

# **MACHINE LEARNING**

- 1) Least Square Error
- 2) Linear regression is sensitive to outliers
- 3) Negative
- 4) Both of them
- 5) High bias and high variance
- 6) Predictive model
- 7) Regularization
- 8) SMOTE
- 9) TPR and FPR
- 10) True
- 11) Apply PCA to project high dimensional data
- 12) It becomes slow when number of features is very large.  
It does not make use of dependent variable
- 13) Regularization is a method used to prevent overfitting by adding additional information or constraints to the model. Overfitting is a state where the model may learn the noise in the training set almost perfectly but it does not have the ability to generate accurate results on new data that it has never seen before.  
Regularization, a statistical method, deals with specifics such as adding some penalty to the cost function (overfit) whenever the predictor is too complex. This constrains the model the energy function seeks to minimize making it simpler.
- 14) Lasso Regularization or L1 regularization , Ridge Regularization or L2 regularization
- 15) A residual variable that appears in statistical model when it does not completely represent the actual combination between independent and dependent variables.

## **PYTHON ASSIGNMENT:**

- 1) %
- 2) 0.666
- 3) 24
- 4) 2
- 5) 6
- 6) the finally block will be executed no matter if the try block raises an error or not.

- 7) It is used to raise an exception
- 8) In for loop
- 9) None of the above
- 10) Yield , raise

### **STATISTICS WORKSHEET-1**

- 1) True
- 2) Central Limit Theorem
- 3) Modeling bounded count data
- 4) D
- 5) Poisson
- 6) False
- 7) Hypothesis
- 8) 0
- 9) C
- 10) Normal Distribution is continuous probability distribution wherein values lie in a symmetrical fashion
- 11) Mean, Median and Mode if there are outliers then mean will not be taken in to count but outliers needs to be treated first
- 12) A/B testing is known as split testing a random experiment where 2 or more variable are shown in different segments
- 13) No its not correct practice as it does not preserve the relationship among variable
- 14) It is used to predict the value of a variable based on the value of another variable
- 15) descriptive statistics and inferential statistics.