

MACHINE LEARNING – Assignment -6

1. (A) High R-squared value for train-set and High R-squared value for test-set.
2. (B) Decision trees are highly prone to overfitting.
3. (C) Random Forest
4. (A) Accuracy
5. (A) Model A
6. (A) Ridge (D) Lasso
7. (C) Random Forest
8. (A) Pruning
9. (B) A tree in the ensemble focuses more on the data points on which the previous tree was not performing well
(C) It is example of bagging technique
10. Whenever we add an independent variable to a model, the R-squared increases, even if the independent variable is insignificant. It never declines. Whereas Adjusted R-squared increases only when independent variable is significant and affects dependent variable.
11. Ridge and lasso regression are two common machine learning approaches for constraining model parameters. Both methods try to get the coefficient estimates as close to zero as possible because minimizing (or shrinking) coefficients can reduce variance dramatically (i.e., overfitting)
12. Variance inflation factor (VIF) is a measure of the amount of multicollinearity in a set of multiple regression variables.
If VIF of three or below is not a cause for concern, but if VIF exceeding 10 indicates high multicollinearity between this independent variable and the others.
13. Scaling of the data makes it easy for a model to learn and understand the problem.
14. Mean Squared Error (MSE). Root Mean Squared Error (RMSE). Mean Absolute Error (MAE)
15. Confusion Matrix ,
 1. sensitivity,

$$\begin{aligned} \text{SN} &= \text{TP} / (\text{TP} + \text{FN}) \\ \text{SN} &= 1000 / (1000 + 250) \\ &= 1000 / 1250 \\ &= 0.8 \\ &= 80\% \end{aligned}$$

2. Specificity

$$\text{specificity} = \text{TN} / (\text{TN} + \text{FP})$$

$$\text{specificity} = 1200 / (1200 + 50)$$

$$= 1200 / 1250$$

$$= 0.96$$

$$= 96\%$$