MACHINE LEARNING – WORKSHEET 2

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. (A) ADABOOST
   (D) XGBOOST
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. In linear regression model R-squared on training data increases as we increase the number of predictors because as we increase the number of predictors we are adding more information to the model so the r-squared increases on the training dataset. But this can lead to overfitting of the model. To prevent the overfitting in spite of r-squared we use adjusted r-squared. Its formula is given as follows:

$$R^2_{adjusted} = 1 - \frac{(1 - R^2)(N - 1)}{N - p - 1}$$

where

R^2 = sample R-square

p = Number of predictors

N = Total sample size.

As we can see from the above formula 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 penalise the presence of unnecessary predictors.

11. **Ridge Regression:**
    It is technique of regularisation in linear regression. It regularises the model by using L2 regularisation which tries to minimise the sum of squares of the magnitude of the coefficients along with the error.
    The cost function in ridge regression looks like this:
    Cost function= MSE+ α * (sum of square of coefficients)
    Where, MSE is the mean squared error.
    α is the regularisation constant. The regularisation increases with increment in α.

    Lasso Regression:
    It is also technique of regularisation in linear regression. It regularises the model by using L1 regularisation which tries to minimise the sum of the magnitude of the coefficients along with the error.
    The cost function in lasso regression looks like this:
    Cost function= MSE+ α * (sum of magnitude of coefficients)
    Where, MSE is the mean squared error.
    α is the regularisation constant. The regularisation increases with increment in α.

12. **VIF :**
    VIF stands for Variance Inflation Factor. VIF determines the strength of the correlation between the independent features. It is predicted by taking a variable and regressing it against every other feature in the dataset. R^2 value is determined to find out how well an independent feature is described by the other independent features. A high value of R^2 means that the feature is highly correlated with the other features. This is captured by the VIF which is denoted below:

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

Where R_i^2 = R^2 score of ith feature against all other features
So, the closer the R^2 value to 1, the higher the value of VIF and the higher the multicollinearity with the particular independent variable.
Generally, if VIF is less than 4, the feature is acceptable to be a part of model otherwise it is dropped.

13. The following are the reasons to scale the data before feeding it to train the model:
    - The gradient descent algorithm which is used to reach the optimal solution in most of the cases, it reached the optimal solution much faster if all the features are at the same scale. That's why scaling helps to reach the optimal solution.
    - If the features in the training dataset are on different scales, then during training the features with large scales will be favoured over there in order to minimise the loss. That's why we do Scaling to puts all the features on the same scale.

14. Some of the metrics used to check the goodness of fit in linear regression are:
    - MSE- mean squared error. As the name suggests it is the average value of squares of the errors made by model on a dataset.

      MSE=(sum(Ytrue-Yexpected)^2)/n
    - R-squared. It is defined as the variance explained by the model/Total variance of the dataset.

      $$R^2 = \frac{\text{Variance explained by the model}}{\text{Total variance}}$$
    - Adjusted R-squared: It takes in to account both the R-squared as well as the number of predictors in the model. That is it considers both the variance explained by the model as well as the number of predictrors used by the model to explain that variance.

      $$R^2_{adjusted} = 1 - \frac{(1 - R^2)(N - 1)}{N - p - 1}$$

      where
      R^2 = sample R-square
      p = Number of predictors
      N = Total sample size.

15.

Actual/Predicted	True	False
True	1000	50
False	250	1200

Fp=False positives=250
Fn= False negatives=50
Tp=True positives=1000
Tn=True negatives=1200
sensitivity= Tp/(Tp+Fn) = 1000/1050 = 0.9523
specificity=Tn/(Tn+Fp) = 1200/1450 = 0.8275
precision=Tp/(Tp+Fp) = 1000/(1000+250) = 0.8
recall=Tp/(Tp+Fn) = 1000/(1000+50) = 0.9523
precision=(Tp+Tn)/(Tp+Tn+Fp+Fn) = (1000+1200)/(1000+1200+250+50) = 0.88