1. What is a parameter?

A **parameter** is a value that defines a model.
 Examples: slope and intercept in linear regression.

2. What is correlation?

Correlation tells how **two variables are related** (how one changes when the other changes).

3. What does negative correlation mean?

Negative correlation means:

- When **one variable increases**, the **other decreases**.
  Example: Speed ↑, Travel time ↓.

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

Machine Learning is a field where computers **learn patterns from data** without being explicitly programmed.

Components:

1. Dataset

2. Features (input variables)

3. Model

4. Loss function

5. Optimizer

6. Training process

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

Loss value shows **how far predictions are from actual values**.

- **Low loss = good model**

- **High loss = bad model**

6. What are continuous and categorical variables?

- **Continuous:** Numerical values that can take any number (height, temperature).

- **Categorical:** Non-numerical groups/labels (gender, city, color).

variables in ML? Common techniques?

We convert categories to numbers using:

- **Label Encoding**

- **One-Hot Encoding**

- **Ordinal Encoding**

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

- **Training data:** Used to teach the model.

- **Testing data:** Used to check the model's performance on unseen data.

9. What is sklearn.preprocessing?

It is a module in sklearn used for **data preparation**, such as scaling, encoding, normalization.

10. What is a Test set?

A test set is a part of the dataset used **only for evaluating the model**, not for training.

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

Using sklearn:

```
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?

1. Understand the problem

2. Collect data

3. Clean the data

4. Perform EDA

5. Preprocess (encode, scale)

6. Choose model

7. Train model

8. Evaluate

9. Improve

10. Deploy

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

EDA helps to:

● Understand patterns

- Find missing values

- Detect outliers

- Understand feature relationships

- Choose correct preprocessing and model

14. What is correlation?

Same as Q2.
 Correlation → Strength of relationship between two variables.

15. What does negative correlation mean?

Same as Q3.
 When one increases, the other decreases.

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

Using pandas:

```
df.corr()
```

Or heatmap:

```
import seaborn as sns
sns.heatmap(df.corr(), annot=True)
```

17. What is causation? Difference between correlation and causation (example)?

- **Causation:** One variable **causes** change in another.

- **Correlation:** Two variables are related but **do not necessarily cause** each other.

Example:
Ice cream sales ↑ and drowning cases ↑ → correlated
But ice cream does **not** cause drowning (cause is summer heat).

18. What is an Optimizer? Types of optimizers? Explain with example.

Optimizer updates model parameters to **reduce loss**.

Types:

- **SGD:** Basic gradient descent

- **Momentum:** Adds velocity to SGD

- **Adam:** Adaptive learning rate, most widely used

- **RMSProp:** Adjusts learning rate using moving averages

Example (Adam in TensorFlow):

```
optimizer = tf.keras.optimizers.Adam()
```

19. What is sklearn.linear_model?

A module that contains linear models such as:

- LinearRegression

- LogisticRegression

- Ridge

- Lasso

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

model.fit() **trains the model** using training data.
Arguments:

- **X_train** (features)

- **y_train** (labels)

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

model.predict() returns **predictions** for new data.
 Arguments:

- **X_test** or any input features

22. What are continuous and categorical variables?

(Repeat of Q6.)
 Continuous → numbers
 Categorical → labels

23. What is feature scaling? How does it help?

Feature scaling means bringing all numerical values to a **similar range**.

Helps by:

- Making training faster

- Improving accuracy

- Necessary for algorithms like SVM, KNN, Logistic Regression

24. How do we perform scaling in Python?

Using StandardScaler or MinMaxScaler:

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

25. What is sklearn.preprocessing?

(Repeat of Q9.)
 Used for preprocessing tasks → encoding, scaling, normalization.

---

26. How do we split data for model fitting in Python?

(Repeat of Q11.)
 train_test_split()

27. Explain data encoding.

Data encoding means converting **text/categorical values into numerical values** so ML models can understand them.

Types:

- Label Encoding

- One-Hot Encoding

- Binary Encoding

- Ordinal Encoding