

MACHINE LEARNING

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

Answer A) Least Square Error

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

Answer A) Linear regression is sensitive to outliers

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

Answer B) Negative

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

Answer A) Regression

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

Answer C) Low bias and high variance

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

Answer B) Predictive modal

7. Lasso and Ridge regression techniques belong to _____?

Answer D) Regularization

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

Answer A) Cross validation

9. The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary

classification problems. It uses _____ to make graph?

Answer A) TPR and FPR

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

curve should be less.

Answer B) False

11. Pick the feature extraction from below:

Answer B) 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?

Answer

A) We don't have to choose the learning rate.

B) It becomes slow when number of features is very large.

13. Explain the term regularization?

Answer

Regularization= when we use regression model to train some data, there is a good chance that the model will overfit the given training data set. regularization helps sort this overfitting problem by restricting the degree of freedom of a given equation i.e. simply reducing the no. of degree of polynomial function by reducing their corresponding weights

14. Which particular algorithms are used for regularization?

Answer

LASSO: (least absolute shrinkage and selection operator)[L1 form]

Ridge regression[L2 form]:

15. Explain the term error present in linear regression equation?

Answer

ERROR/RESIDUALS=(ACTUAL - PREDICTED)

The error terms (i.e., $\epsilon_1, \epsilon_2, \dots, \epsilon_n$) follow a "multivariate normal distribution" with a mean of zero and a constant variance of σ^2 .

PYTHON – WORKSHEET 1

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

Answer C) %

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

Answer 0

3. In python, 6<<2 is equal to?

Answer C) 24

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

Answer A) 2

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

Answer D) 6

6. What does the finally keyword denotes in python?

Answer C) 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?

Answer A) It is used to raise an exception

8. Which of the following is a common use case of yield keyword in python?

Answer A) in defining an iterator

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

Answer

A) _abc

B) 1ab

C) abc2

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

Answer A) yield

B) raise

STATISTICS WORKSHEET-1

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

Answer a) 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?

Answer a) Central Limit Theorem

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

Answer b) Modeling bounded count data

4.

Point out the correct statement.

Answer

d) All of the mentioned

5. _____ random variables are used to model rates.

Answer c) Poisson

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

Answer b) False

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

Answer b) Hypothesis

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

original data.

Answer a) 0

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

Answer c) Outliers cannot conform to the regression relationship

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

Answer

Normal distribution, also known as the Gaussian distribution, is a probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean

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

Handling Missing Data.

Data Dropping.

Mean/Median Imputation.

Random Sample Imputation.

Multiple Imputation.

12. What is A/B testing?

Answer

A/B testing (also known as split testing or bucket testing) is a methodology for comparing two versions of a webpage or app against each other to determine which one performs better.

13. Is mean imputation of missing data acceptable practice?

Mean imputation (MI) is one such method in which the mean of the observed values for each variable is computed and the missing values for that variable are imputed by this mean. This method can lead into severely biased estimates even if data are MCAR ,

14. What is linear regression?

Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable.

15. What are the various branches of statistics?

Answer

descriptive and Inferential Statistics

The two major areas of statistics are known as descriptive statistics, which describes the properties of sample and population data, and inferential statistics, which uses those properties to test hypotheses and draw conclusions.

