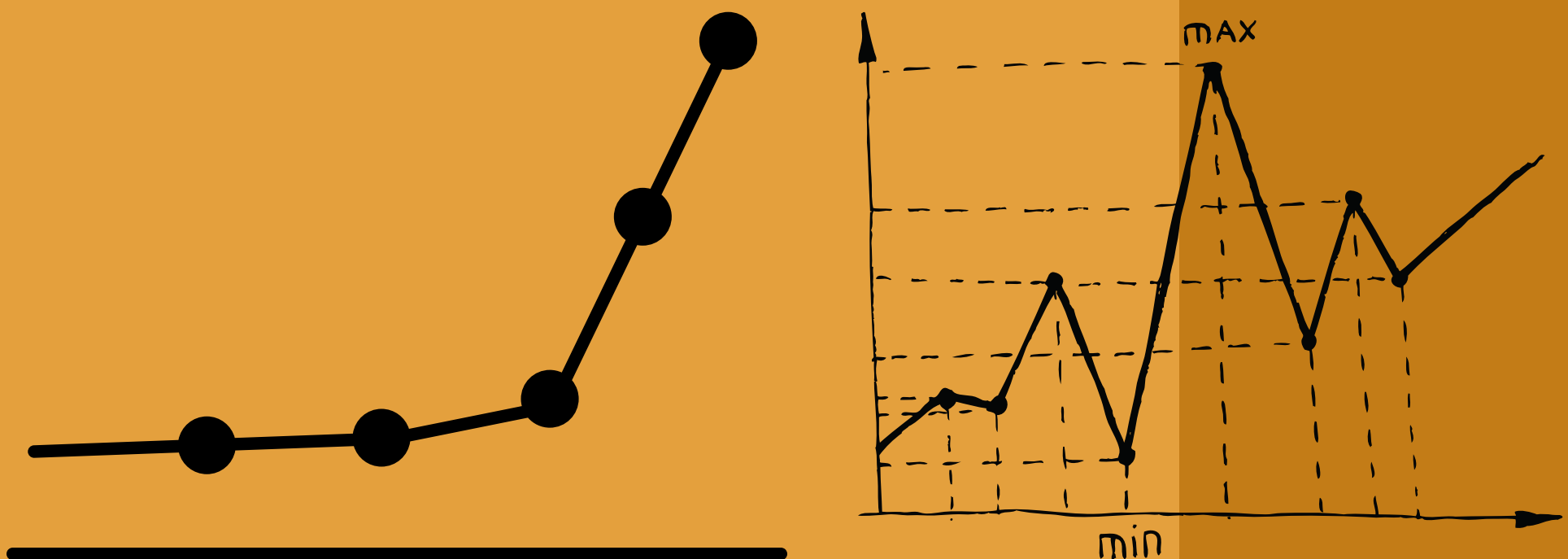


# DATA SCIENCE

## TOP 10

### INTERVIEW

### QUESTIONS



# **1. What is the difference between supervised and unsupervised learning?**

Supervised learning involves training a model on labeled data to make predictions, while unsupervised learning involves finding patterns and structures in unlabeled data without specific target outcomes.



## **2. What is the Central Limit Theorem and why is it important?**

The Central Limit Theorem states that the distribution of sample means approximates a normal distribution, regardless of the shape of the population distribution. It is crucial in inferential statistics as it allows us to make inferences about a population based on a sample.



### **3. Explain the concept of regularization in machine learning.**

Regularization is a technique used to prevent overfitting in machine learning models. It adds a penalty term to the loss function, discouraging complex models and promoting simplicity, ultimately improving the model's generalization performance.



## **4. What is feature selection and why is it important?**

Feature selection is the process of selecting the most relevant and informative features from a dataset. It helps improve model performance by reducing dimensionality, eliminating noise, and enhancing interpretability.



## **5. What is the difference between bagging and boosting?**

Bagging and boosting are ensemble learning techniques. Bagging involves training multiple models independently on different subsets of the training data and aggregating their predictions, while boosting focuses on sequentially training models, giving more weight to misclassified instances to improve overall performance.

## **6. What is cross-validation and why is it useful?**

Cross-validation is a technique used to evaluate a model's performance by partitioning the data into multiple subsets. It helps assess the model's ability to generalize to unseen data and aids in hyperparameter tuning and model selection.



## **7. Explain the bias-variance tradeoff.**

The bias-variance tradeoff refers to the balance between a model's ability to accurately capture the true underlying patterns (low bias) and its sensitivity to fluctuations in the training data (low variance). Finding the optimal tradeoff is crucial to avoid underfitting or overfitting.





## **8. What is A/B testing and how is it used in data science?**

A/B testing is a statistical technique used to compare two versions of a variable (A and B) to determine which performs better. In data science, it is commonly used to evaluate the impact of changes in a product or system, helping in decision-making and optimization.



## 9. How do you handle missing values in a dataset?

Missing values can be handled through techniques such as imputation (replacing missing values with estimated values), deletion (removing instances with missing values), or using algorithms specifically designed to handle missing data.



# 10. What is the difference between overfitting and underfitting?

Overfitting occurs when a model performs well on the training data but poorly on unseen data due to capturing noise or irrelevant patterns. Underfitting happens when a model is too simple and fails to capture the underlying patterns in the data.



Thank  
You!

Make sure to follow for  
more interesting contents.



[www.linkedin.com/in/prashantkr57](https://www.linkedin.com/in/prashantkr57)