MachineLearning_Assignment

1. A)
2. A)
3. B)
4. B)
5. A)
6. B)
7. D)
8. D)
9. A)
10. B)
11. B)
12. B) & C)
13. Regularization helps to reduce the effects of overfitting or underfitting issues. Overfitting occurs mostly due to presence of outliers and noise. The model performance is poor for the unseen data. It helps to nullify or reduce the effect of coefficients by applying the shrinkage factor (lamda).There are 2 types of regularization techniques (LASSO /L1 or Ridge/L2)
    In LASSO, it gives the zero weightage to the unnecessary features thereby eliminating them.I t also act as a feature selection method.
    In Ridge less weightage is given to the features which contributes less to the model.
14. LASSO and Ridge
15. In linear regression equation, the term error is defined as the Mean Squared error. It is the summation of square of a distance between actual & predicted value.

Python Worksheet 1

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

Statistics Worksheet 1

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

8. a)
9. c)
10. A normal distribution has a bell-shaped curve. When we collect the data viz height or weight of the population it is more likely to be normally distributed. It has two important parameters mean (mu) and standard deviation. Normal distributions are symmetrical. 68 % of data falls within 1 std dev of the mean followed by 95 % within 2 std dev and 99.7 % within 3 std dev.
11. Dealing with the missing data in datasets is important as it may have a significant impact on the model building. While dealing with large data, we may choose to delete the NAN values however it may not be an efficient option and we may choose other methods viz Mean, Median and Mode for replacing missing data.
12. A/B technique is used to determine which technique/versions will perform better or more precisely will make a significant contribution.
13. No. Since it doesn't consider the correlation factor.
14. Linear regression uses linear equation to model the given dataset. It has an equation y = mx+c where y(dependent variable) is the value to be predicted for given x (independent variable) , m is the slope and c is the intercept. It uses the least square method for selecting the best fitting line.
15. There are mainly two main branches viz Descriptive and inferential Statistics. The descriptive statistics is generally used where the population is small or finite for e.g avg weight of the class, Avg scores of Batch. Inferential statistics is performed when the data is large for e.g election opinion poll by news channel. In other words, a sample is drawn from the population to make the inference.