PYTHON – WORKSHEET 1

Q1 to Q8 have only one correct answer. Choose the correct option to answer your question.

1. Which of the following operators is used to calculate remainder in a division?
A) # B) &
C) % D) $
Solution: C

2. In python 2//3 is equal to?
A) 0.666 B) 0
C) 1 D) 0.67
Solution: B

3. In python, 6<<2 is equal to?
A) 36 B) 10
C) 24 D) 45
Solution: C

4. In python, 6&2 will give which of the following as output?
A) 2 B) True
C) False D) 0
Solution: A

5. In python, 6|2 will give which of the following as output?
A) 2 B) 4
C) 0 D) 6
Solution: D

6. What does the finally keyword denotes in python?
A) It is used to mark the end of the code
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.
C) the finally block will be executed no matter if the try block raises an error or not.
D) None of the above
Solution: C

7. What does raise keyword is used for in python?
A) It is used to raise an exception. B) It is used to define lambda function
C) it's not a keyword in python. D) None of the above
Solution: A

8. Which of the following is a common use case of yield keyword in python?
A) in defining an iterator B) while defining a lambda function
C) in defining a generator D) in for loop.
Solution: C

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 B) 1abc
C) abc2 D) None of the above

Solution: A, C

10. Which of the following are the keywords in python?
A) yield B) raise
C) look-in D) all of the above
Solution: A, B

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
b) False
Solution: A

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 b) Central Mean Theorem
c) Centroid Limit Theorem d) All of the mentioned
Solution: A

3. Which of the following is incorrect with respect to use of Poisson distribution?
a) Modeling event/time data b) Modeling bounded count data
c) Modeling contingency tables d) All of the mentioned
Solution: B

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
Solution: D

5. _____ random variables are used to model rates.
a) Empirical b) Binomial
c) Poisson d) All of the mentioned
Solution: C

6. 10. Usually replacing the standard error by its estimated value does change the CLT.
a) True b) False
Solution: B

7. 1. Which of the following testing is concerned with making decisions using data?
a) Probability b) Hypothesis
c) Causal d) None of the mentioned
Solution: B

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

a) 0 b) 5

c) 1 d) 10

Solution: A

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

a) Outliers can have varying degrees of influence

b) Outliers can be the result of spurious or real processes

c) Outliers cannot conform to the regression relationship

d) None of the mentioned

Solution: C

Q10and Q15 are subjective answer type questions, Answer them in your own words briefly.

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

Solution: It is also called as Gaussian distribution, this is one among the imp topics of statistics, this is imp because when you work on any Machine learning algorithms the data should satisfy the normal distribution, so that the model gets trained well and we will receive good predictions

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

Solution: Missing data can be dealt with many number of ways,as per my view if the missing data is very less number in count we can drop those rows , if not we can use the mean imputation where we calculate the mean and assign to missing values

12. What is A/B testing?

Solution: It is used to test/compare between different products based on the user assumptions /behaviour

Example: Watching datascience content in youtube videos shows the adds related to datascience course.

13. Is mean imputation of missing data acceptable practice?

Solution: As per my view this can be decided based on the data,it is not acceptable for some data which ignore feature corelation, for **Example** we have a table with age and fitness scores a small child has the fitness score missing when we replace the missing value with mean it basically shows the high fitness level.

14. What is linear regression in statistics?

Solution: Linear Regression is a basic and commonly used type of predictive analaysis,this comes under supervised learning under the categeory regression,basically we will be using it whenever continuous data is present

The major useses of regression analaysis

1.determining strength of predictors

2.forecasting an effect

3.trend forecasting

15. What are the various branches of statistics?

Solution: These are the various branches of statistics

- **Descriptive(**Organizing and summarizing the data**)**
  - **Central Tendency**
    - **Mean**
    - **Median**
    - **Mode**
  - **Despersion of data**
    - **Range**
    - **Variance**
    - **Standarad Deviation**
    - **Skewness**
    - **Percentile**
- **Inferential(Tec**hnique used for the data that we have measured to form conclusions**)**
  - **Hypothesis Testing**
    - **T –test**
    - **Chi-square test**
    - **Correaltion Test**
    - **Annova Test**

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 B) Maximum Likelihood
C) Logarithmic Loss D) Both A and B

Solution: A

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

A) Linear regression is sensitive to outliers B) linear regression is not sensitive to outliers
C) Can't say D) none of these

Solution: A

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

A) Positive B) Negative
C) Zero D) Undefined

Solution: B

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

A) Regression B) Correlation
C) Both of them D) None of these

Solution: B

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

A) High bias and high variance B) Low bias and low variance
C) Low bias and high variance D) none of these

Solution: C

6. If output involves label then that model is called as:
A) Descriptive model B) Predictive modal
C) Reinforcement learning D) All of the above
Solution: B

7. Lasso and Ridge regression techniques belong to _____?
A) Cross validation B) Removing outliers
C) SMOTE D) Regularization
Solution: D

8. To overcome with imbalance dataset which technique can be used?
A) Cross validation B) Regularization
C) Kernel D) SMOTE
Solution: D

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 B) Sensitivity and precision
C) Sensitivity and Specificity D) Recall and precision
Solution: A

10. In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less.
A) True B) False
Solution: B

11. Pick the feature extraction from below:
A) Construction bag of words from a email B) Apply PCA to project high dimensional data
C) Removing stop words D) Forward selection
Solution: B

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.
B) It becomes slow when number of features is very large.
C) We need to iterate.
D) It does not make use of dependent variable.
Solution: A,B

MACHINE LEARNING

Q13 and Q15 are subjective answer type questions, Answer them briefly.
13. Explain the term regularization?
Solution: Regularizations are techniques used to reduce the error by fitting a function appropriately on the given training set and avoid overfitting. The word regularize means to make things regular or acceptable. This is exactly why we use it for.

In simple words Regularization is a technique used in regression to reduce the complexity of the model and to shrink the coefficients of the independent features.

14. Which particular algorithms are used for regularization?

Solution: The commonly used regularization techniques are :

L1 regularization

L2 regularization

Dropout regularization

- A regression model which uses L1 Regularization technique is called LASSO(Least Absolute Shrinkage and Selection Operator) regression.
    - Lasso Regression adds "absolute value of magnitude" of coefficient as penalty term to the loss function(L).
- A regression model that uses L2 regularization technique is called Ridge regression.
    - Ridge regression adds "squared magnitude" of coefficient as penalty term to the loss function(L).

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

Solution: An error term is a residual variable produced by a statistical or mathematical model, which is created when the model does not fully represent the actual relationship between the independent variables and the dependent variables. As a result of this incomplete relationship, the error term is the amount at which the equation may differ during empirical analysis.

The error term is also known as the residual, disturbance, or remainder term, and is variously represented in models by the letters e, ε, or u.