

### 1. What is a parameter?

A parameter is a variable in a machine learning model that the algorithm learns from data during training (e.g., weights in linear regression).

### 2. What is correlation?

Correlation is a statistical measure that shows how two variables move in relation to each other.

### 3. What does negative correlation mean?

It means as one variable increases, the other decreases. They move in opposite directions.

### 4. Define Machine Learning. What are the main components in Machine Learning?

Machine Learning is the ability of machines to learn from data without being explicitly programmed.

**Main components:**

- Data
- Model
- Loss Function
- Optimizer
- Evaluation Metric

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

A lower loss value indicates the model's predictions are close to the actual values, suggesting better performance.

### 6. What are continuous and categorical variables?

- **Continuous:** Numeric values with infinite possibilities (e.g., height, salary).
- **Categorical:** Values that represent categories or labels (e.g., gender, color).

### 7. How do we handle categorical variables in Machine Learning? What are the common techniques?

Common techniques:

- **Label Encoding**
- **One-Hot Encoding**
- **Ordinal Encoding**

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

- **Training Set:** Used to fit the model.
- **Testing Set:** Used to evaluate model performance on unseen data.

### 9. What is sklearn.preprocessing?

A module in Scikit-learn for data preprocessing (e.g., scaling, encoding).

### 10. What is a Test set?

A separate portion of data used to evaluate the model after training.

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

```
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)
```

### 12. How do you approach a Machine Learning problem?

Steps:

- Understand the problem
- Collect and clean data
- Perform EDA
- Feature Engineering
- Train model
- Evaluate and tune
- Deploy

### 13. Why do we have to perform EDA before fitting a model to the data?

EDA helps understand patterns, detect outliers, and prepare data appropriately for modeling.

### 14. What is correlation?

(Repeated – see Q2)

### 15. What does negative correlation mean?

(Repeated – see Q3)

### 16. How can you find correlation between variables in Python?

```
df.corr() # For numeric columns
```

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

- **Causation:** One variable causes a change in another.
- **Correlation:** Two variables move together, but one doesn't cause the other.  
**Example:** Ice cream sales and drowning are correlated but not causally related (temperature is the hidden cause).

### 18. What is an Optimizer? What are different types of optimizers?

Optimizers adjust model parameters to reduce loss.

**Examples:**

- **SGD:** Simple gradient descent
- **Adam:** Adaptive moment estimation

- **RMSprop**: Maintains moving average of squared gradients

**19. What is sklearn.linear\_model?**

A Scikit-learn module for linear models like LinearRegression, LogisticRegression, etc.

**20. What does model.fit() do? What arguments must be given?**

Trains the model using training data.

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

**21. What does model.predict() do? What arguments must be given?**

Predicts target values for input data.

```
model.predict(X_test)
```

**22. What are continuous and categorical variables?**

(Repeated – see Q6)

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

Scaling standardizes feature ranges (e.g., 0–1). It improves convergence and accuracy in models like SVM, KNN.

**24. How do we perform scaling in Python?**

```
from sklearn.preprocessing import StandardScaler
```

```
scaler = StandardScaler()
```

```
X_scaled = scaler.fit_transform(X)
```

**25. What is sklearn.preprocessing?**

(Repeated – see Q9)

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

(Repeated – see Q11)

**27. Explain data encoding.**

Data encoding transforms categorical data into numeric form.

Examples: Label Encoding, One-Hot Encoding.