

Data Science Question Bank

1. What do you mean by regression?
2. What is the equation of simple linear regression?
3. What is the equation of multiple linear regression?
4. Where is non-linear regression used?
5. What do you mean by the Polynomial regression?
6. What is logistic regression?
7. What do you mean by k-nn approach?
8. Why is weighted k-nn used?
9. What do you mean by the r-near algorithm?
10. What is the disadvantage of near neighbour approach?
11. What is branch and bound algorithm? What is its advantage?
12. What do you mean by a projection algorithm?
13. What is the Bayes theorem?
14. What is the naïve Bayes approach?
15. Numerical based on Naïve Bayes approach
16. What are the salient features of decision trees?
17. What are axis parallel, oblique split and non-linear tests?
18. State various impurities in decision trees
19. Numerical based on impurity calculations
20. What is decision tree pruning?
21. What are the advantages of decision tress?
22. What do you mean by inter cluster and intra cluster distance?
23. What are the various types of clustering approaches?
24. What is polythetic clustering?
25. What is monothetic clustering?
26. What are the advantages and disadvantages of agglomerative clustering?
27. What is the k-means clustering approach?
28. What do you mean by holdout, resubstitution, cross-validation and bootstrapping strategies for training and testing?
29. What do you mean by a confusion matrix?
30. What are various terms like accuracy, precision, recall, sensitivity, specificity, f-score?