PYTHON – WORKSHEET 1

Answers

1. C)
2. B)
3. C)
4. A)
5. D)
6. C)
7. A)
8. C)
9. A), C)
10. A), B)

MACHINE LEARNING- WORK SHEET 1

1. B)
2. C)
3. B)
4. C)
5. D)
6. B)
7. C)
8. B)
9. A), B), C)
10. B), C)
11. C), D)
12. Mini Batch Gradient Descent (MBGD) can be used as the algorithm splits the training datasets into smaller pockets of data. These small batches of data are further used to calculate the necessary outcomes.

Likewise, Ensemble method such as bagging or boosting can be made use of which popularly utilizes the decision tree to outline an optimal predictive model.

13. Normal equations will not need any feature scaling techniques like min-max normalization or standardization.
    While any algorithms using gradient descent will be affected drastically causing skewness in the plot if feature scaling is not implemented.

STATISTICS- WORK SHEET 1

1. a)
2. a)
3. b)
4. d)
5. c)
6. b)
7. b)
8. a)
9. c)

10. Normal Distribution:
    - Normal distribution has a characteristic probability bell curve.
    - Mean, median and mode are equal
    - Normal distribution curves are symmetrical but not all symmetrical curves are normal.
    - Normal distribution is a limit case of Poisson distribution where n=infinite.
    - Normal distribution has skewness =0

11. Missing data in a dataset causes skewness hampering the quality of the dataset.
    Although, there aren't any proven techniques to handle missing data. The two primary methods to resolve the error would be to delete data with the missing values or use imputation techniques.
    My recommendation for an imputation technique would be the time series methods that makes assumption for the missing data based on the observation of the adjacent

data. This technique will not result in major losses in data variation like that in mean, median and mode method.

12. A/B testing or split testing is a performance evaluator that compares two versions of various assets like web pages, apps or emails against each other to evaluate their performances.

13. Mean Imputation of missing data should not be an ideal choice for imputation despite their popular use. This method draws no direct relationship between parameters and the data set loses its essence in acquiring variation in the data.

14. Linear Regression:
    - Linear regression is a type of predictive analysis.
    - A statistical model that analyses linear relationship between dependent variable and the independent variable.
    - Linear regression assumes the mean of the residuals is zero.
    - The error terms are assumed to be normally distributed.

15. Two branches of statistics are:
    - Descriptive statistics
    - Inferential statistics