

Internship
13-09-2021
Amit Kumar Tiwari

Index

1. **Python Worksheet-1**
2. **Statistics worksheet-1**
3. **Machine Learning – 1**

1. Python Worksheet-1

Remark – kindly read the answers in “Value (option number)” pattern.

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-0(B)

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

ANS-24(C)

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

ANS – 2(A)

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

ANS – 6(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-(A)

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

ANS-(A)

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

ANS- (A) & (C)

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

ANS – (A) & (B)

12. Write a python program to find the factorial of a number.

ANS

```

13. x = int(input("Enter a number: "))
14. factorial = 1
15.
16. if x<0:
17.     print("factorial doesnt exist for -ve number")
18. elif x==0:
19.     print("factorial for 0 is 1")
20. else:
21.     for i in range(1, x+1):
22.         factorial = factorial*i
23. print("factorial of", x, "is", factorial)

```

24. Write a python program to find whether a number is prime or composite.

ANS

```

x = int(input("Enter a number: "))
if x<1:
    for i in range(1, int(x/2)+1):
        if (x%i==0):
            print(x, "Number is not a prime number")
            break
        else:
            print(x, "Number is a prime number")
else:
    print(x, 'Number is not a prime number')

```

25. 13. Write a python program to check whether a given string is palindrome or not.

ANS

```

26. word = input()
27. if word==word[::-1]:
28.     print("The word is Palindrome")
29.
30. else:
31.     print("The word is not a Palindrome")

```

32. 14. Write a Python program to get the third side of right-angled triangle from two given sides.

ANS

```

x = int(input("x = "))
y = int(input("y = "))

z = x**2+y**2
import math
print(math.sqrt(z))

```

33. 15. Write a python program to print the frequency of each of the characters present in each string.

ANS

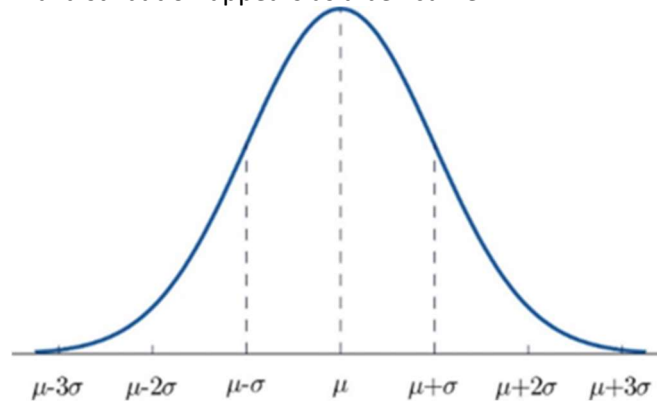
```
34. test_str = input()
35.
36. all_freq = {}
37.
38. for i in test_str:
39.     if i in all_freq:
40.         all_freq[i] += 1
41.     else:
42.         all_freq[i] = 1
43. print ("Count of all characters in GeeksforGeeks is :\n"+ str(all_freq))
```

2. Statistics Worksheet-1

1. Bernoulli random variables take (only) the values 1 and 0.
ANS – True(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 – (B)
5. 5. _____ random variables are used to model rates.
ANS – (A)
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 – (B)
10. What do you understand by the term Normal Distribution?
ANS – Normal distribution, otherwise known as the Gaussian distribution, is a probability distribution that denotes the symmetric property of the mean. The idea behind this function is

that the data near the mean occurs more frequently than the data away from the mean. It infers that the data around the mean represents the entire data set.

Like PDF, the normal distribution appears as a bell curve:

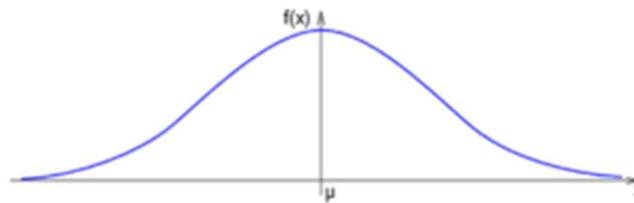


Normal Distribution – Statistics and Probability

The graph of the Normal Distribution depends on two factors: the Mean and the Standard Deviation

- Mean: Determines the location of the center of the graph
- Standard Deviation: Determines the height of the graph

If the standard deviation is large, the curve is short and wide:



Standard Deviation Curve

If the standard deviation is small, the curve is tall and narrow:



Standard Deviation Curve – Statistics and Probability

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

ANS –

- A. Mean or Median Imputation
- B. Multivariate Imputation by Chained Equations (MICE)
- C. Random Forest

12. What is A/B testing?

ANS –

A/B testing is a shorthand for a simple controlled experiment. in which two samples (A and B) of a single vector-variable are compared. These values are similar except for one variation which might affect a user's behavior. A/B tests are widely considered the simplest form of controlled experiment. However, by adding more variants to the test, its complexity grows.

A/B tests are useful for understanding user engagement and satisfaction of online features like a new feature or product. Large social media sites like LinkedIn, Facebook, and Instagram use A/B testing to make user experiences more successful and as a way to streamline their services.

13. Is mean imputation of missing data acceptable practice?

ANS –

Yes, but in recommendation it should be properly observed & tasted of other imputations are good we may go ahead with them as well.

Imputing the mean preserves the mean of the observed data. So, if the data are missing completely at random, the estimate of the mean remains unbiased.

14. What is linear regression in statistics?

ANS -

Linear regression quantifies the relationship between one or more predictor variable(s) and one outcome variable. Linear regression is commonly used for predictive analysis and modeling.

For example, it can be used to quantify the relative impacts of age, gender, and diet (the predictor variables) on height (the outcome variable). Linear regression is also known as multiple regression, multivariate regression, ordinary least squares (OLS), and regression.

15. What are the various branches of statistics?

ANS -

- A. Descriptive statistics
- B. Inferential statistics

C.Machine Learning-1

1. The computational complexity of linear regression is:

ANS- (B)

2. Which of the following can be used to fit non-linear data?

ANS- (B)

3. Which of the following can be used to optimize the cost function of Linear Regression?

ANS- (B)

4. Which of the following method does not have closed form solution for its coefficients?

ANS- (D)

5. Which gradient descent algorithm always gives optimal solution?

ANS- (A)

6. Generalization error measures how well a model performs on training data.

ANS- (A)

7. The cost function of linear regression can be given as $J(w_0, w_1) = \frac{1}{2m} \sum_{i=1}^m (w_0 + w_1 x(i) - y(i))^2$. The half term at start is due to:

ANS- (A)

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

ANS- (D)

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

ANS- (A) & (B)

10. Which of the following statement/s are true if we generated data with the help of polynomial features with 5 degrees of freedom which perfectly fits the data?

ANS- (A) & (B)

11. Which of the following sentence is false regarding regression?

ANS- (C) & (D)

12. Which Linear Regression training algorithm can we use if we have a training set with millions of features?

ANS- (C) & (D)

Since there are lots of features, we cannot use Normal Equations (it will be very, very computationally expensive). Instead, we can use Gradient Descent.

13. Which algorithms will not suffer or might suffer, if the features in training set have very different scales?

ANS- (C) & (D)

The Gradient Descent suffers from features of different scales because the model will take a longer time to reach the global maximum. We can always scale the features to eliminate this problem.