Name –kajal panwar

Intership no-28

From-data trained acadmy

Batch no=1841

PYTHON – WORKSHEET

Answers of the worksheet one of python

1) Option C -----(%)
2) Option B-----(0)
3) Option C-----(24)
4) Option A-----(2)
5) Option D-----(6)
6) Option C----the finally block will be executed no matter if the try block raises an error or not.
7) Option A---- It is used to raise an exception.
8) Option C--- in defining a generator
9) Qption A,B,C all are correct
10) Option A and B----yield and raise

Question 11 to 15 are in Jupiter notebook

STATISTICS WORKSHEET-1

1) takes 0 and 1 because it arise as a result of a binary outcome------(a)

2) central limit theorem ---(a)

3) modeling bounded data---(b)

4) All of the above-----(d)

5) passion----(c)

6) false---(b)

7) hypothesis---(b)

8) Zero----(a)

9) outliers cannot confirm to the regression relationship----(c)

10) Normal distribution -----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) The real-world data often has a lot of missing values. The cause of missing values can be data corruption or failure to record data. The handling of missing data is very important during the preprocessing of the dataset as many machine learning algorithms do not support missing valuesThere are 7 ways to handle missing values in the dataset:

1. Deleting Rows with missing values
2. Impute missing values for continuous variable

3. Impute missing values for categorical variable
4. Other Imputation Methods
5. Using Algorithms that support missing values
6. Prediction of missing values
7. Imputation using Deep Learning Library — Datawig

12) 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) he process of replacing null values in a data collection with the data'smean is known as mean imputation. Mean imputation is typically considered terrible practice since it ignores feature correlation mean imputation decreases the variance of our data while increasing bias. As a result of the reduced variance, the model is less accurate and the confidence interval is narrower.

14) Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable.

15) There are three real branches of statistics:

a) data collection

b) descriptive statistics and

c)inferential statistics.

MACHINE LEARNING

1) option A-Least square error

2) option A-Linear regression is sensitive to outliers

3) option –B-negative

4) option –A-regression

5) option-C-Low bias and high variance

6) option-C) Reinforcement learnin

7)option- D) Regularization

8)option- D) SMOTE

9)option- C) Sensitivity and Specificity

10)option-) B-true

11)option- B) Apply PCA to project high dimensional data

12)option 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) Regularization refers to techniques that are used to calibrate machine learning models in order to minimize the adjusted loss function and prevent overfitting or underfitting

14) Ridge Regression.

LASSO (Least Absolute Shrinkage and Selection Operator)
Elastic-Net Regression.

15) The standard error of the regression is the average distance that the observed values fall from the regression line.