

Q1. What is a parameter?

A **parameter** is a value inside a model that the model learns from training data.
Example: weights and bias in linear regression.

Q2. What is correlation?

Correlation measures the relationship between two variables—how one variable changes when another changes.

Q3. What does negative correlation mean?

Negative correlation means when one variable increases, the other decreases.
Example: As price increases, demand decreases.

Q4. Define Machine Learning. What are the main components in Machine Learning?

Machine Learning (ML) is a branch of Artificial Intelligence where machines learn from data and improve performance without being explicitly programmed.

Main components of Machine Learning:

1. Data
2. Model
3. Features
4. Algorithm
5. Loss function
6. Training process

Q5. How does loss value help in determining whether the model is good or not?

The **loss value** shows how far the model's predictions are from actual values.

- **Low loss** → Model is good
- **High loss** → Model is poor

Q6. What are continuous and categorical variables?

- **Continuous variables:** Numerical values that can take any value
 - Example: Height, Weight, Salary
- **Categorical variables:** Values representing categories
 - Example: Gender, City, Color

Q7. How do we handle categorical variables in Machine Learning? What are common techniques?

Categorical data must be converted into numbers.

Common techniques:

1. **Label Encoding** – Assigns numbers to categories
2. **One-Hot Encoding** – Creates separate columns for each category
3. **Ordinal Encoding** – Used when categories have an order

Q8. What do you mean by training and testing a dataset?

- **Training dataset:** Used to train the model
- **Testing dataset:** Used to evaluate the model's performance on unseen data

Q9. What is `sklearn.preprocessing`?

`sklearn.preprocessing` is a module in **Scikit-learn** used for data preprocessing such as:

- Scaling
- Encoding
- Normalization

Example tools: `StandardScaler`, `LabelEncoder`, `OneHotEncoder`

Q10. What is a Test set?

A **test set** is a portion of data used only to check how well the trained model performs on new data.

Q11. How do we split data for model fitting (training and testing) in Python?

We use `train_test_split` from `sklearn.model_selection`.

Example:

```
from sklearn.model_selection import train_test_split  
  
X_train, X_test, y_train, y_test = train_test_split(  
    X, y, test_size=0.2, random_state=42  
)
```

Q12 Why do we have to perform EDA before fitting a model to the data?

Exploratory Data Analysis (EDA) helps us understand the data before training a model.
It is used to:

- Find missing values and outliers
- Understand data distribution
- Detect relationships between variables
- Choose the right features and model

☞ Without EDA, the model may give wrong or misleading results.

. What is correlation?

Correlation measures the strength and direction of the relationship between two variables.
Its value ranges from -1 to +1.

Q13. What does negative correlation mean?

Negative correlation means when one variable increases, the other decreases.
Example: As speed increases, travel time decreases.

Q14. How can you find correlation between variables in Python?

We use the `.corr()` function in Pandas.

Example:

```
import pandas as pd  
df.corr()
```

This gives the correlation matrix of numerical variables.

Q15. What is causation? Explain difference between correlation and causation with an example.

Causation means one variable directly causes a change in another variable.

Difference:

- Correlation: Variables move together
- Causation: One variable causes the change

Example:

Ice cream sales and drowning deaths are correlated, but ice cream does not cause drowning. Hot weather causes both → correlation ≠ causation.

Q16. What is an Optimizer? What are different types of optimizers? Explain each with an example.

An optimizer updates model parameters to minimize the loss function.

Common optimizers:

1. Gradient Descent – Updates parameters using full dataset
2. Stochastic Gradient Descent (SGD) – Uses one data point at a time
3. Mini-Batch Gradient Descent – Uses small batches
4. Adam Optimizer – Adaptive learning rate, fast and efficient

Example: Adam optimizer is widely used in deep learning.

Q17. What is `sklearn.linear_model`?

`sklearn.linear_model` is a module in Scikit-learn that provides linear models such as:

- Linear Regression
- Logistic Regression
- Ridge, Lasso

Q18. What does `model.fit()` do? What arguments must be given?

`model.fit()` trains the model using data.

Arguments:

- `X_train` → Input features
- `y_train` → Target variable

Example:

```
model.fit(X_train, y_train)
```

Q19. What does `model.predict()` do? What arguments must be given?

`model.predict()` generates predictions using the trained model.

Argument:

- `X_test` → Input data for prediction

Example:

```
y_pred = model.predict(X_test)
```

Q20. What are continuous and categorical variables?

- Continuous variables: Numerical values with range
 - Example: Age, Salary
- Categorical variables: Represent categories
 - Example: Gender, City

Q21. What is feature scaling? How does it help in Machine Learning?

Feature scaling converts features to the same scale.

Benefits:

- Improves model performance
- Speeds up training
- Important for distance-based algorithms

Q22. How do we perform scaling in Python?

Using `StandardScaler` or `MinMaxScaler`.

Example:

```
from sklearn.preprocessing import StandardScaler  
  
scaler = StandardScaler()  
X_scaled = scaler.fit_transform(X)
```

Q23. What is **sklearn.preprocessing**?

`sklearn.preprocessing` is a module used for:

- Scaling
- Encoding
- Normalization

Examples: `StandardScaler`, `LabelEncoder`, `OneHotEncoder`

Q24. How do we split data for model fitting (training and testing) in Python?

Using `train_test_split`.

Example:

```
from sklearn.model_selection import train_test_split  
  
X_train, X_test, y_train, y_test = train_test_split(  
    X, y, test_size=0.2, random_state=42  
)
```

Q25. Explain data encoding.

Data encoding converts categorical data into numerical form so ML models can understand it.

Common encoding techniques:

1. Label Encoding – Assigns numbers to categories
2. One-Hot Encoding – Creates separate columns
3. Ordinal Encoding – Used for ordered categories

If you want, I can also provide:

- 2-line short answers
- Hindi explanation
- Interview MCQs
- Practical code examples