PATTERN RECOGNITION
(CSEN 4233)

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) When you find noise in data, which of the following option would you consider in k-NN classification?
(a) Value of k can be increased (b) Value of k can be decreased
(c) Noise cannot be dependent on value of k (d) None of (a), (b) & (c).

(ii) Which of the following measure is not a metric?
(a) Euclidean Distance (b) Cosine Similarity
(c) KL divergence (d) None of these

(iii) The numerical output of a sigmoid node in a neural network
(a) Is unbounded, encompassing all real numbers
(b) Is unbounded, encompassing all integers
(c) Is bounded between 0 and 1
(d) Is bounded between -1 and 1.

(iv) Suppose that X_1, ..., X_m are categorical input attributes and Y is categorical output attribute. Suppose we plan to learn a decision tree without pruning, using the standard algorithm. The maximum depth of the decision tree must be
(a) less than m+1 (b) greater than m+1
(c) both (a) and (b) can be true (d) Neither (a) nor (b) is true

(v) Which of the following is limitation of Expectation Maximization Clustering?
(a) Flexible Cluster Shapes (b) Soft Assignment
(c) Handles Overlapping Data (d) Computational Complexity

(vi) Hierarchical agglomerative based clustering is a
(a) bottom up approach (b) top down approach
(c) both (a) and (b) (d) none of the above.

(vii) DBSCAN cannot be used (with high accuracy) for datasets that are
(a) Convex (b) Uniform density
(c) Non-uniform density (d) None of (a), (b) & (c)

(viii) When performing regression or classification, which of the following is the correct way to pre-process the data?
(a) Normalize the data → PCA → training
(b) PCA → normalize PCA output → training
(c) Normalize the data → PCA → normalize PCA output → training
(d) None of the above.

(ix) If the variance covariance matrix is diagonal matrix then the features are_____
(a) independent (b) dependent
(c) zero mean value (d) uniformly distributed.

(x) Consider the following two statements:
Statement 1: Independent Component Analysis is an unsupervised method
Statement 2: Independent Component Analysis works very well with Gaussian data distribution
(a) Only Statement 1 is correct (b) Only Statement 2 is correct
(c) Both statements are incorrect (d) Both statements are correct

Fill in the blanks with the correct word

(xi) In a decision tree node for binary class, the probability of one class is 0.75. The entropy of the node is _____.

(xii) The null hypothesis in the Chi-Square test for feature selection states that the feature and the target variable are _____ with respect to each other

(xiii) Fisher's linear discriminant analysis is used in _____.

(xiv) When two classes can be separated by a separate line, they are known as_____

(xv) In L1 regularization (Lasso), the coefficients of irrelevant features will shrunk to _____.

Group - B

2. (a) 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. *[(CO1)(CO2)(CO3)(CO5)Apply/IOCQ)]*

(b) What is the basic principle behind the Minimum Distance Classifier? How does Minimum Distance Classifier assume data is distributed in feature space?
[(CO1)(CO2)(CO3)(CO5)Remember/LOCQ)]

6 + 6= 12

3. (a) Prove that, in high dimensions, most points in a dataset become nearly equidistant from each other. *[(CO1)(CO2)(CO3)(CO5)Analyse/HOCQ)]*

(b) Assume a scenario where a very limited amount training data, but with well-understood data distribution is available for training. In such case, which approach (parametric/nonparametric) would you choose, and why? *[(CO1)(CO2)(CO3)(CO5)Analyze/IOCQ)]*

(c) How does the k-Nearest Neighbors (k-NN) algorithm exemplify nonparametric learning? *[(CO1)(CO2)(CO3)(CO5)/ Understand/LOCQ)]*

4 + 4 + 4 = 12

Group - C

4. (a) Consider the following Table 1 containing dataset, where the weekend activity of a student has been given for 10 consecutive weeks.

Week#	Weather type	Humidity	Pocket Money	Weekend Activity
Week1	Hot	High	500	Movie
Week2	Cold	Low	2000	Shopping
Week3	Rainy	Low	1500	Restaurant
Week4	Rainy	High	500	Movie
Week5	Hot	Low	2000	Restaurant
Week6	Cold	High	1500	Shopping
Week7	Hot	Low	2000	Shopping
Week8	Cold	Low	500	Restaurant
Week9	Cold	High	2000	Shopping
Week10	Rainy	High	500	Movie

The weekend activity (i.e., Class/ Label) of the given dataset can depend only on some or all of the following three predictors of the table: Weather type, Humidity, and pocket money.

Show all steps to determine which of these predictors will be used for the root node of each of the decision trees, constructed using the following algorithms, to predict the weekend activity:

(i) Using ID3 algorithm, based on Information Gain using GINI impurity index

(ii) Using ID3 algorithm, based on Information Gain using Entropy. *[(CO1)(CO2)(CO3)(CO5) (Apply/IOCQ)]*

(b) Briefly discuss about Gini Impurity function? *[(CO1)(CO2)(CO3)(CO5) (Understand/LOCQ)]*

(5 + 5) + 2 = 12

5. (a) Derive the weight update equation in MLP (Multi layered Perceptron) Classifier using sigmoid function as the activation function for classifying binary-class dataset. *[(CO1)(CO2)(CO3)(CO5))(Derive/IOCQ)]*

(b) Discuss why stochastic gradient prevents over-fitting of data and regularises well in learning a multi layer perceptron model? *[(CO1)(CO2)(CO3)(CO5))(Analyse/HOCQ)]*

(c) What is vanishing and exploding gradient problem. *[(CO1)(CO2)(CO3)(CO5))(Analyze/IOCQ)]*

6 + 3 + 3 = 12

Group - D

6. (a) For the given data, compute two clusters using K-medoids algorithm for clustering where k=2 and the initial cluster centers are (1.0, 1.0) and (5.0, 7.0). Execute for two iterations considering that distance between two points is defined by Manhattan distance (i.e., L1-norm).

Record Number	A	B
R1	1.0	1.0
R2	1.5	2.0
R3	3.0	4.0
R4	5.0	7.0
R5	3.5	5.0
R6	4.5	5.0
R7	3.5	4.5

[(CO1)(CO2)(CO4)(CO5))/Apply/IOCQ)]

(b) What are the major drawback of k-means algorithm?

[(CO1)(CO2)(CO4)(CO5)/Remember/LOCQ)]

9 + 3 = 12

7. (a) Explain the Fuzzy C-means clustering algorithm.

[(CO1)(CO2)(CO4)(CO5))/Understand/IOCQ]

(b) What is the key difference between Fuzzy C-Means (FCM) and K-means clustering? *[(CO1)(CO2)(CO4)(CO5))/Understand/LOCQ]*

(c) Explain the role of the fuzziness parameter (m) in FCM.

[(CO1)(CO2)(CO4)(CO5))/Understand/ IOCQ)]

6 + 3 + 3 = 12

Group - E

8. (a) What is the primary goal of dimension reduction in pattern recognition? How does it improve classifier performance? *[(CO1)(CO2)(CO5)(CO6))/Understand/IOCQ)]*

(b) How does PCA relate to the singular value decomposition (SVD)?

[(CO1)(CO2)(CO5)(CO6))/Analyze/IOCQ)]

(c) Why is PCA considered an "unsupervised" technique?

[(CO1)(CO2)(CO5)(CO6))/Analyze/IOCQ)]]

3 + 6 + 3 = 12

9. (a) Explain how Convolution Neural Network perform feature extraction

[(CO1)(CO2)(CO5)(CO6))Apply/LOCQ)]

(b) How Linear Discriminant Analysis can be used in dimension reduction

[(CO1)(CO2)(CO5)(CO6))Understand/LOCQ)]

7 + 5 = 12

Cognition Level	LOCQ	IOCQ	HOCQ
Percentage distribution	31	62	7