ASSIGNMENT-6 (MACHINE LEARNING)

1. C

2. B

3. C

4. A

5. B

6. A, D

7. C, B

8. A, C

9. A, B, C

10. Explain how does the adjusted R-squared penalize the presence of unnecessary predictors in the model?

Ans: The adjusted R-squared compensates for the addition of variables and only increases if the new predictor enhances the model above what would be obtained by probability. Conversely, it will decrease when a predictor improves the model less than what is predicted by chance.

11. Differentiate between Ridge and Lasso Regression.

Ans: Lasso known as (L1- Technique) and Ridge(L2-Technique) are the different types of regularization techniques used in Linear regression models where we penalize the coefficients to find the optimum solution.

In ridge, the penalty function is defined by the sum of the squares of the coefficients.

In Lasso, we penalize the sum of the absolute values of the coefficients. Another type of regularization method is Elastic-Net, it is a hybrid penalizing function of both lasso and ridge.

12. What is VIF? What is the suitable value of a VIF for a feature to be included in a regression modelling?

Ans: The Variance Inflation Factor (VIF) is a measure of collinearity among predictor variables within a multiple regression. It is calculated by taking an independent variable and regressing it against every other predictor in the model.

Including highly correlated variables in your model can lead to overfitting. If we overfit, then the model performs extraordinarily well on the training data but doesn't generalize well when we try to use it on new data. Small VIF values, VIF < 3, indicate low correlation among variables under ideal conditions. The default VIF cut-off value is 5; only variables with a VIF less than 5 will be included in the model. However, note that many sources say that a VIF of less than 10 is acceptable

13. Why do we need to scale the data before feeding it to the train the model?

Ans: Scaling has two process one is standardization and normalization. It helps to bring down the data under one scale. So that algorithms can take less time and can perform better.

14. What are the different metrics which are used to check the goodness of fit in linear regression?

Ans: Metrics that can be used for the linear Regression are:

I. Mean Absolute error
II. Root Mean Squared Error
III. Relative Absolute Error
IV. Relative Squared Error
V. Coefficient

15. From the following confusion matrix calculate sensitivity, specificity, precision, recall and accuracy.

Actual/Predicted	True	False
True	1000 True Positive (TP)	50 False Positive (FN)
False	250 False Negative (FN)	1200 True Negative (TN)

- **Accuracy=TruePostive+TrueNegative/TruePositive+TrueNegative+FalsePositive+FalseNegative**
  Solution: 1000+1200/1000+1200+250+50 =2200/2500 = 0.88
- **Recall = TruePositive/TruePostive+FalseNegative**
  Solution: 1000/1000+50 = 0.95
- **Precision = TruePositive/TruePositive+FalsePositive**
  Solution: 1000/1000+250 = 0.8
- **Sensitivity = TruePositive/TruePositive+FalseNegative**
  Solution: 1000/1000+50 = 0.95
- **Specificity = TrueNegative/FalsePositive+TrueNegative**
  Solution: 1200/50+1200 = 1200/1250 = 0.96