C.C.O'
26-09-22

Suggestion Questions :-

1. Explain the process of ETL in brief?
2. What is stacking in data science?
3. What do you mean by population and sample in statistics?
4. What is the importance of central tendancy of data?
5. Explain normalization and orthogonality of data?
   ↓
   (Standanization of data in PCA)
6. Explain Null and Alternate hypothesis.
7. What is meant by t-value?
8. What is k-Fold Cross validation?
9. Explain ROC Curve with diagram
10. Explain list slicing and list comprehension in python with example.
11. What is packing and unpacking of data in python?

* Null Hypothesis Or Alternate Hypothesis

The Null or the Alternative Hypothesis are two competing claims that researchers weigh evidence for and against using a statistical test.

Null hypothesis (H0) :-
There is no effect in the population.

C.W.
27-09-22

17. What are the data types in python Explain mutable and Immutable Objects in python?
18. Explain List comprehension and slicing in python
19. List down the condition for Over-fitting and underfittig?
~~20~~ What is meant by Imbalance data
~~20~~ ~~List down the~~
20. How can we select an appropriate value of K in K-means Clustering algorithm.
Define Precision and recall.

Camlin Page
Date / /

Long Answer Questions (15 Marks)

1. (a) Explain the data science life cycle with diagram
   (b) What are the different techniques of data engineering?
   (c) Differentiate between covariance and co-relation of data.

2. (a) Explain Nominal and, ordinal, interval and ratio variables with examples.
   (b) Discuss the different techniques of dimensionality reduction.
   (c) What is Vector space Model (V.S.M.)

3. (a) Discuss Normal, chi-square and t-student distribution of data
   (b) What is the role of Eigen value and Eigen vector in PCA-Algorithm.

4. (a) Given the following data set use PCA to reduce the dimension from 2 to 1. Draw the one dimensional data distribution chart after dimensionality reduction.

(b) What is Co-variance Matrix ?

(c) What is L.D.A. ?

5. (a) Write down the steps of PCA Algorithm?

(b) What are the applications of PCA Algorithm?

(c) Name the graphical techniques used in EDA.

6. (a) Discuss the different techniques used in EDA.

(b) Explain normalization and Orthogonality of data of variables.

7. (a) Discuss the different techniques used in descriptive and inferential statistics.

(b) Explain the following:
Standard deviation, Kurtosis and Skewness of data.

8. (a) Describe the steps of obtaining a random sample of data.

(b) What is Hypothesis testing.

(c) Briefly Explain Confidence Interval.

9. (a) Write down the differences between descriptive and Inferential statistics.

(b) Explain Null and Alternative Hypothesis.

(c) What is the importance of p value in statistics.

10. (a) Explain 1- sample, 2- sample and paired t Sample test.

(b) What is meant by p-value

11. (a) Draw the generic workflow of any machine learning Algorithm.

(b) Discuss the applications of machine learning in detail, banking & Healthcare sector.

(c) What is meant by discretization and Binarization of data?

12. (a) Explain Parametric & Non-parametric Statistical learning methods.

(b) What is regression Analysis.
Write down the different techniques of regression Analysis.

(c) Discuss the differences between classification and Regression.

13. (a) Find the means of x and Y variables and the coefficient of Co-relation between them from the following two Regression Equations.

$$2y - x - 50 = 0$$
$$3y - 2x - 10 = 0$$

(b) What are the assumptions that must hold for a Regression Model?

(c.) What do you Understand by Logistic Regression?

14. (b.) What is Bias in data science?

(c.) What is K-fold cross Validation?

(a.) Mention the different techniques used for sampling.

15 (a.) What is F1 score and how to Calculate it.

(b.) Define the Similarity Functions :-
Jaccard Index, Co-sine Based similarity and Manhatten distance similarity.

(c.) What is the difference between an error and residual error.

(d.) What is the Bias - Variance trade off in Data science.

16. (a.) What is R.M.S.E (Root mean square error)?

(b.) How to Calculate Binary classification error using its Confusion matrix?

(c.) What is Ensemble learning?

(d.) What is kernel Function in SVM?

17. (a.) Explain univariate, Bivariate and multivariate Analysis.

(b.) How can we handle missing data?

(c.) How to deal with Outliers.

18. (a.) ~~How~~ Explain Bagging in data science.

(b.) What are the popular python libraries used in data science?

(c.) What are the support vectors in SVM?

(d.) What is the Basic Principle of a support vector Machine?

19. (a.) What are the Hard margin and soft margin in SVM?

(b.) What is the role of C - Hyper Parameter in SVM?

(c.) Explain different types of kernel functions in SVM?

20. (a) What effects the decision boundary in SVM?

(b.) What is ~~Good~~ boot strapping in Bagging and Random forest?

(c.) Write down the steps of Random Forest Algorithm.