DATE
PAGE

Machine Learning -6

1. (A)
2. (B)
3. (c) Random forest
4. (B) Sensitivity
5. (B) Model B
6. (A) Ridge
   (D) Lasso.
7. (B) Decision Tree
   (C) Random forest
8. (A), (C)
9. (A), (B)
10. The adjusted R squared is a modified version of R-squared that has been adjusted for the number of predictors in the model.

- The adjusted R-squared increases only if the new term improves the model more than would be expected by chance
- It decreases when a predictor improves the model by less than expected by chance.

(11) The difference between ridge and lasso regression is that it tends to make coefficients to absolute zero as compared to Ridge which never sets the value of coefficient to absolute zero

(12) Variance inflation factor (VIF) is a measure of the amount of multicollinearity in a set of multiple regression variables.

- The VIF of a regression model is equal to the ratio of overall model variance to the variance of a model that includes only that single independent variable
- A high VIF indicates that the associated independent variable is highly collinear with other variables of model.

(13) We need to scale the data before feeding it to train the model because variables that are measured at different scales do not contribute equally to the analysis & might end up creating a bias.

(14) The different metrics used to check the goodness of fit in linear regression are! —

- R-Squared
- Overall f-test
- Root Mean Square Error (RMSE)

(15)

- Sensitivity = recall $= \frac{tp}{tp + fn}$

$= \frac{1000}{1000 + 50} = \frac{1000}{1050}$

$= 0.952$

- Specificity $= \frac{tn}{tn + fp}$

$= \frac{1200}{1200 + 250} = 0.827$

DATE
PAGE

- Precision $= \frac{tP}{tP + fP} = \frac{1000}{1000 + 250} = 0.8$

- Accuracy $= \frac{TP + TN}{TP + TN + FP + FN} = \frac{1000 + 1200}{1000 + 1200 + 250 + 50}$

$= \frac{2200}{2500}$

$= 0.88$