Logistic Regression

**Interview Questions:**

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

ANS =

|  |  |
| --- | --- |
| Precision | Recall |
| A) When calculating a model's precision, it is important to consider both the positive and negative samples being classified. | A) We only need positive samples to calculate the Recall for a model. All negative samples will be ignored. |
| B) This allows us to measure the ability to classify the positive samples in the model. | B) This allows us to determine how many positive samples have been correctly classified using the ML model. |
| C) Precision of the ML model depends on both the negative as well as positive samples. | C) Recall of the ML model is dependent upon positive samples and independent of negative samples. |
| D) In Precision, all positive samples that are classified as positive should be considered either correctly or incorrectly. | D) Recall cares about accurately classifying all positive samples. It doesn't care if any negative samples are classified as positive. |

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

Cross-validation is a technique used to assess the performance of a machine learning model by splitting the data into multiple subsets and training the model on different combinations of these subsets. It's crucial in binary classification to provide a more reliable estimate of the model's performance and prevent overfitting.

**Why it's important in binary classification:**

* **Reliable performance estimation:**

Cross-validation provides a more robust estimate of how well the model will generalize to unseen data compared to a single train-test split. It utilizes the entire dataset for both training and testing, reducing the variance in performance estimates.

* **Overfitting detection:**

By training and evaluating the model on different subsets of the data, cross-validation helps identify if the model is overfitting the training data. Overfitting occurs when the model learns the training data too well, including noise, and performs poorly on new, unseen data.

* **Model comparison:**

Cross-validation facilitates the comparison of different models or algorithms by providing a consistent evaluation across different subsets of the data, ensuring a fair comparison.

* **Hyperparameter tuning:**

Cross-validation can be used to optimize the hyperparameters of a model, which are parameters that are set before training, by evaluating the model's performance with different hyperparameter settings.

* **Imbalanced datasets:**

In binary classification, datasets may have an unequal number of examples for each class (imbalanced data). Cross-validation techniques like stratified cross-validation ensure that each fold maintains the same class distribution as the original dataset, preventing biased model training.

* **Reduced variance:**

By using multiple training and validation sets, cross-validation reduces the variance in the performance estimates, providing a more stable and reliable evaluation of the model.