MACHINE LEARNING

In Q1 to Q11, only one option is correct, choose the correct option:

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?

D) none of these

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 variable?

C) Both of them

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

A) High bias and high variance

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

A) Descriptive model

7. Lasso and Ridge regression techniques belong to _____?

C) SMOTE

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

C) Kernel

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:

B) Apply PCA to project high dimensional data

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

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

Regression?

B) It becomes slow when number of features is very large.

C) We need to iterate.

Q13 and Q15 are subjective answer type questions, Answer them briefly.

13. Explain the term regularization?

Answer-

Making the resulting maps* more smooth, i.e. regular. Techniques that are used to calibrate machine learning models in order to minimize the adjusted loss function and prevent overfitting or under fitting.

14. Which particular algorithms are used for regularization?

Answer— There are three main regularization techniques, namely:

Ridge Regression (L2 Norm)

Lasso (L1 Norm)

Dropout

Ridge and Lasso can be used for any algorithms involving weight parameters.

Dropout is primarily used in any kind of neural networks e.g. ANN, DNN, CNN or RNN.

Ridge regression is also called L2 norm or regularization.

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

A type I error (false-positive) occurs if an investigator rejects a null hypothesis that is actually true in the population; a type II error (false-negative) occurs if the investigator fails to reject a null hypothesis that is actually false in the population.

STATISTICS WORKSHEET-1

Q1 to Q9 have only one correct answer. Choose the correct option to answer your question.

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

a) True

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?

d) All of the mentioned

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

a) Modeling event/time data

4. Point out the correct statement.

d) All of the mentioned

5. _____ random variables are used to model rates.

c) Poisson

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

b) False

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

b) Hypothesis

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

a) 0

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

c) Outliers cannot conform to the regression relationship

Q10and Q15 are subjective answer type questions, Answer them in your own words briefly.

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

Normal distribution is a statistic which had been widely applied for mathematical concepts.one of the first application of the normal distribution was used in astronomical observation where they found

errors of measurements. Normal distribution also knows as Gaussian distribution. Normal distributions are symmetric, unimodal, and asymptotic, and the mean, median, and mode are all equal.

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

The deletion of the rows or columns having null values. If any columns have more than half of the values as null then we can drop the entire column. In the same way, rows can also be dropped if having one or more columns values as null.

imputation techniques-

1. Delete the Data
2. Imputing Averages.
3. Assign New Category
4. Certain Algorithms.

12. What is A/B testing?

A/B testing, also known as split testing. In A/B testing, A refers to control or the original testing variable. Whereas B refers to variation or a new version of the original testing variable.it refers to a randomized experimentation process wherein two or more versions of a variable. A/B testing helps marketers observe how one version of a piece of marketing content performs alongside another

13. Is mean imputation of missing data acceptable practice?

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. Mean imputation does not preserve relationships between variables such as correlations.

14. What is linear regression in statistics?

It predicts the value of unknown data by using another related and known data value. For eg. data about your expenses and income for last year. Linear regression techniques analyze this data and determine that your expenses are half your income. They then calculate an unknown future expense by halving a future known income. Linear regression models are relatively simple and provide an easy to interpret mathematical formula to generate predictions.

15. What are the various branches of statistics.

statistics are known as descriptive statistics, which describes the properties of sample and population data and inferential statistics, which uses those properties to test hypotheses and draw conclusions.

Descriptive statistics are also categorized into four different categories:

Measure of frequency

Measure of dispersion

Measure of central tendency

Measure of position

Inferential Statistic-

the data has been collected, analyzed and summarized then we use these stats to describe the meaning of the collected data.

PYTHON – WORKSHEET 1

Q1 to Q8 have only one correct answer. Choose the correct option to answer your question.

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

C) %

2. In python 2//3 is equal to?

B) 0

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

C) 24

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

A) 2

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

D) 6

6. What does the finally keyword denotes in python?

D) None of the above

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

A) It is used to raise an exception

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

C) in defining a generator

Q9 and Q10 have multiple correct answers. Choose all the correct options to answer your question.

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

A) _abc B) 1abc

C) abc2

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

A) yield B) raise

Q11 to Q15 are programming questions. Answer them in Jupyter Notebook.

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

```
# Python program to find the factorial of a number provided by the user.


num =5

factorial = 1


if num < 0:

   print(“ factorial does not exist for negative numbers")

elif num == 0:

   print("The factorial of 0 is 1")

else:

   for i in range(1,num + 1):

       factorial = factorial*i

   print("The factorial of",num,"is",factorial)
```

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

```
number = int(input("Enter any number to find it is a prime or composite : "))
```

```
if number > 1:
   for i in range(2, number):
      if (number % i) == 0:
         print(number, "is NOT a prime number")
         break
   else:
      print(number, "is a PRIME number")
elif number == 0 or 1:
   print(number, "is a neither prime NOR composite number")
else:
   print(number, "is NOT a prime number it is a COMPOSITE number")
```

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

```
string = input("Please enter your text : ")


if(string == string[:: - 1]):
   print("This is a Palindrome String")
else:
   print("This is Not a Palindrome String")
```

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

```
from math import sqrt

print("Enter the lengths of triangle sides:")

a = float(input("a:"))

b = float(input("b:"))

c = sqrt(a**2 + b**2)

print("The length of the hypotenuse is:", c )
```

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

```
freq = {}

string=str(input("enter string: "))

for i in string:

    if i in freq:

        freq[i] += 1

    else:

        freq[i] = 1

print("frequency of string :\n" +  str(freq))
```