Printed Pages – 6 Roll No. 300012821042.

# C128571(022)

## B. Tech. (Hon's) (Fifth Semester) Examination, Nov.-Dec. 2023

### (AICTE Scheme)

### (Data Science Engg. Branch)

### PATTERN RECOGNITION and MACHINE LEARNING

*Time Allowed : Three hours*

*Maximum Marks : 100*

*Minimum Pass Marks : 35*

*Note : Each question contains four parts. Part (a) of each question is compulsory. Attempt any two parts from (b), (c) and (d) of each question. The figure in the right-hand margin indicates marks.*

### Unit-I

1. (a) What is Learning? Design a learning system for checkers problem. 4

   (b) What do you mean by loss function machine learning? Explain loss function in linear regression. 8

(c) State Bayes theorem. Consider a scenario where your yearly checkup is done, the doctor has bad news and good news. The bad news is that you tested positive for a serious disease, and that the test is 99% accurate (i.e., the probability of testing positive given that you have the disease is 0·99, as it is the probability of testing negative given that you have the disease). This is a rare disease, striking only one in 10,000 people. Why is it good news that the disease is rare? What are the chances that you actually have the disease? 8

(d) Define multi-dimensional space. Explain Normed vector and dot product space in brief. 8

## Unit-II

2. (a) Explain Maximum entropy estimation in brief. 4

(b) Suppose we wish to include the continuous-valued attribute temperature in describing the training example days in the learning task of following table. Incorporate this continuous value to discrete value for constructing decision tree. 8

| F | f1 | f2 | f3 | f4 | f5 | f6 |
|---|---|---|---|---|---|---|
| Temperature | 40 | 48 | 60 | 72 | 80 | 90 |
| Play Tannis | No | No | Yes | Yes | Yes | No |





(c) Explain maximum likelihood estimation? Consider an unfair coin is flipped 100 times. 61 head are observed. The coin either has a probability 1/3, 1/2 and 2/3 of flipping a head each time. Find which of the three is MLE? 8

(d) What is Mutual information estimation? Explain Akaike and Bayesian information criterion. 8

## Unit-III

3. (a) Briefly explain discriminant function. 4

(b) Suppose that the initial seeds are $A1$, $A4$ and $A7$. Run the k-means clustering for following 8 examples. Iterate this algorithm till convergence and mention centroid with cluster in each iteration. 8

$A1 = (2, 10)$, $A2 = (2, 5)$, $A3 = (8, 4)$, $A4 = (5, 8)$, $A5 = (7, 5)$, $A6 = (6, 4)$, $A7 = (1, 2)$, $A8 = (4, 9)$

(c) Explain Fishers linear discrimination brief. Discuss its merits over PCA. 8

(d) Explain Probabilistic Discriminative Models. Write down difference between Generative and Discriminative probabilistic model. 8

## Unit-IV

4. (a) Briefly explain semi supervised algorithm with suitable example. 4

(b) Consider following data sample on effect of hours of mixing on temperature of wood pulp. Predict temperature of wood pulp for hour ($x$) = 13. 8

| S. No. | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| Hours of mixing ($X$) | 2 | 4 | 6 | 8 | 10 | 12 |
| Temperature of wood pulp (Y) | 21 | 27 | 29 | 64 | 86 | 92 |

(c) Consider following data sample with two cluster. Implement fuzzy c means clustering. Iterate this algorithm for at least two iteration. 8

| | (1, 3) | (2, 5) | (4, 8) | (7, 9) |
|---|---|---|---|---|
| Cluster 1 | 0·8 | 0·7 | 0·2 | 0·1 |
| Cluster 2 | 0·2 | 0·3 | 0·8 | 0·9 |

(d) What is rough k-means? Explain its algorithm. 8

## Unit-V

5. (a) What is two class problem? Explain different method to solve this problem in brief. 4



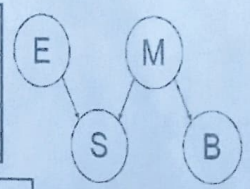


(b) What is Support Vector Machine? Derive an expression for maximum marginal hyperplane. 8

(c) What is Kernal function? Explain the linear and Fisher Kernal methods in machine learning. 8

(d) Explain conditional independence? Consider a scenario, where a smell of Sulphur (S) can be caused either by rotten eggs (E) or as a sign of the doom brought by the Mayan Apocalypse (M). The Mayan Apocalypse also causes the oceans to boil (B). The Bayesian network and corresponding conditional probability tables for this situation are shown below. 8

| P (E) | |
|---|---|
| $+e$ | 0·4 |
| $-e$ | 0·6 |




| P (M) | |
|---|---|
| $+m$ | 0·1 |
| $-m$ | 0·9 |

| $P(S \mid E, M)$ | | | |
|---|---|---|---|
| $+e$ | $+m$ | $+s$ | 1·0 |
| $+e$ | $+m$ | $-s$ | 0·0 |
| $+e$ | $-m$ | $+s$ | 0·8 |
| $+e$ | $-m$ | $-s$ | 0·2 |
| $-e$ | $+m$ | $+s$ | 0·3 |
| $-e$ | $+m$ | $-s$ | 0·7 |
| $-e$ | $-m$ | $+s$ | 0·1 |
| $-e$ | $-m$ | $-s$ | 0·9 |

| $P(B \mid M)$ | | |
|---|---|---|
| $+m$ | $+b$ | 1·0 |
| $+m$ | $-b$ | 0·0 |
| $-m$ | $+b$ | 0·1 |
| $-m$ | $-b$ | 0·9 |

Then find :

(i) What is the probability that the oceans boil?

(ii) What is the probability that the Mayan Apocalypse is occurring, given that there is a smell of Sulphur, the oceans are boiling, and there are rotten eggs?

(iii) What is the probability that the Mayan Apocalypse is occurring, given that the oceans are boiling?

Printed Pages – 4 Roll No. : 300012821042

# C128572(022)

## B. Tech. (Hon's) (Fifth Semester) Examination
## Nov.-Dec. 2023

## (AICTE Scheme)

## (Data Science Engg. Branch)

## INTELLIGENT DATA ANALYSIS

***Time Allowed : Three hours***

***Maximum Marks : 100***

***Minimum Pass Marks : 35***

***Note :*** *Attempt all questions. Part (a) from each question is compulsory and answers any* ***two*** *of the remaining (b), (c) and (d).*

### Unit-I

1. (a) What is Data Mining? Explain in brief. 4

   (b) What do you mean by data? Explain various types of data used in machine learning. 8

(c) Why central tendency is an important measure in data analysis? Explain all the measure under central tendency calculation. 8

(d) What is data visualization? Explain different techniques used for data visualization. 8

## Unit-II

2. (a) What do you mean by pre-processing of data. Explain in brief. 4

(b) How to measure Data Similarity and Dissimilarity. Explain different tools to measure similarity and dissimilarity. 8

(c) What is data transformation? Explain different techniques of data transformation. 8

(d) How artificial neural network (ANN) is different from biological neural network? List different types of ANN model. 8

## Unit-III

3. (a) Briefly explain association analysis in data mining. 4





(b) Explain FP growth algorithm in brief. 8

(c) Consider following data sample. Execute k-nearest neighbour algorithm to classify data sample. Classify a new entry where brightness is 20 and saturation is 35. 8

| S. No. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| Brightness | 40 | 50 | 60 | 10 | 70 | 60 | 25 |
| Students | 20 | 50 | 90 | 25 | 70 | 10 | 80 |
| Class | Red | Blue | Blue | Red | Blue | Red | Blue |

(d) Explain Maximal Frequent Itemset and Closed Frequent Itemset. 8

## Unit-IV

4. (a) What is cluster? Explain different technique used to make cluster. 4

(b) Consider following data and execute DBSCAN algorithm. Assume = 3.5 and MinPts = 3. Find core, boundary and noise point. 8

| Feature | D1 | D2 | D3 | D4 | D5 | D6 | D7 | D8 |
|---|---|---|---|---|---|---|---|---|
| X | 5 | 8 | 3 | 4 | 3 | 6 | 6 | 5 |
| Y | 7 | 4 | 3 | 4 | 7 | 7 | 1 | 5 |

(c) Discuss issue associated with k-means algorithms. Explain different approach to improve the performance of k-means algorithm. 8

(d) Differentiate between agglomerative and Divisive Hierarchical clustering algorithm. 8

## Unit-V

5. (a) What do you mean by anomalies data mining? Describe different types of anomalies. 4

(b) Discuss different types of anomalies detection techniques. 8

(c) Explain classification-based anomalies detection technique with suitable example. 8

(d) What are the different challenges in anomalies detection? Explain in brief. 8

Printed Pages– 4 Roll No. 300012821042

# C127573(022)

## B. Tech. (Hon's) (Fifth Semester) Examination, Nov.-Dec. 2023

### (Computer Science and Engg. Branch - Artificial Intelligence)

### CRYPTOGRAPHY and NETWORK SECURITY

*Time Allowed : Three hours*

*Maximum Marks : 100*

*Minimum Pass Marks : 35*

*Note : All questions are compulsory. Part (a) of each unit is compulsory and carries 4 marks. Attempt any **two** parts from (b), (c) and (d) and carries 8 marks each.*

### Unit-I

1. (a) Define symmetric cipher model. 4

(b) Explain how steganography can be used to enhance security in communication. 8

(c) Discuss the Limitations of Perfect Secrecy. 8

(d) Describe Shannon's theorem and its significance in cryptography. 8

**Unit-II**

2. (a) What is modular arithmetic and give a simple example? 4

(b) How does prime factorization underpin the security of modern cryptographic methods? 8

(c) Explain discrete logarithms and their applications in cryptography. 8

(d) Describe computations in finite fields and their relevance to cryptography. 8

**Unit-III**

3. (a) Briefly explain what a pseudorandom function is. 4

(b) Summarize the DES encryption process. 8





(c) Describe different modes of operation in block ciphers. 8

(d) Discuss the vulnerabilities of DES and methods to increase its security. 8

**Unit-IV**

4. (a) Define public-key cryptography. 4

(b) Outline the Diffie-Hellman Key Agreement process. 8

(c) Explain the RSA algorithm and discuss its security. 8

(d) Compare and contrast private and public-key encryption. 8

**Unit-V**

5. (a) What is a hash function? 4

(b) Explain the concept of Collision-Resistant Hash Functions. 8

(c) Describe the role of Secure Message Authenticate Codes in network security. 8

(d) Discuss the SHA-512 hash algorithm and its advantages over its predecessors. 8

Printed Pages – 3 Roll No. : 3000128210472

# C128574(022)

## B. Tech. (Hon's) (Fifth Semester) Examination

## Nov.-Dec. 2023

## (AICTE Scheme)

## (Data Science/Artificial Intelligence)

## NATURAL LANGUAGE PROCESSING

***Time Allowed : Three hours***

***Maximum Marks : 100***

***Minimum Pass Marks : 35***

***Note :*** *Attempt all questions. Part (a) from each question is compulsory each question 4 marks. answers any* **two** *of the remaining (b), (c) and (d) each question 8 marks..*

### Unit-I

1. (a) Define Named Entity Recognition. 4

   (b) Explain different levels of Natural Language Processing. 8

(c) What is stemminrg in NLP and how it is different from lemmatization? 8

(d) Explain the challenges of NLP. 8

**Unit-II**

2. (a) Define Derivational Morphology. 4

(b) Explain the process of dealinrg with various spelling errors. 8

(c) Explain major phonetics classes. 8

(d) Explain N Gram Model. 8

**Unit-III**

3. (a) Define parts of speech. 4

(b) Explain Tree Bank and different types oef Tree Bank. 8

(c) Explain the parsinrg approach and its type. 8

(d) Explain scope ambiguity and attachment ambiguity resolution. 8





**Unit-IV**

4. (a) What do you mean by Dialogue Analysis? 4

(b) Differentiate information extraction and information retrieval. 8

(c) Write approaches to discourse analysis and its example. 8

(d) Explain Anaphora Resolution. 8

**Unit-V**

5. (a) Define Shallow Parsing. 4

(b) Explain application of Question Answering based system. 8

(c) Explain process of machine translation. 8

(d) Explain application of sentiment analysis. 8

Printed Pages – 6 Roll No. : 30901282042

# C127532(022)

## B. Tech. (Hon's) (Fifth Semester) Examination, Nov.-Dec. 2023

### (Artificial Intelligence)

### COMPUTATIONAL COMPLEXITY

*Time Allowed : Three hours*

*Maximum Marks : 100*

*Minimum Pass Marks : 40*

***Note :*** *Attempt all questions. All question carries equal marks.*

### Unit-I

1. (a) Explain the concept of polynomial time in the context of computational complexity. 4

   (b) Write a non-deterministic algorithm to search an element from a given set of elements. 8

(c) Discuss the concept of reduction in the context of NP-completeness. Explain how a problem is reduced to another problem and how this relates to NP-complete problems? 8

## Unit-II

2. (a) Consider a scenario where you have a knapsack with a maximum weight capacity of $W$ and a set of $N$ items, each with a weight ($w[i]$) and a value ($v[i]$). You are asked to maximize the total value of items that you can place in the knapsack without exceeding its weight capacity. 4

(i) Write the dynamic programming algorithm to solve the 0/1 knapsack.

(ii) Provide the time and space complexity of your algorithm.

(b) Calculate the minimum no. of multiplication and placing of parenthesis for the given chain matrix multiplication. $A1 = 2 \times 4$, $A2 = 4 \times 6$, $A3 = 6 \times 7$, $A4 = 7 \times 8$. 8





(c) Write Huffman code algorithm and solve the given problem $A : 20$, $B : 13$, $C : 45$, $D : 34$, $E : 16$, $F : 27$, $G : 19$. 8

## Unit-III

3. (a) Define Finger printing algorithm with an example. 4

(b) (i) Define the concept of randomized algorithms and explain why they are used in computational problems.

(ii) Describe at least two de-randomization techniques, such as the method of conditional probabilities and pseudo random generators. Provide a step-by-step explanation of how these techniques work.

(iii) Discuss the advantages and limitations of de-randomization in the context of algorithm design and analysis. 8

(c) Briefly describe about Algebraic methods with examples. 8

## Unit-IV

4. (a) Consider a weighted, directed graph $G$ with non-negative edge weights. Explain the following aspects of shortest path algorithms : 4

(i) Define and explain the concept of a "shortest path" in a graph.

(ii) Provide a brief overview of Dijkstra's algorithm, including its main idea and conditions under which it can be applied.

(iii) Mention one limitation or scenario where Dijkstra's algorithm may not be the most suitable choice for finding the shortest path.

(b) Consider a connected, undirected graph with the following edge weights : 8

(1, 2) : 3

(1, 3) : 2

(1, 4) : 1

(2, 3) : 4

(2, 4) : 5

(3, 4) : 6





(i) Apply Kruskal's algorithm to find the minimum spanning tree (MST) of the given graph. Show each step, including the edges considered, and the evolving MST.

(ii) Apply Prim's algorithm to find the minimum spanning tree (MST) of the given graph, starting from vertex 1. Show each step, including the vertices added and the evolving MST.

(c) Consider a flow network with the following capacities : Source

(S) $\rightarrow$ A : 10

Source

(S) $\rightarrow$ B : 15

A $\rightarrow$ C : 20

A $\rightarrow$ D : 25

B $\rightarrow$ C : 5

B $\rightarrow$ D : 10

C $\rightarrow$ Sink (T) : 30

D $\rightarrow$ Sink (T) : 20

(i) Apply Ford-Fulkerson algorithm (using the Edmonds-Karp implementation) to find the maximum flow in the given flow network. Show

each step, including the residual graph and the evolving flow.

(ii) Discuss the concept of minimum cut in the context of flow networks. Identify the minimum cut in the provided flow network and explain its significance. 8

## Unit-V

5. (a) Explain the concept of decision trees in machine learning. 4

(b) Illustrate the Red Black Tree property and using the following elements create a Red Black Tree. 8

(c) Explain the concept of Fibonacci Heaps in detail.

(i) Define the structure and properties of Fibonacci Heaps, including the concept of nodes, key values, and the potential degrees of nodes.

(ii) Describe the Fibonacci Heap operations, specifically the processes of insertion, union, decrease key, and extract minimum. Provide a step-by-step walk through for each operation. 8