MACHINE LEARNING

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. C) Precision

5. B) Model B

6. A) Ridge, D) Lasso

7. B) Decision Tree, C) Random Forest

8. D) All of the above

9. B) A tree in the ensemble focuses more on the data points on which the previous tree was not performing well, A) We initialize the probabilities of the distribution as 1/n, where n is the number of data-points.

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

* 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. Differentiate between Ridge and Lasso Regression.

**Ridge Regression:** It is used to solve multi collinearity in OLS regression models through the incorporation of shrinkage parameter (it is vital in ridge regression). The assumptions for the model is same as OLS model like linearity, constant variance and independence and normality not need to be assumed.

**Lasso Regression:** It is more similar to Ridge Regression but perform automatic variable selection. It allows regression coefficient to be zero whereas Ridge does not.

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

* The Variance Inflation Factor (VIF) measures the severity of multicollinearity in [regression analysis](https://corporatefinanceinstitute.com/resources/knowledge/finance/regression-analysis/). It is a statistical concept that indicates the increase in the variance of a regression coefficient as a result of collinearity.
* 1 = not correlated.
* Between 1 and 5 = moderately correlated.
* Greater than 5 = highly correlated.

. What is known is that the more your VIF increases, the less reliable your regression results are going to be. In general, a VIF above 10 indicates high correlation and is cause for concern. Some suggest a more conservative level of 2.5 or above.

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

* To ensure that the gradient descent moves very smoothly towards the minima and that the steps for gradient descent are updated at the same rate for all the features, we have to scale the data before feeding it to the model. Having features on a similar scale can help the gradient descent converge more quickly towards the minima.

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

* Mean Squared Error
* Root Mean Squared Error
* R-Squared
* Adjusted R-Squared.

15. From the following confusion matrix calculate sensitivity, specificity, precision, recall and accuracy. Actual/Predicted True False True 1000 50 False 250 1200.

* Sensitivity= TP/TP+FN=1000/1000+50=1000/1050=0.9523
* Specificity=TN/FP+TN=1200/250+1200=1200/1450=0.8275
* Precision=TP/TP+FP=1000/1000+250=1000/1250=0.8
* Recall=TP/TP+FN=1000/1000+50=1000/1050=0.9523
* Accuracy= F- measure= 2\*[(precision\*recall/precision+recall)]= 2\*[0.8\*0.9523]/[0.8+0.9523]=2\*[0.7618/1.7523]=2\*0.4347=0.8694.