**Interview Questions for Logistic Regression:**

1. What is the difference between precision and recall?  
Ans.

|  |  |
| --- | --- |
| Precision | Recall |
| The ratio of true positive predictions to all positive predictions. | The ratio of true positive predictions to all actual positive. |
| Measures how many of the predicted positives are actually correct. | Measures how well the model identifies all actual positives. |
| Minimizes false positives (FP) | Minimizes false negatives (FN) |
| Important when the cost of false positives is high (e.g., spam detection). | Important when the cost of false negatives is high (e.g., disease diagnosis). |
| Precision= True Positives / (True Positives + False Positives ​) | Recall= True Positives / (True Positives + False Negatives ​) |

### 2. What is cross-validation, and why is it important in binary classification? Ans. Cross-validation is a statistical technique used to evaluate the performance and generalizability of a machine learning model. It involves dividing the dataset into multiple subsets (or "folds"), training the model on some of these subsets, and validating it on the remaining subsets. This process is repeated several times, and the results are averaged to obtain a more robust estimate of the model's performance.

**Common Types of Cross-Validation:**

1. **K-Fold Cross-Validation**:
   * The data is split into *k* equally sized folds.
   * The model is trained on *k-1* folds and validated on the remaining fold.
   * This process is repeated *k* times, with each fold serving as the validation set once.
   * The average performance across all folds is calculated.
2. **Stratified K-Fold Cross-Validation**:
   * Similar to K-Fold, but the folds are created so that each fold has the same proportion of classes as the original dataset. This is particularly important in binary classification to ensure that both classes are well-represented in each fold.
3. **Leave-One-Out Cross-Validation (LOOCV)**:
   * A special case of K-Fold where *k* equals the number of data points. Each data point is used once as a validation set, and the model is trained on all other points.

**Why is Cross-Validation Important in Binary Classification?**

1. **Avoids Overfitting**:
   * Cross-validation helps to ensure that the model is not just memorizing the training data but is learning to generalize to unseen data. This is crucial in binary classification, where the model must accurately classify new, unseen instances into one of two classes.
2. **Provides a More Reliable Estimate of Model Performance**:
   * By using different subsets of the data for training and validation, cross-validation provides a more comprehensive assessment of the model’s performance, reducing the bias that could result from a single train-test split.
3. **Works Well with Limited Data**:
   * In binary classification, especially when data is limited, cross-validation maximizes the use of available data by allowing each data point to be used for both training and validation.
4. **Helps in Model Selection and Hyperparameter Tuning**:
   * Cross-validation is often used to compare different models or tune hyperparameters by assessing how well they perform across multiple folds. This ensures that the selected model and its parameters are optimal and not biased by a particular data split.
5. **Handles Class Imbalance**:
   * Stratified K-Fold Cross-Validation is particularly useful in binary classification when dealing with imbalanced classes. It ensures that each fold has a similar distribution of the two classes, leading to more consistent performance evaluation.