5 QUESTIONS on “Accuracy” for Data Science and AI Interviews

DINESHLAL

01

What is Accuracy in the context of Machine Learning?

Easy Explanation: Accuracy measures how often the model correctly predicts the outcome. It is a simple and commonly used metric in classification tasks.

Detailed Explanation: Accuracy is the ratio of correct predictions (both positive and negative) to the total number of predictions.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Where:

- TP (True Positives): Correctly predicted positive cases.
- TN (True Negatives): Correctly predicted negative cases.
- FP (False Positives): Incorrectly predicted as positive.
- FN (False Negatives): Incorrectly predicted as negative.

01

What is Accuracy in the context of Machine Learning?

Example: If a model predicts whether an email is spam, and out of 100 emails, it correctly predicts 90 (both spam and non-spam), the accuracy would be:

$$Accuracy = \frac{90}{100} = 0.9 \text{ or } 90\%$$

Key Points:

- Accuracy works best when the data is balanced (equal number of classes).
- It may not be the best metric when dealing with imbalanced datasets, as it can be misleading.

02

Why might accuracy not be a good measure in some cases?

Easy Explanation: While accuracy seems like a good indicator, it can be misleading, especially in cases where the data is imbalanced (i.e., one class is much larger than the other).

Detailed Explanation:

- Imbalanced datasets: If one class dominates the dataset, a model predicting the majority class can have high accuracy but poor performance for the minority class.
    - Example: In a dataset with 95% "Not Fraud" cases and 5% "Fraud" cases, if the model predicts all cases as "Not Fraud," it will still have 95% accuracy but won't detect any fraud, which is critical.

02

Why might accuracy not be a good measure in some cases?

Detailed Explanation:

Alternative Metrics: In such cases, you should consider metrics like:

- Precision: Focuses on how many of the predicted positives are actual positives.
- Recall: Focuses on how many actual positives are correctly predicted.
- F1 Score: Harmonic mean of precision and recall, useful when seeking balance.
    -

03

What is the difference between accuracy and precision?

Easy Explanation: Accuracy tells us how many predictions were correct overall, while precision focuses on the quality of positive predictions.

Detailed Explanation:

- Accuracy: Measures the proportion of correct predictions out of all predictions.
    - Example: In a binary classification of whether an animal is a dog, accuracy measures how often the model is right (both for dogs and non-dogs).

03

What is the difference between accuracy and precision?

- Precision: Measures how many of the predicted positives (dogs) were actually correct.
  - Example: If the model predicts 10 animals as dogs but only 6 of them are actually dogs, the precision is:

$$Precision = \frac{6}{10} = 0.6 \text{ or } 60\%$$

Key Difference:

- Accuracy considers both true positives and true negatives.
- Precision focuses only on the correctness of the positive predictions.

04

What is the relationship between accuracy and F1 score?

Easy Explanation: F1 score is a balanced metric that combines both precision and recall, and is useful when accuracy might be misleading due to imbalanced classes.

Detailed Explanation:

- Accuracy: Gives the overall correctness of the model but doesn't distinguish between classes.
- Example: A model can have high accuracy but poor F1 score if it misses many true positives.
- F1 Score: The harmonic mean of precision and recall, balancing both metrics. It is useful when you need to balance the trade-off between precision and recall, especially in imbalanced datasets.

04

What is the relationship between accuracy and F1 score?

Detailed Explanation:

- Example: In a medical diagnosis case, if the model predicts few true positives (sick patients), it can still have good accuracy, but the F1 score will be low, showing that it isn't performing well for detecting sick patients.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

05

Can a model have high accuracy but still be ineffective? Explain with an example.

Easy Explanation: Yes, especially in imbalanced datasets, a model can appear to perform well based on accuracy but fail to identify important cases.

Detailed Explanation:

- Imbalanced datasets: In cases where one class dominates, accuracy can be high without the model performing well for the minority class.
- Example: Suppose you have a dataset where 95% of the data points are labeled "Not Spam" and only 5% are "Spam." If the model predicts all emails as "Not Spam," it will have 95% accuracy, but it will completely fail to detect any spam emails. This makes the model ineffective for the real task.

05

Can a model have high accuracy but still be ineffective? Explain with an example.

Detailed Explanation:

- Better Metrics:
- Precision and Recall should be used to evaluate performance on the minority class.
- The F1 score provides a more balanced evaluation in such cases.

FOLLOW ALONG

for learning more on Data Science, AI and Generative AI

Dinesh Lal