# Worksheet set -1

# Machine learning

1.(A) Least square error

2.(A) Linear regression is sensitive to outliers

3.(B) Negative

4.(B) Correlation

5.(C)Low bias and high variance

6.(B)Predictive model

7.(D) Regularization

8. (D) SMOTE

9. (A) TPR and FPR

10. (B) False

11.(B) Apply PCA to project high dimensional data

12.(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 is a technique used to prevent overfitting by adding a penalty term to the loss function, discouraging the model from assigning too much importance to individual features or coefficients.

14. Ridge Regression:-Ridge regression is a method for analyzing data that suffer from multi-collinearity.

- LASSO (Least Absolute Shrinkage and Selection Operator) Regression:- LASSO is a regression analysis method that

performs both feature selection and regularization in order to enhance the prediction accuracy of the model.

Elastic-Net Regression:- Elastic-Net is a regularized regression method that linearly combines the L1 and L2 penalties of the LASSO and Ridge methods respectively.

15.it measure the distance of the observed y-values from the predicted y-values at each value of x; squaring each of these distances; calculating the mean of each of the squared distance.

# PYTHON

1.(C) %

2.(A) 0.666

3.(c)24

4.(A) 2

5.(D)6

6.(C) finally block will be executed no. matter if the try block raises an error

7(A) It is used to raise an exception

8. (C) in defining a generator

9(C) abc2

10. (D) all of the above

# Stastistics

1. (A) True
2. (A) central limit theorem
3. .(C) Modeling bounded count data
4. (D) All of the mentioned
5. (C) Poisson distribution
6. (B) False
7. (B) Hypothesis
8. (A) 0
9. . (C) outliers can’t conform to the regression

10.Normal Distribution: Normal Distribution is the most common or normal form of distribution of Random Variables, hence the name “normal distribution.” It is also called Gaussian Distribution in Statistics or Probability. We define Normal Distribution as the probability density function of any continuous random variable for any given system.

11. The most common methods of imputing values when dealing with missing data. In cases where there are a small number of missing observations, data scientists can calculate the mean or median of the existing observations and insert them in place of the missing observations open in new.

12. AB testing—also called split testing or bucket testing—compares the performance of two versions of content to see which one appeals more to visitors/viewers.

13. Mean imputation reduces the variance of the imputed variables. Mean imputation shrinks standard errors, which invalidates most hypothesis tests and the calculation of confidence interval. Hence mean imputation of missing data is not acceptable practice.

14. Linear regression analysis is used to predict the value of a variable based on the value of another variable.

15. There are three real branches of statistics: data collection, descriptive statistics and inferential statistics.