MACHINE LEARNING

ASSIGNMENT - 04

Answer No.01:- Option(c) -> High R-squared value for train-set and Low R-squared value for test-set.

Answer No.02:- Option(b) -> Decision trees are highly prone to overfitting

Answer No.03:- Option(c) -> Random Forest

Answer No.04:- Option(a) -> Accuracy

Answer No.05:- Option(b) -> Model B

Answer No.06:- Option(a&d) -> Ridge & Lasso

Answer No.07:- Option(b&c) -> Decision Tree & Random Forest

Answer No.08:- Option(a&c) -> Pruning & Restricting the max depth of the tree

Answer No.09:- Option(a&b) -> We initialize the probabilities of the distribution as 1/n, where n is the number of data-points & A tree in the ensemble focuses more on the data points on which the previous tree was not performing well.

Answer No.10:- 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.

Answer No.11:- Lasso is a modification of linear regression,where the model is penalized for the sum of absolute values of the weights.Thus the absolute values of weight will be reduced and many will tend to be zeros.

Ridge takes a step further and penalizes the model for the sum of squared value of the weights.Thus the weights not only tend to have smaller absolute values, but also really tend to penalize the extremes of the weights,resulting in a group of weights that are more evenly distributed.

Answer No.12:- Variance inflation factor (VIF) is a measure of the amount of multicollinearity in the regression analysis.As a rule of thumb,a VIF of three or below is not a cause for concern.As VIF increases the less reliable your regression results are going to be.

Answer No.13:- To ensure that the gradient descent moves smoothly towards the minima and that the steps for gradient descent are updated at the same rate for all the features , we scale the data before feeding it to the model.

Answer No.14:- Three statistics are used in ordinary least square regression to evaluate model fit –

R-squared

Overall F-test

Root Mean Square Error.

Well there are multiple types of goodness of fit tests but the most common is the chi square test.

Answer No.15:- Accuracy = 0.88

Sensitivity = 0.8

Specificity = 1.04

Precision = 0.95

Recall = 0.8