**Question 1**: What is Logistic Regression, and how does it differ from Linear Regression?

**Answer1:** Logistic Regression is a statistical method used to predict the probability of a categorical outcome, such as yes/no or 0/1. It uses the sigmoid function to convert predictions into probabilities between 0 and 1.

It differs from Linear Regression, which predicts continuous values like height or price. While Linear Regression gives any numeric output, Logistic Regression predicts the likelihood of an event happening, making it suitable for classification problems.

**Question 2**: Explain the role of the Sigmoid function in Logistic Regression

**Answer2:** The Sigmoid function in Logistic Regression is used to convert the output of a linear equation into a value between 0 and 1.

This helps interpret the result as a probability — for example, the probability that an event will happen.

If the output is greater than 0.5, it is usually classified as 1 (Yes), and if it's less than 0.5, it's classified as 0 (No).

In short, the sigmoid function helps Logistic Regression make clear yes/no or true/false predictions.

**Question 3**: What is Regularization in Logistic Regression and why is it needed?

**Answer3 :**Regularization in Logistic Regression is a method used to reduce overfitting by adding a penalty to large coefficient values in the model.

It helps the model stay simple and generalize better to new data.

Without regularization, the model might fit the training data too closely and perform poorly on unseen data.

Common types are L1 (Lasso) and L2 (Ridge) regularization, which keep the model's weights small and stable

**Question 4**: What are some common evaluation metrics for classification models, and why are they important?

**Answer4:** Some common evaluation metrics for classification models are Accuracy, Precision, Recall, F1-Score, and the Confusion Matrix.

- Accuracy measures how often the model predicts correctly.

- Precision shows how many of the predicted positives are actually correct.
- Recall tells how many actual positives the model was able to find.
- F1-Score is the balance between precision and recall.
- The Confusion Matrix shows the counts of correct and incorrect predictions.

These metrics are important because they help us understand how well the model performs, especially when dealing with imbalanced datasets or different types of classification errors.

**Question 10:** Imagine you are working at an e-commerce company that wants to predict which customers will respond to a marketing campaign. Given an imbalanced dataset (only 5% of customers respond), describe the approach you'd take to build a Logistic Regression model — including data handling, feature scaling, balancing classes, hyperparameter tuning, and evaluating the model for this real-world business use case.

## **Answer10: Approach to Build Logistic Regression on Imbalanced Dataset**

1. Understand the Problem and Data:
    - Target variable: customer response (Yes/No)
    - Imbalance: only 5% positive responses → need careful handling

2. Data Preprocessing:
    - Handle missing values and outliers
    - Encode categorical features (one-hot encoding)
    - Feature scaling (e.g., StandardScaler) to standardize numeric features

3. Handling Class Imbalance:
    - Use oversampling (SMOTE) or undersampling to balance classes
      Or set `class_weight='balanced'` in Logistic Regression

4. Model Training & Hyperparameter Tuning:
    - Use Logistic Regression with regularization (L1/L2)
    - Tune C (regularization strength) and penalty using GridSearchCV
    - Optionally try different solvers (liblinear, saga)

5. Evaluation Metrics:
    - Accuracy is misleading for imbalanced data
    - Use metrics like Precision, Recall, F1-score, ROC-AUC
    - Use a Confusion Matrix to visualize true positives, false positives, etc.

6. Validation:
    - Use cross-validation to ensure the model generalizes
    - Check precision-recall curve to choose the optimal threshold for predicting "responded"

7. Deployment & Monitoring:
    - Use the model to score customers for marketing
    - Monitor real-world performance and update the model periodically