Sakshatar

Welcome to the Sakshatar Experience! ■

Introduction

This report contains an overview of the interview questions and the corresponding answers provided. The insights gathered aim to enhance the evaluation process and improve overall performance.

Question 1: What is the role of the k value in the k-NN algorithm?
Answer: The k value in the k-NN algorithm represents the number of nearest neighbors to consider when making a prediction. A smaller k value leads to more flexible and potentially noisy predictions, while a larger k value leads to more stable predictions but may overlook local patterns.

Question 2: Why is Logistic Regression termed as Regression and not classification?
Answer: Despite having the word "Regression" in its name, Logistic Regression is actually a classification algorithm. The term "Regression" comes from the linear regression algorithm it's based on, used to predict continuous values, while Logistic Regression is specifically used for binary classification problems, predicting the probability of an instance belonging to a particular class.

Question 3: What are the disadvantages of Information Gain?
Answer: Information Gain tends to favor attributes with a large number of values, which can lead to overfitting. It may also struggle with continuous-valued attributes, as it partitions them into discrete intervals, potentially losing valuable information during this process.

Question 4: Can we solve the multiclass classification problems using Logistic Regression? If Yes then How?
Answer: Yes, we can solve multiclass classification problems using Logistic Regression by applying techniques such as one-vs-rest (OvR) or multinomial methods. In OvR, we train a separate binary logistic regression model for each class, while in the multinomial method, a single model is trained to predict probabilities for each class.

Question 5: What do you understand about Information Gain? Also, explain the mathematical formulation associated with it.
Answer: "Information Gain measures the reduction in uncertainty of the target variable after splitting the data by an attribute. It is used in decision tree algorithms to choose the best attribute for splitting at each node.
Mathematically, Information Gain is calculated as:
IG(D, A) = H(D) - H(D|A)
Where:
- IG(D, A) is the Information Gain
- H(D) is the entropy of the original dataset D
- H(D|A) is the conditional entropy of D given the attribute A"

Conclusion

This report summarizes the insights gained from the interview process.

Thank You! ■

We appreciate your effort and time! ■