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=B)0

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

ANS=C)24

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

ANS=A) 2

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

ANS=D,

6. What does the finally keyword denotes in python?

ANS=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?

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.

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

ANS=A) _abc, B) abc2

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

ANS=A) YIELD, B) RAISE

MACHINE LEARNING WORKSHEET 1

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

ANS=D) Both A and B

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=c) Both of them

5)Which of the following is the reason for over fitting condition?
ANS= 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) TPR AND FPR

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

ANS=TRUE

11)Pick the feature extraction from below:
Ans=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?

ANS=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?
Ans= Regularization is a process for regularizing or calibrating machine learning models to reduce loss function and prevent overfitting or underfitting.
There are two types of regularization:-
L1 Regularization and L2 Regularization.

14. Which particular algorithms are used for regularization?
Ans= Lasso Regularization
Ridge Regularization
ElasticNet Regularization
15. Explain the term error present in linear regression equation?
Ans= In linear regression, the error term is the difference between the expected price at a particular time and the price that was actually observed. In instances where the price is exactly what was anticipated at a particular time, the price will fall on the trend line and the error term will be zero.
For equation we write as :-
a=c+bx+e, where e is the error term.

Statistics worksheet 1

1)Bernoulli random variables take (only) the values 1 and 0.
Ans=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=Modelling bounded count data.
4) Point out the correct statement.
Ans=All of the mentioned.
5) _____ random variables are used to model rates.
Ans=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= 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.

Q10and Q15 are subjective answer type questions, Answer them in your own words briefly. 10) What do you understand by the term Normal Distribution?

Ans= Normal distribution or Gaussian Distribution are where mean=median=mode. In Normal distribution all the data are more frequently around mean then far from mean i.e., outliers are under the value 3. Normal distribution is also called as bell curve in graph.

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

ANS=Missing Data is too problematic situation for analysing data. It can be handled either by choosing the average values of the data near the missing values or the most common Data occurring in the dataset. It can be either mean, median or mode. We can also delete the cells of missing data. Imputation techniques: 1)Complete Case Analysis. 2)Arbitrary Analysis. 3)Frequent Category Imputation.

12) What is A/B testing?

Ans= A/B is similar to hypothesis testing where two datasets are compared and based on that it determines the statistical significant relationship between them.it shows the relationship between two datasets. It is an analytical method for making decisions that estimates population parameters based on sample statistics. It can also be called as statistical inference.

13) Is mean imputation of missing data acceptable practice?

Ans=Yes mean imputation is an acceptable practice but it has its own limitations and boundraies. Mean imputation of missing data can only be worked where there are random missing data and where data should not be more informative and does not bias the data.

14) What is linear regression in statistics?

Ans=Linear Regression is the linear approach of the modelling the relationship between the dependant variables and independent variables .It shows the relation between variables with a straight line and it is also derived by linear equation.

15) What are the various branches of statistics?

Ans=Branches of statistics:-
a) Inferential Statistics.
b) Descriptive Statistics.