DAA questions

What is the Fibonacci sequence of numbers?

The Fibonacci sequence is a series of numbers in which each number is the sum of the two numbers before it.

It starts like this:

👉 0, 1, 1, 2, 3, 5, 8, 13, 21, 34, …

- **Formula:**

$$F(n) = F(n-1) + F(n-2)$$

What is the Golden Ratio?

The Golden Ratio (also called the Divine Proportion) is a special number that appears in nature, art, architecture, and mathematics — and it's closely connected to the Fibonacci sequence.

- **Definition**

The Golden Ratio is usually written as the Greek letter ϕ (phi) and equals approximately:

$$\phi = 1.6180339887\ldots$$

It's an irrational number, meaning it goes on forever without repeating.

What is the real application for Fibonacci series

Flower petals often follow Fibonacci numbers (e.g., lilies have 3 petals, daisies 34, sunflowers 55 or 89).

🌲 Tree branching and leaf arrangements (phyllotaxis) follow Fibonacci spirals for efficient sunlight exposure.

🐚 Shells and hurricanes form logarithmic spirals based on Fibonacci proportions.

what is recursive and non recursive program

1. Recursive Program

A recursive program is one that calls itself to solve smaller parts of the same problem.

✅ **Advantages:**

- **Code is simple and easy to understand.**
- **Good for problems naturally divided into subproblems (like trees, Fibonacci, factorial).**

⚠️ **Disadvantages:**

- **Slower due to many function calls.**
- **Uses more memory (stack space).**

2. Non-Recursive (Iterative) Program

A non-recursive program (or iterative) uses loops (for/while) instead of self-calls.

✅ **Advantages:**

- **Faster execution.**
- **Uses less memory.**

⚠️ **Disadvantages:**

- **Code can be longer and less elegant.**

What is Huffman Encoding?

Huffman Encoding is a data compression algorithm used to reduce the size of data without losing any information (called lossless compression).

It works by assigning shorter binary codes to frequent characters and longer codes to less frequent ones — so the overall size of the encoded message becomes smaller.

Real-World Example

Used in:

- 📁 **ZIP / RAR file compression**
- 🖼️ **JPEG / PNG images**
- 📜 **Text and communication protocols**

Which tree is used in Huffman encoding?Give one Example

The **Huffman Tree** is a special kind of **binary tree** —
specifically a **full binary tree**, meaning every internal node has **exactly two children**.

Each **leaf node** represents a **character** from the message,
and the **path from root to leaf** gives that character's **binary code**.

Why Huffman coding is lossless compression?

Huffman coding replaces each character with a **unique binary code** (shorter for frequent symbols, longer for rare ones).

Because these binary codes are designed carefully, the encoded data can always be **perfectly reversed**.

What is a Greedy Strategy?

A Greedy Strategy (or Greedy Algorithm) is a problem-solving approach that makes the best possible choice at each step, hoping that these local (immediate) choices will lead to the global (overall) optimal solution.

- **Key Idea:**

At every step:

1. **Choose the option that looks best at the moment (locally optimal choice).**
2. **Never reconsider previous choices.**
3. **Continue until the entire problem is solved.**

Explain concept of fractional knapsack

The Fractional Knapsack Problem is an optimization problem in which a set of items, each having a value and a weight, must be placed into a knapsack of limited capacity so as to maximize total value.

in the **Fractional Knapsack**, you can **take fractions of items** (not necessarily the whole item).

- **Example**

Item	Value (₹)	Weight (kg)
1	60	10
2	100	20
3	120	30

Knapsack capacity = **W = 50 kg**

- **Step 1: Compute value/weight ratio**

Item	Value	Weight	Value/Weight
1	60	10	6
2	100	20	5

Item	Value	Weight	Value/Weight
3	120	30	4

🔹 **Step 2: Sort items by value/weight ratio (descending order)**

Highest ratio first → take items that give **maximum profit per kg**.

Order:
Item 1 (6) → Item 2 (5) → Item 3 (4)

🔹 **Step 3: Fill the knapsack greedily**

Step	Item	Weight Taken	Value Gained	Remaining Capacity
1	1	10	60	40
2	2	20	100	20
3	3	20 (partial)	(20/30)*120 = 80	0

✅ **Total Value = 60 + 100 + 80 = ₹240**
Knapsack full at 50 kg.

Explain concept of 0/1 knapsack

The 0/1 Knapsack Problem is an optimization problem where you have a set of items, each with a weight and a value, and a knapsack that can carry a limited maximum weight (capacity).

You must select items to maximize the total value — but you can either:

- **Take the whole item (1), or**
- **Leave it (0)**
  — you cannot take fractions of an item.

What is Dynamic Approach?

The Dynamic Approach, also called Dynamic Programming (DP), is a problem-solving technique used to solve complex problems by breaking them into smaller overlapping subproblems, solving each subproblem only once, and storing their results to avoid repeated work.

Explain Branch and Bound Approach. Which is best?

Branch and Bound is a search-based optimization technique that divides the problem into smaller parts (branches) and uses bounds to eliminate (or prune) parts of the search space that cannot lead to the optimal solution.

What is Quick sort

Quick Sort is a divide-and-conquer sorting algorithm that works by selecting a pivot element, partitioning the array around that pivot, and then recursively sorting the left and right subarrays.

It is one of the fastest and most widely used sorting algorithms in practice.

- **Basic Idea:**
  1. **Pick an element from the array — called the pivot.**
  2. **Rearrange (partition) the array so that:**
     - **Elements smaller than the pivot are on the left,**
     - **Elements greater than the pivot are on the right.**
  3. **Recursively apply the same process to the left and right parts.**

- **Example:**

Input:

[10, 7, 8, 9, 1, 5]

Step 1:

Pick pivot = 5
Partition → [1, 5, 8, 9, 10, 7]
(after placing all smaller elements before 5)

Step 2:

Now sort left side [1] and right side [8, 9, 10, 7] recursively.

Continue until every subarray has 1 element.

✅ **Final sorted array: [1, 5, 7, 8, 9, 10]**

what is deterministic and randomized variant.

Deterministic Quick Sort

➤ Definition:

In the deterministic variant, the pivot is chosen in a fixed (predefined) way, such as:

- Always the first element
- Always the last element
- Always the middle element

Hence, it's deterministic — the pivot choice does not change between runs for the same input.

🔹 Example:

Array: [10, 7, 8, 9, 1, 5]
If we always choose the last element as pivot (5),
the result will be same every time for this input.

🔹 Pros:

✅ Simple and predictable.

✅ Easy to implement.

🔹 Cons:

❌ If the array is already sorted or nearly sorted,
the pivot may divide the array unevenly → worst case $O(n^2)$ time.

Randomized Quick Sort

➤ Definition:

In the randomized variant, the pivot is chosen randomly from the array before partitioning.

This means each time the algorithm runs, the pivot may be different, even for the same input.

🔹 Example:

Array: [10, 7, 8, 9, 1, 5]
Randomly pick pivot = 8 → partition.
Next run, pivot might be 1 → different partitioning.

🔹 Pros:

✅ Reduces the chance of hitting worst case ($O(n^2)$).

✅ Average performance stays O(n log n) regardless of input order.

✅ More stable and efficient for real-world data.

🔹 Cons:

❌ Requires generating random numbers.

❌ Slightly more code complexity.

ML Questions

What is data preprocessing?

Data preprocessing is the process of cleaning, transforming, and organizing raw data before using it for analysis or machine learning.

The main goal is to make the data accurate, consistent, and ready for modeling**.**

Define Outliers?

An outlier is a data point that is significantly different from the other values in a dataset.

It lies far away from the normal range of data and can distort analysis or model accuracy

- Example:

Suppose you have the following ages of people in a group:
[22, 24, 27, 25, 23, 120]

Here, 120 is an outlier, because it's unusually large compared to the rest (most are between 20–30).

What is Linear Regression?

Linear Regression is a supervised learning algorithm used to predict a continuous value based on one or more input (independent) variables.

It finds a straight-line relationship between the input variables (X) and the output variable (Y).

What is Random Forest Algorithm?

Random Forest is an ensemble algorithm that creates multiple decision trees and aggregates their results to enhance prediction accuracy and reduce overfitting

Explain: pandas, numpy.

• Pandas: A Python library for data manipulation and analysis, providing tools to work with data in tables.

• NumPy: A library for numerical computing in Python, used for working with arrays and performing mathematical operations.

Data Preprocessing

Data preprocessing involves cleaning, transforming, and organizing raw data to make it suitable for analysis or machine learning models.

Binary Classification

Binary classification is a type of classification where there are only two possible outcomes, such as "yes" or "no," "true" or "false."

explain K-Nearest Neighbours (K-NN)

K-Nearest Neighbours (K-NN) is a supervised learning algorithm used for classification and regression.
It's a **simple yet powerful** method that makes predictions based on the **similarity** between data points.

K-NN assumes that **similar things exist close to each other**.
So, when you want to predict the class (or value) of a new data point, the algorithm looks at its **K nearest neighbors** in the training data and uses **their labels** to decide the output.

Support Vector Machine (SVM)

SVM is a supervised learning algorithm that finds the best boundary (hyperplane) to separate data into different classes with maximum margin.

Train, Test, and Split Procedure

This process divides a dataset into training and testing sets, where the training set trains the model, and the test set evaluates its performance.

What is Normalization?

Normalization is a **data preprocessing technique** used to **adjust the scale of features** so that they all contribute equally to a machine learning model.

In simple terms —

Normalization means **bringing all feature values to a common scale** (usually between 0 and 1) **without distorting differences** in the data.

What is Standardization?

Standardization (also called **Z-score normalization**) is a data preprocessing technique used to **rescale features** so that they have:

- **Mean (μ) = 0**
- **Standard deviation (σ) = 1**

In simple words — it **centers** the data around zero and gives it a **unit variance**, making different features **comparable in scale**.

Confusion Matrix

A confusion matrix is a table that shows the performance of a classification model by comparing actual vs. predicted values for each class, helping to evaluate accuracy, precision, and recall.

One Example of Confusion Matrix?

📊 **Example Scenario**

Suppose you built a model to detect whether an email is **Spam** or **Not Spam**.
You test it on **10 emails**, and here are the results:

Email	Actual	Predicted
1	Spam	Spam
2	Spam	Spam
3	Spam	Not Spam
4	Not Spam	Spam
5	Not Spam	Not Spam
6	Spam	Spam
7	Not Spam	Not Spam
8	Not Spam	Spam
9	Not Spam	Not Spam
10	Spam	Not Spam

🧩 **Construct the Confusion Matrix**

	Predicted: Spam	**Predicted: Not Spam**
Actual: Spam	**True Positive (TP) = 3**	**False Negative (FN) = 2**
Actual: Not Spam	**False Positive (FP) = 2**	**True Negative (TN) = 3**

What is the basic principle of a Support Vector Machine?

A Support Vector Machine (SVM) is a supervised machine learning algorithm used for classification and regression, but it's most commonly used for classification tasks.

🧠 **Core Idea / Basic Principle**

The basic principle of SVM is to find the best separating boundary (called a hyperplane) that maximally separates the data points of different classes.

In simple words:

SVM tries to draw the widest possible line (or margin) between two classes.

What is k-Means clustering/ hierarchical clustering

1️⃣ **K-Means Clustering**

⚙️ **Definition:**

K-Means is an unsupervised learning algorithm used to group data into K distinct clusters based on feature similarity.

It tries to minimize the distance between data points and their cluster centers (centroids).

2️⃣ **Hierarchical Clustering**

⚙️ **Definition:**

Hierarchical Clustering builds a hierarchy of clusters — either from bottom to top or top to bottom.

It doesn't need you to pre-specify the number of clusters.

What is the Elbow Method?

The Elbow Method is a technique used to find the optimal number of clusters (K) in K-Means clustering.

The Elbow Method helps you find a value of K where adding more clusters doesn’t significantly improve performance.