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

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

Ans:C

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

Ans:A

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:C

7.What does raise keyword is used for in python?

Ans:A

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

Ans:C

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

Ans:A,C

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

Ans:A,B

STATISTICS WORKSHEET-1

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

Ans: 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?

Ans: a

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

Ans: b

4.Point out the correct statement

Ans: d

5._____ random variables are used to model rates.

Ans: c

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

Ans: b

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

Ans: b

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

Ans: a

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

Ans: c

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

Ans: Normal distribution, also known as the Gaussian distribution, is a probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean. In graph form, normal distribution will appear as a bell curve.

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

Ans:

1.Use deletion methods to eliminate missing data.

2.Use regression analysis to systematically eliminate data

3.Using imputation techniques

Various imputation techniques:

1. Imputation Using (Mean/Median) Values.
2. Imputation Using (Most Frequent) or (Zero/Constant) Values.
3. Imputation Using k-NN

12. What is A/B testing?

Ans: A/B testing is a basic randomized control experiment. It is a way to compare the two versions of a variable to find out which performs better in a controlled environment.

13.Is mean imputation of missing data acceptable practice?

Ans: Bad practice in general.Mean reduces a variance of the data.Mean ignores feature correaltion.

14. What is linear regression in statistics?

Ans: Linear regression attempts to model the relationship between two variables by fitting a linear equation to observed data. A linear regression line has an equation of the form Y = a + bX, where X is the explanatory variable and Y is the dependent variable.

15. What are the various branches of statistics?

Ans: 1.Descriptive statistics

2.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

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

Ans: A

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

Ans: B

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

Ans: B

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

Ans: C

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

Ans: B

7.Lasso and Ridge regression techniques belong to _____?

Ans: D

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

Ans: D

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

Ans: A

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

Ans: B

11.Pick the feature extraction from below:

Ans: A,B,C

12.Which of the following is true about Normal Equation used to compute the coefficient of the Linear Regression?

Ans: A,B

13.Explain the term regularization?

Ans: Regularization is a technique used for tuning the function by adding an additional penalty term in the error function. The additional term controls the excessively fluctuating function such that the coefficients don't take extreme value

14.Which particular algorithms are used for regularization?

Ans: There are three main regularization techniques, namely: Ridge Regression (L2 Norm) Lasso (L1 Norm) Dropout

15Explain the term error present in linear regression equation?

Ans: Within a linear regression model the error term is the difference between the expected price at a particular time and the price that was actually observed.