Machine Learning Worksheet-6

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. B) Sensitivity

5. B) Model B

6. A) Ridge, D) Lasso

7. C) Random Forest, B) Decision Tree

8. A) Pruning, C) 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, B) A tree in the ensemble focuses more on the data points on which the previous tree was not

performing well

10. Mathematically **adjusted R-squared** is given as:

$$1-\frac{(\mathbf{1}-\boldsymbol{R}^{2})(\boldsymbol{N}-\mathbf{1})}{(\boldsymbol{N}-\boldsymbol{M}-\mathbf{1})}$$

Here,

$\boldsymbol{R^2}$ = R-squared value determined from the model

N = Number of independent variables

M = Number of data points

From the above expression for R-squared, we see that, for a fixed number of data points (***M***), if the model's performance (in other words, R-squared) does not increase significantly by increasing the number of features, then the overall adjusted R-squared value may decrease, penalizing the effect of unnecessary predictors present in the model.

11.

Sl no.	**Lasso Regression**	**Ridge Regression**
1	It is known as L1 regularization, as the objective cost function contains sum of absolute weights as a penalty term.	It is known as L2 regularization, as the objective cost function contains sum of squares of weights as a penalty term.
2	Lasso regression tend to make coefficient/weights to absolute zero.	In contrast to Lasso regression, Ridge regression only reduces the coefficients to smaller values.
3	Lasso regression can also be used as a feature selection/reduction process for large features in the data as it reduces feature coefficient to zero for penalizing higher cost.	Ridge regression is useful if only relevant features are present in the dataset, which we may not want to

		remove from our model.

12. VIF (Variance Inflation Factor) is used to determine the strength of relationship between the independent variables and therefore the overall multicollinearity between these variables. VIF calculations for a single variable involves taking that variable as a target and regressing it against every other independent variable in the dataset. Mathematically, VIF is given as follows:

VIF = $\frac{1}{1-R^2}$

Here, R^2 is the R-squared value for a particular independent variable as a target.

Also, from the above equation, VIF score increases with increase in R-squared value, implying higher **multicollinearity** for that particular independent variable.

Setting a VIF threshold for a particular variable is subjective, and rather depends on the relevance of that variable for solving that problem or domain expertise, however, in practice generally VIF greater than 10 is considered highly correlated and is likely to get removed.

13. We need to scale data mainly due to following reasons:

- It increases the computational speed of the algorithm.
- It may increase convergence speed of the algorithm.

14. Metrics which are used to check goodness of fit in linear regression are:

Root Mean Squared Error (RMSE): RMSE is a very common evaluation metric. It can range between 0 and infinity. Lower values are better.

Mathematically it is denoted by:

square root of (1/n * (∑ (y -ŷ) ²))

Here,

n = Number of observations

y = Actual value

ŷ = Predicted value

Mean Absolute Error (MAE): Mean Absolute Error (MAE) is simply the average of the absolute value of the errors.

Mathematically it is denoted by:

(1 / n) * (∑ |y - ŷ|)

R-Squared (R^2): R^2 represents the proportion of variance explained by the model.

Mathematically it is denoted by:

1 - (SSE/SST)

Here,

SSE = The sum of the squared differences between the

actual values and predicted values.

SST = The sum of the squared differences between the actual values and the

mean of the all the observations of predicted values.

15. Recall = Sensitivity = $\frac{\mathbf{1000}}{\mathbf{1000+250}}$ **= 0.80**

Specificity = $\frac{\mathbf{1200}}{\mathbf{1200+50}}$ **= 0.96**

Precision = $\frac{\mathbf{1000}}{\mathbf{1000+50}}$ **= 0.95**

Accuracy = $\frac{\mathbf{1000+1200}}{\mathbf{(1000+1200)+(250+50)}}$ **= 0.88**