

4. What are some of the techniques used for sampling? What is the main advantage of sampling?
5. List down the conditions for Overfitting and Underfitting.
6. Differentiate between the long and wide format data.
7. What are Eigenvectors and Eigenvalues?
8. What does it mean when the p-values are high and low?
9. When is resampling done?
10. What do you understand by Imbalanced Data?
11. Are there any differences between the expected value and mean value?
12. What do you understand by Survivorship Bias?
13. What is a Gradient and Gradient Descent?
14. Define confounding variables.
15. Define and explain selection bias?
16. Define bias-variance trade-off?
17. Define the confusion matrix?
18. What is logistic regression? State an example where you have recently used logistic regression.
19. What is Linear Regression? What are some of the major drawbacks of the linear model?
20. What is a random forest? Explain it's working.
21. In a time interval of 15-minutes, the probability that you may see a shooting star or a bunch of them is 0.2. What is the percentage chance of you seeing at least one star shooting from the sky if you are under it for about an hour?
22. What is deep learning? What is the difference between deep learning and machine learning?

Data Science Interview Questions for Experienced ▾

1. How are the time series problems different from other regression problems?
2. What are RMSE and MSE in a linear regression model?
3. What are Support Vectors in SVM (Support Vector Machine)?
4. So, you have done some projects in machine learning and data science and we see you are a bit experienced in the field. Let's say your laptop's RAM is only 4GB and you want to train your model on 10GB data set.
5. Explain Neural Network Fundamentals.
6. What is Generative Adversarial Network?
7. What is a computational graph?
8. What are auto-encoders?
9. What are Exploding Gradients and Vanishing Gradients?
10. What is the p-value and what does it indicate in the Null Hypothesis?
11. Since you have experience in the deep learning field, can you tell us why TensorFlow is the most preferred library in deep learning?
12. Suppose there is a dataset having variables with missing values of more than 30%, how will you deal with such a dataset?
13. What is Cross-Validation?
14. What are the differences between correlation and covariance?
15. How do you approach solving any data analytics based project?
16. How regularly must we update an algorithm in the field of machine learning?