# MACHINE LEARNING

## Answer sheet

1. C) High R-squared value for train-set and Low R-squared value for test-set.
2. B) Decision trees are highly prone to overfitting.
3. C) Random Forest
4. A) Accuracy
5. B) Model B
6. A) Ridge & D) Lasso
7. B) Decision Tree & C) Random Forest
8. A) Pruning & C) Restricting the max depth of the tree
9. C) It is example of bagging technique & A) We initialize the probabilities of the distribution as 1/n, where n is the number of data-points
10. 
11. Ridge & Lasso regression both are doing work similar. while lasso regression takes the magnitude of the coefficients, ridge regression takes the square. Lasso is L1 Form and Bridge is L2 form.
12. VIF- A variance inflation factor (VIF) is a measure of the amount of multicollinearity in regression analysis. Multicollinearity exists when there is a correlation between two or more independent variables in a multiple regression model.

    Most research papers consider a VIF (Variance Inflation Factor) > 10 as an indicator of multicollinearity, but some choose a more conservative threshold of 5 or even 2.5.

13. scaling of the data makes it easy for a model to learn and understand the problem. Scaling of the data comes under the set of steps of data pre-processing when we are performing machine learning algorithms in the data set.
14. The adjusted R-square statistic is generally the best indicator of the fit quality when you add additional coefficients to your model. The adjusted R-square statistic can take on any value less than or equal to 1, with a value closer to 1 indicating a better fit. A RMSE value closer to 0 indicates a better fit.
15. 