PYTHON – WORKSHEET 1

1. C-%

2.B--0

3. C---24

4. A--2

5. D--6

6. A

7. A

8. C

9. A & C

10. A & B

STATISTICS WORKSHEET-1

1, True

2. a) Central Limit Theorem

3. b) Modeling bounded count data

4. d) All of the mentioned

5. c) Poisson

6. b) False

7. b) Hypothesis

8. a) 0

9. c) Outliers cannot conform to the regression relationship

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

Normal Distribution is a statitical phenomenon representing a symmetric bell-shaped

curve, where mean, median and mode are equal, Data are spread equally towards left and right.

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

Missing data are handled by Imputation or Removing data, Removing data is not a good technique, In this case model may not performe well. Imputation is a good technique to handle missing data, there are 2 type of Imputation (Single imputation and Multiple imputation(Nominal,Ordinal,scale)). Singtle imputation handle by mean,median and mode, mean and median use in case of numeric data and mode is use in case of charecter data.

Best imputation technique is Multiple Imputation.

12. What is A/B testing?

A/B testing help us to compare two versions of something to learn which is more effective.

13. Is mean imputation of missing data acceptable practice?

No, Mean imputataion ignore the relationships between variables.

14. What is linear regression in statistics?

Linear regression is a linear approach for modelling the relationship between two variables. There are two types of variable(Dependent and Indepedent) in statitics.

15. What are the various branches of statistics?

Biostatistics, Population statistics, Official Statistics, Econometric, Psychometrics, Industrial Statistics, Computint Statistics, Actuarial, Physics Statistics...

MACHINE LEARNING

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 _____?

B) Negative

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

B) Correlation

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

C) Low bias and high variance

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

B) Predictive modal

7. Lasso and Ridge regression techniques belong to _____?

D) Regularization

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

D) SMOTE

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.

B) False

11. Pick the feature extraction from below:

A) Construction bag of words from a email

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

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

13. Explain the term regularization?

Regularization is one of the techinique through which Lasso, Ridge or ElasticNet algorithm will try to reduce down the underfitting or overfitting present in any of the model or Bias-variance having in that model could be try to reduce.

14. Which particular algorithms are used for regularization?

Lasso, Ridge and Elastic-Net Regressions are use for regularization.

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

Error means the model is not completely accurate, Error is the difference between the actual value and Predicted value.

Y=a+bx+e

Y=Label/Output/Target(Dependent variable(

b=coefficient value

x=Slope(Independent variable)

e=error value.