Worksheet 1_SuC_Python

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 - B) 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 - B) It encloses the lines of code which will be executed if any error occurs while executing the lines of code in the try block.

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 - C) in defining a generator

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

A) _abc C) abc2

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

Answer - A) yield B) raise

Worksheet 1_SuC_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 - C) Zero

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

Answer - B) Correlation

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 - D) SMOTE

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

Answer - C) Sensitivity and Specificity

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?

Answers -

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?

Answer - While training a machine learning model, the model can easily be overfitted or under fitted. To avoid this, we use regularization in machine learning to properly fit a model onto our test set. Regularization techniques help reduce the chance of overfitting and help us get an optimal model.

14. Which particular algorithms are used for regularization?

Answer - LASSO, Ridge, and Elastic-Net regression.

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

Answer - An error term represents the margin of error within a statistical model; it refers to the sum of the deviations within the regression line, which provides an explanation for the difference between the theoretical value of the model and the actual observed results.

Worksheet 1_SuC_Statistics

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.

a) The exponent of a normally distributed random variables follows what is called the log- normal distribution

b) Sums of normally distributed random variables are again normally distributed even if the variables are dependent

c) The square of a standard normal random variable follows what is called chi-squared distribution

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 - a) True

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?

c) Outliers cannot confirm to the regression relationship

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

Answer - **Normal distribution**, the most common distribution function for independent, randomly generated variables. Its familiar bell-shaped curve is everywhere in statistical reports, from survey analysis and quality control to resource allocation.

The graph of the normal distribution is characterized by two parameters: the mean, or average, which is the maximum of the graph and about which the graph is always symmetric; and the standard deviation, which determines the amount of dispersion away from the mean.

A small standard deviation (compared with the mean) produces a steep graph, whereas a large standard deviation (again compared with the mean) produces a flat graph.

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

Answer - According to data scientists, there are three types of missing data. These are Missing Completely at Random (MCAR) – when data is completely missing at random across the dataset with no discernable pattern. There is also Missing At Random (MAR) – when data is not missing randomly, but only within sub-samples of data. Finally, there is Not Missing at Random (NMAR), when there is a noticeable trend in the way data is missing.

Ways -

- Mean Impulsion
- Multivariate Imputation by Chained Equations
- Random Forest

12. What is A/B testing?

Answer - It is a method of 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?

Answer - True, imputing the mean preserves the mean of the observed data. So if the data are missing completely at random, the estimate of the mean remains unbiased.

14. What is linear regression in statistics?

Answer - Linear regression quantifies the relationship between one or more predictor variable(s) and one outcome variable. Linear regression is commonly used for predictive analysis and modeling.

15. What are the various branches of statistics?

Answer - There are three real branches of statistics: data collection, descriptive statistics and inferential statistics.