**1. How does logistic regression differ from linear regression?**

* **Linear Regression:**  
  Predicts a **continuous** numeric output.  
  Example: Predicting house prices.
* **Logistic Regression:**  
  Predicts a **probability** for **classification** problems (usually binary).  
  It uses the **sigmoid function** to squash output between 0 and 1.  
  Example: Predicting whether an email is spam or not (0 or 1).

**2. What is the sigmoid function?**

The **sigmoid function** is used in logistic regression to map any real-valued number into a value between **0 and 1**:

σ(x)=11+e−x\sigma(x) = \frac{1}{1 + e^{-x}}

* If output is **close to 1** → predicts **positive class**
* If output is **close to 0** → predicts **negative class**

**3. What is precision vs recall?**

| **Metric** | **Definition** | **Formula** |
| --- | --- | --- |
| **Precision** | Of all predicted positives, how many were correct? | TP / (TP + FP) |
| **Recall** | Of all actual positives, how many were caught? | TP / (TP + FN) |

* **Precision** focuses on **quality** (fewer false positives).
* **Recall** focuses on **completeness** (fewer false negatives).

**4. What is the ROC-AUC curve?**

* **ROC (Receiver Operating Characteristic)** curve:  
  Plots **True Positive Rate (Recall)** vs. **False Positive Rate** at different thresholds.
* **AUC (Area Under the Curve):**  
  Measures overall model performance (higher is better).  
  AUC = 1 → perfect model, AUC = 0.5 → random guessing.

**5. What is the confusion matrix?**

A table to evaluate classification performance:

|  | **Predicted Positive** | **Predicted Negative** |
| --- | --- | --- |
| **Actual Positive** | True Positive (TP) | False Negative (FN) |
| **Actual Negative** | False Positive (FP) | True Negative (TN) |

Helps compute metrics like accuracy, precision, recall, F1-score.

**6. What happens if classes are imbalanced?**

* The model may **favor the majority class**, leading to **misleading accuracy**.
* For example, if 95% of samples are class A, predicting everything as A gives 95% accuracy but is **useless**.
* Use techniques like:
  + **Resampling** (oversample minority or undersample majority)
  + **Class weights**
  + **Alternative metrics** (precision, recall, F1-score, AUC)

**7. How do you choose the threshold?**

* Default threshold is **0.5** (if probability > 0.5 → positive class).
* But you can adjust it to:
  + **Increase precision** (raise threshold)
  + **Increase recall** (lower threshold)
* Use tools like **ROC curve** or **Precision-Recall curve** to find the optimal threshold for your business goal.

**8. Can logistic regression be used for multi-class problems?**

Yes.

* Logistic regression can be extended using:
  + **One-vs-Rest (OvR):** Train a binary classifier for each class.
  + **Multinomial logistic regression (Softmax):** Predicts probabilities for **all classes simultaneously**.

Most libraries (like scikit-learn) support this with multi\_class='multinomial'.