WORKSHEET-4

MACHINE_LEARNING_ANSWERSHEET_4

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 over lifting.**
3. **(c) Random Forest**
4. **(b) Sensitivity**
5. **(b) Model B**
6. **(a) & (d) Ridge & Lasso**
7. **(b) & (c) Decision Tree & Random Forest**
8. **(a) & (c) Pruning & Restricting the max depth of the tree**
9. **(a) We initialize the probabilities of the distribution as 1/n, where n is the number of data points.**
10. **R-squared comes with an inherent problem – additional input variables will make the R-squared stay the same or increase (this is due to how the R-squared is calculated mathematically). Therefore, even if the additional input variables show no relationship with the output variables, the R-squared will increase. if R-squared remains constant and number of predictors increase then adjusted R-Squared decreases. So, if R-squared is not increased significantly on adding predictors, adjusted R**

Squared will decrease. So, in this way adjusted r-squared will penalize the presence of unnecessary predictors.

11. Ridge and Lasso regression are some of the simple techniques to reduce model complexity and prevent over-fitting which may result from simple linear regression. However, There are few difference between Ridge and Lasso which is as follows.

- **Lasso regression takes the magnitude of the coefficients whereas Ridge regression takes the square.**
- **Lasso is a modification of linear regression, in which the absolute values of weight will be reduced, and many will tend to be zeros whereas in Ridge group of weights that are more evenly distributed.**

12. Variance inflation factor (VIF) is a measure of the amount of multicollinearity in regression analysis. Multicollinearity exists when there is a correlation between multiple independent variables in a multiple regression model.

The default VIF cutoff value is 5; only variables with a VIF less than 5 will be included in the model. However, note that many sources say that a VIF of less than 10 is acceptable.

13. Below are the reasons to scale the data before feeding it to train the model :-

- **It is 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.**
- **Feature scaling is essential for machine learning algorithms that calculate distances between data.**

14. R-squared, the overall F-test, and the Root Mean Square Error (RMSE) are used to check the goodness of fit in linear regression. 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.

Actual / Predicted	TRUE	FALSE
TRUE	1000	50
FALSE	250	1200

F_P =250

F_n = 50

T_p = 1000

T_n = 1200

- Accuracy (all correct / all) = **T_p** + **T_n** / **T_p** + **T_n** + **F_P** + **F_n**
  (1000+1200)/(1000+1200+250+50) = 0.88
- Specificity (true negatives / all actual negatives) = **T_n** / **T_n** + **F_P**
  1200/ (1200+250) =0.8275
- Precision (true positives / predicted positives) = **T_p** / **T_p** + **F_P**
  1000/(1000+250) = 0.8
- Recall (true positives / all actual positives) = **T_p** / **T_p** + **F_n**
  1000/(1000+50) = 0.9523