Python Worksheet 1

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

Ans: C %

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 output?

Ans: 2

5. In Python 6|2 will give which output?

Ans: 6

6. What does the finally keyword denotes in python?

Ans: 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?

Ans: A) It is used to raise an exception

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

Ans: C) in defining a generator

# WORKSHEET STATISTICS WORKSHEET-1

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

Ans: 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?

Ans: a) Central Limit Theorem

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

Ans: c) Modeling contingency tables

4. Point out the correct statement.

Ans: d) All of the mentioned

5. \_\_\_\_\_\_ random variables are used to model rates.

Ans: c) Poisson

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

Ans: b) False

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

Ans: b) Hypothesis

8. Normalized data are centered at\_\_\_\_\_\_and have units equal to standard deviations of the original data.

Ans: a) 0

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

Ans: c) Outliers cannot conform to the regression relationship

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

Ans: The distribution of data is normal when it is symmetric around mean.

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

Ans: we can use few techniques to handle missing data. Eg. Deletion method to eliminate missing data, or regression analysis, data imputation technique.

Would recommend average imputation technique where we can use average value from the other data entries to fill missing values.

12. What is A/B testing?

Ans: In A/B testing, we do compare two data sets and will check which one is better.

13. Is mean imputation of missing data acceptable practice?

Ans: No, mean imputation is not a good practice for missing data.

14. What is linear regression in statistics?

Ans: It is a regression model where one variable is dependent on another variable

15. What are the various branches of statistics?

Ans: Data collection, Descriptive statistics and inferential statistics

# MACHINE LEARNING

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

Ans: A) Least Square Error

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

Ans: A) Linear regression is sensitive to outliers.

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

Ans: B) Negative

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

Ans: B) Correlation

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

Ans: C) Low bias and high variance

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

Ans: B) Predictive modal

7. Lasso and Ridge regression techniques belong to \_\_\_\_\_\_\_\_\_?

Ans: D) Regularization

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

Ans: D) SMOTE

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

Ans: C) Sensitivity and Specificit

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

Ans: B) False

11. Pick the feature extraction from below:

Ans: C) Removing stop words

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.

D) It does not make use of dependent variable.

13. Explain the term regularization?

Ans: In Machine Learning, regularization is a technique used to minimize the chances of overfitting or underfitting. It helps to reduce the errors.

Two types of regularization

1. Ridge Regularization
2. Lasso Regularization

14. Which particular algorithms are used for regularization?

Ans:

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

Ans: Error 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.