

## Linear Regression

1. What is Linear Regression? What are the Assumptions involved in it?
2. How to find the parameters of the linear model?
3. Assume you need to generate a predictive model using multiple regression. Explain how you intend to validate this model
4. **What do you mean by Regularisation? What are L1 and L2 Regularisation?**

## Logistic Regression

5. What is Logistic Regression? What is the loss function in LR?
6. What is confusion matrix, precision, recall, TPR, FPR and how is the ROC AUC score calculated ?

## ML Fundamentals

7. What is the difference between supervised and unsupervised learning?
8. What is bias-variance tradeoff and how is the overfitting and underfitting related to it ?
9. What is meant by Cross Validation. Name some common cross Validation techniques?
10. Difference between convex and non-convex cost function; what does it mean when a cost function is non-convex?
11. What are some of the steps for data wrangling and data cleaning before applying machine learning algorithms?
12. What is the difference between bagging and boosting?

## Naive Bayes Theorem

13. What is bayes theorem - how do we use it to do classification?
14. How is multi-class classification done with naive bayes?
15. What are some drawbacks of naive bayes algorithm?

## Visualisations

16. What is the difference between a histogram and bar plot?

## Decision Tree

17. What is entropy and information gain?
18. How do you control for overfitting in a decision tree?
19. How are prediction probabilities calculated in a decision tree?
20. Can we do regression with decision tree? How?

## K Means Algorithms

1. Is Feature Scaling required for the K means Algorithm?
2. How to find the K in k-means clustering?
3. What is a centroid point in K means Clustering?
4. Does centroid initialization affect K means Algorithm?
5. Is there a way of finding the accuracy of k-means algorithm?

## Mathematics/Probability

1. What is the distribution of sample Means of a distribution with mean  $\mu$  and variance  $\sigma^2$ ?
2. What is eigenvalue and eigenvector?
3. What is a null hypothesis? What is type 1 and type 2 error and how is it controlled ? What is the power of a hypothesis test?

## Programming

4. Write a program to count the frequency of occurrence of a character in a string?
5. Given a word, reverse only the vowels in the word keeping the consonants in the same place? Input : "Saudi" Output : "Siuda"

Input : "kousik" Output : "kiosk"

6. What is the complexity of linear search across an array of size  $n$ ? What is the complexity of binary search across an array of size  $n$ ? What property must the array have to do binary search?
7. How can we estimate the value of  $\pi$  using simulation?