

Batch DS2311

Name- Akash Yadav.

MACHINE LEARNING

1. Which of the following methods do we use to find the best fit line for data in Linear Regression?

Ans-Least Square Error

2-Which of the following statement is true about outliers in linear regression?

Ans- Linear regression is sensitive to outliers

3. A line falls from left to right if a slope is _____?

Ans -Negative

4-Which of the following will have symmetric relation between dependent variable and independent variable?

Ans- Regression

5-Which of the following is the reason for over fitting condition?

Ans --High bias and high variance

6. If output involves label then that model is called as:

Ans-- Descriptive model

7-Lasso and Ridge regression techniques belong to _____?

Ans ---Regularization

8- To overcome with imbalance dataset which technique can be used?

Ans-- Cross validation

9- The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary classification problems. It uses _____ to make graph?

Ans --Sensitivity and Specificity

10. In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less.

Ans – False

11. Pick the feature extraction from below

Ans- Apply PCA to project high dimensional data

12. Which of the following is true about Normal Equation used to compute the coefficient of the Linear Regression?

Ans --A) We don't have to choose the learning rate. B) It becomes slow when number of features is very large. C) We need to iterate.

13-Explain the term regularization?

Ans --Regularization is a technique to resolve the issues of overfitting and underfitting by the help of Lasso and Ridge.

14. Which particular algorithms are used for regularization?

Ans -Ridge Regression (L2),Lasso (L1).

15 Explain the term error present in linear regression equation?

Ans $y=a+bx+e$

Y= outcase

b= coeff value . slope

a= intercept

x= independent variable

e = error

error is the mean absolute error . occurs by the difference of testing and training data.

STATISTICS WORKSHEET-1

1. Bernoulli random variables take (only) the values 1 and 0

Ans- True

2. Which of the following theorem states that the distribution of averages of iid variables, properly normalized, becomes that of a standard normal as the sample size increases?

Ans -) Central Limit Theorem

3. Which of the following is incorrect with respect to use of Poisson distribution?

Ans Modeling bounded count data

4. Point out the correct statement.

Ans - All of the mentioned

5. _____ random variables are used to model rates

Ans- Poission

6. 10. Usually replacing the standard error by its estimated value does change the CLT.

Ans. False

7. 1. Which of the following testing is concerned with making decisions using data?

ANS- Hypothesis

8. - 4. Normalized data are centered at _____ and have units equal to standard deviations of the original data

Ans- 0

9. Which of the following statement is incorrect with respect to outliers?

Ans- Outliers cannot conform to the regression relationship

10. What do you understand by the term Normal Distribution?

Ans- Also Known as Gaussian distribution. The distribution which is symmetric about mean . The data near mean is having more occurrence then the data far away from mean.

11. How do you handle missing data? What imputation techniques do you recommend?

Ans. Substitution, Hot deck imputation, Cold deck imputation, Regression imputation.

12. What is A/B testing?

Ans. Its is the comparisons of two variable to find out the better one.

13. Is mean imputation of missing data acceptable practice?

Ans. No its is not a acceptable practice

14. What is linear regression in statistics?

Ans $y=a+bx$

Y= dependent variable

b= coeff value . slope

a= intercept

x= independent variable

The relationship between dependent variable and independent variable

15. What are the various branches of statistics?

Ans. There are two branches of statistic:-

descriptive statistics and inferential statistics.

PYTHON – WORKSHEET 1

1. Which of the following operators is used to calculate remainder in a division?

Ans- %

2- In python 2//3 is equal to?

Ans- 0

3 In python, 6<<2 is equal to

Ans – 24

4- In python, 6&2 will give which of the following as output?

Ans- 2

5- In python, 6|2 will give which of the following as output?

Ans- 6

6- What does the finally keyword denotes in python?

Ans - the finally block will be executed no matter if the try block raises an error or not.

7- What does raise keyword is used for in python?

Ans- It is used to raise an exception

9- . Which of the following are the valid variable names?

Ans- 1abc

10. Which of the following are the keywords in python?

Ans yield , raise