PYTHON – WORKSHEET 1

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

2. In python 2//3 is equal to?
A) 0

3. In python, 6<<2 is equal to?
A)24

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

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

6. What does the finally keyword denotes in python?
A) 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?
A) It is used to raise an exception.

8. Which of the following is a common use case of yield keyword in python?
A) in defining an iterator

Q9 and Q10 have multiple correct answers. Choose all the correct options to answer your question.
9. Which of the following are the valid variable names?
A) _abc

10. Which of the following are the keywords in python?
A) Yield B) raise

STATISTICS WORKSHEET-1

Q1 to Q9 have only one correct answer. Choose the correct option to answer your question.

1. Bernoulli random variables take (only) the values 1 and 0.
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?
a) Central Limit Theorem

3. Which of the following is incorrect with respect to use of Poisson distribution?
a) Modeling bounded count data

4. Point out the correct statement.
a) All of the mentioned

5. _____ random variables are used to model rates.
a) All of the mentioned

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

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

8. 4. Normalized data are centered at_____and have units equal to standard deviations of the original data.
a) 0

9. Which of the following statement is incorrect with respect to outliers?
a) Outliers cannot conform to the regression relationship

10. What do you understand by the term Normal Distribution?
A. it is a type of continuous probability distribution in which most data points cluster toward the middle of the range, while the rest taper off symmetrically toward either extreme

11. How do you handle missing data? What imputation techniques do you recommend?
A. Missing data, or missing values, occur when you don't have data stored for certain variables or participants. Data can go missing due to incomplete data entry, equipment malfunctions, lost files, and many other reasons.

- Ignore the records with missing values.
- Predict missing values.

12. What is A/B testing?
A. A/B testing in its simplest sense is an experiment on two variants to see which performs better based on a given metric
A/B testing is a form of statistical and two-sample hypothesis testing.

- Statistical hypothesis testing is a method in which a sample dataset is compared against the population data.
- Two-sample hypothesis testing is a method in determining whether the differences between the two samples are statistically significant or not.

13. Is mean imputation of missing data acceptable practice?
A. Typically it is terrible practices since it ignores the feature correlation and potentially creates unrealistic values

14. What is linear regression in statistics?
A. Linear regression is commonly used for predictive analysis these regression estimates are used to explain the relationship between one dependent variable and one or more independent variables. is defined by the formula y = c + b*x, where y = estimated dependent variable score, c = constant, b = regression coefficient, and x = score on the independent variable

15. What are the various branches of statistics?
A. There are 2 types of branches they are

- Descriptive Statistics:
  The first aspect of statistics is descriptive statistics, which deals with the presentation and collection of data. It is not as simple as it appears. Which must know how to design and experiment, select the appropriate focus group, and prevent biases that are too easy to introduce into the experiment?
- Inferential Statistics:
  Inference statistics are statistical techniques that allow statisticians to utilize data from a sample to conclude, predict the behavior of a given population, and make judgments or decisions.

MACHINE LEARNING

In Q1 to Q11, only one option is correct, choose the correct option:
1. Which of the following methods do we use to find the best fit line for data in Linear Regression?
A) Least Square Error

2. Which of the following statement is true about outliers in linear regression?
A) Linear regression is sensitive to outliers

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

4. Which of the following will have symmetric relation between dependent variable and independent variable?
A) None of these

5. Which of the following is the reason for over fitting condition?
A) Low bias and high variance

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

7. Lasso and Ridge regression techniques belong to _____?
A) Cross validation

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

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

10. In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less.
A) True
11. Pick the feature extraction from below:
A) Construction bag of words from a email
B) Apply PCA to project high dimensional data

In Q12, more than one options are correct, choose all the correct options:

12. Which of the following is true about Normal Equation used to compute the coefficient of the Linear Regression?
A) We don't have to choose the learning rate.
C) We need to iterate.
D) It does not make use of dependent variable.

13. Explain the term regularization?
A. It is one of the most important concept of ML. It is used to prevent from over fitting or under fitting the machine learning model. To understand it this achieves in technical way which we need to understand the concept of bias and variance.

- Bias- the difference between the actual an predicted values. A machine learning model with high bias gives small consideration to the data pattern, resulting in under fit models.

- Variance- It measures of flexibility of the model. Which is opposite to bias and decides how the sensitive of the model to be change based on the pattern in the input data?

14. Which particular algorithms are used for regularization?
A. There are three commonly used regularization techniques to control the complexity of machine learning models, as follows:

- L2 regularization(Ridge)
- L1 regularization(Lasso)
- Elastic Net

15. Explain the term error present in linear regression equation?
A. error term represents the margin of error within a statistical model, which shows an explanation for the difference between the theoretical value of the model and the actual.