Python worksheet -1

Q-1- Which of the following operators is used to calculate remainder in a division?

- ⇨ C) %

Q-2 -In python 2//3 is equal to?

- ⇨ B) 0

Q-3- In python, 6<<2 is

- ⇨ C) 24

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

- ⇨ A) 2

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

- ⇨ D)6

Q-6-What does the finally keyword denotes in python?

- ⇨ C) the finally block will be executed no matter if the try block raises an error or not.

Q-7-What does raise keyword is used for in python?

- ⇨ A) It is used to raise an exception.

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

- ⇨ C) in defining a generator

Q-9- Which of the following are the valid variable names?

- ⇨ _abc
- ⇨ Abc2

Q-10-Which of the following are the keywords in python?

- ⇨ Yield
- ⇨ Raise

In [73]:
```
#Q-11 : Find the factorial of a number
def factorial(n):
    if n <0:
        return 1
    elif n ==0 or n==1:
        return 1
    else :
        fact =1
        while(n>1):
            fact = fact*n
            n=n-1


        return fact

a = int(input("Enter the  number :"))
print("Factorial of", a , ' :' , factorial(a))
```

In [72]:
```
#Q-12 : Check if thr number is prime or composite
def checkPrime(n):
    if n >1:
        for i in range(2,int(n/2)+1):
            if (n%i)==0:
                print(n ," not prime")
                break
            else:
                print("number is prime")
    else:
        print("number is not prime")

number=5
print(checkPrime(number))
```

```
number is prime
None
```

In [70]:
```
#Q-13:Check wether string is palindrome or not
def checkPalindrome(str):
     for i in range(0, int(len(str)/2)):
        if str[i] != str[len(str)-i-1]:
            return False
     return True

s = "Sameer"
isPalindrome = checkPalindrome(s)
if(isPalindrome):
    print(str , " is palindrome")
else:
    print(str , " is not palindrome")
```

```
Sameer  is not palindrome
```

In [75]:
```python
#Q-14:get the third side of right angled triangle
import numpy as np

def thirdSide(a,b):
    c= np.sqrt((a*a)+(b*b))
    return c

print("third side of triangle is : " , thirdSide(10,20))
```

```
third side of triangle is :  22.360679774997898
```

In [77]:
```python
#Q-15: get the frquency of each characer in the string
def frequency(str):
    freq = {}

    for i in str:
        if i in freq:
            freq[i] += 1
        else:
            freq[i] = 1
    return freq



Str = "Sameer"
print( "Counts of each character is : " , frequency(str))
```

```
Counts of each character is :  {'S': 1, 'a': 1, 'm': 1, 'e': 2, 'r': 1}
```

Statistics Worksheet 1

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

- ⇨ a) True

Q-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

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

- ⇨ c) Modeling contingency tables

Q-4-Point out the correct statement

- ⇨ d) All of the mentioned

Q-5-_____ random variables are used to model rates

- ⇨ c) Poisson

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

- ⇨ b) False

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

- ⇨ b) Hypothesis

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

- ⇨ a) 0

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

- ⇨ c) Outliers cannot conform to the regression relationship

Q-10-What do you understand by the term 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.

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

- ⇨ We handle missing data sets using imputing techniques. These methods are used as the process using means , mode of the particular type of columns and replacing the missing data sets using such techniques. There are different type of imputing techniques but most commonly used are –
  a)KNN Imputer
  b)Iterative Imputer

Q-12-What is A/B testing?

- ⇨ **A/B testing** is one of the most popular controlled experiments used to optimize web marketing strategies. It allows decision makers to choose the best design for a website by looking at the analytics results obtained with two possible alternatives A and B.

Q-13- Is mean imputation of missing data acceptable practice?

- ⇨ Yes, For regression data or classification data , mean imputation can be used as it gives a suitable value for the missing data. But its not recommended to use it for categorical data as it can disturb the data set.

Q-14-What is linear regression in statistics?

- ⇨ Linear Regression contains both dependent and independent variables where we can find predict the dependent/output variable by plotting the independent variables. Its also represents a straight line with the equation as -
  Y= MX+C.
  Y - dependent variable
  M-Slope
  C – Intercept
  X- independent variable

Q-15-What are the various branches of statistics?

- ⇨ There are three real branches of statistics: data collection, descriptive statistics and inferential statistics

Machine Learning Worksheet-1

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

- ⇨ A) Least Square Error

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

- ⇨ A) Linear regression is sensitive to outliers

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

- ⇨ B) Negative

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

- ⇨ C) Both of them (Regression and correlation)

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

- ⇨ C) Low bias and high variance

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

- ⇨ B) Predictive modal

Q-7-Lasso and Ridge regression techniques belong to _____?

- ⇨ D) Regularization

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

- ⇨ D) SMOTE

Q-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

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

- ⇨ B) False

Q-11-Pick the feature extraction from below:

- ⇨ C) Removing stop words

Q-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.

Q-13-Explain the term regularization?

- ⇨ Regularization is a technique used to reduce the errors by fitting the function appropriately on the given training set and avoid overfitting

Q-14-Which particular algorithms are used for regularization?

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

Q-15-Explain the term error present in linear regression equation?

- ⇨ an error term is the term in a model regression equation that tallies up and accounts for the unexplained difference between the actually observed values of the independent variable and the results predicted by the model.