; Student=No; Credit Rating= Fair)

Age	income	student	credit_rating	buys_computer
< = 30	high	no	fair	no
< = 30	high	no	excellent	no
31 ... 40	high	no	fair	yes
> 40	medium	no	fair	yes
> 40	low	yes	fair	yes
> 40	low	yes	excellent	no
31 ... 40	low	yes	excellent	yes
< = 30	medium	no	fair	no
< = 30	low	yes	fair	yes
> 40	medium	Yes	fair	yes
< = 30	medium	Yes	excellent	yes
31 ... 40	medium	No	excellent	yes
31 ... 40	high	Yes	fair	Yes
> 40	medium	No	excellent	No

- (b) Write the working principle of KNN classification algorithm. [[CO5)(Understand/LOCQ]]

7 + 5 = 12

7. (a) Consider the following table with predicted results obtained after applying 2 classification model on a dataset, having two classes malignant (positive class) and benign (negative class). Calculate the TP, FP, TN, and FN of each model and next Compare the sensitivity of the two models and determine whose performance is better.

	Model 1 (Predicted Class)	Model 2 (Predicted Class)	Actual Class
Sample 1	+	+	+
Sample 2	-	+	+
Sample 3	-	-	+
Sample 4	+	-	+
Sample 5	+	-	-
Sample 6	-	+	-
Sample 7	+	-	-

[[CO4)(Evaluate/HOCQ]]

- (b) Justify with example for or against the statement
 “The concept of neighbors in higher dimension does not exist” [[CO4)(Analyse/IOCQ]]

7 + 5 = 12

Group - E

8. (a) Create clusters of the following spatial data objects using DBSCAN with minpts = 3 and Epsilon = 3.2

01	4	6
02	7	4
03	3	3
04	5	5
05	4	6
06	5	7
07	6	2
08	6	6

[[CO1,CO3,CO6)(Create/HOCQ]]

- (b) Discuss the advantages of DBSCAN over kmean's algorithm. [[CO6)(Remember/LOCQ]]

8 + 4 = 12

9. (a) Explain how the membership matrix in fuzzy C means clustering algorithm gets updated. [[CO6)(Remember/LOCQ]]

- (b) Write the back propagation based neural network algorithm. Also show the working of the same for designing the 2 i/p 1o/p XOR network. [[CO4)(Apply/IOCQ]]

3 + (5 + 4) = 12

Cognition Level	LOCQ	IOCQ	HOCQ
Percentage distribution	25	41.66	33.33