Logistic Regression

**Interview Questions:**

1. What is the difference between precision and recall?

 **Precision**: Measures the accuracy of positive predictions. It’s the ratio of true positives (correctly predicted positives) to the sum of true positives and false positives.  
**Formula:** *Precision = TP / (TP + FP)*

* **Use case:** Important when false positives are costly (e.g., email spam detection).

 **Recall (Sensitivity or True Positive Rate):** Measures how well the model captures all actual positives. It’s the ratio of true positives to the sum of true positives and false negatives.  
**Formula:** *Recall = TP / (TP + FN)*

* **Use case:** Crucial when missing a positive case is costly (e.g., disease detection).
* **Key Difference:** Precision focuses on the correctness of positive predictions, while recall focuses on identifying as many positives as possible.

1. What is cross-validation, and why is it important in binary classification?

 **Cross-Validation:** A technique for assessing how well a model generalizes to an independent dataset. It involves splitting the dataset into multiple folds (commonly 5 or 10), training the model on some folds, and testing it on the remaining fold. This process repeats, and results are averaged.

* **Common method:** *k-Fold Cross-Validation* (data is split into *k* subsets).

 **Why It’s Important in Binary Classification:**

* **Reduces Overfitting:** Helps ensure the model doesn’t just memorize the training data.
* **Reliable Performance Estimate:** Provides a more accurate measure of model performance compared to a simple train-test split.
* **Efficient Use of Data:** All data points get to be in both the training and testing sets.