MACHINE LEARNING_Worksheet_4 -ANSWERS

1) **A) High R-squared value for train-set and High 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 B)Lasso**
7) **B) Decision Tree C) Random Forest**
8) **A) Pruning C) Restricting the max depth of the tree**
9) **B) A tree in the ensemble focuses more on the data points on which the previous tree was not performing well C) It is example of bagging technique**

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

Answer= The adjusted R-squared compensates for the addition of variables and only increases if the new predictor enhances the model above what would be obtained by probability. Conversely, it will decrease when a predictor improves the model less than what is predicted by chance. Adjusted R2 is a corrected goodness-of-fit (model accuracy) measure for linear models. It identifies the percentage of variance in the target field that is explained by the input or inputs. R2 tends to optimistically estimate the fit of the linear regression

11) **Differentiate between Ridge and Lasso Regression ?**

Answer= lasso regression, ridge regression puts a similar constraint on the coefficients by introducing a penalty factor. However, while lasso regression takes the magnitude of the coefficients, ridge regression takes the square. Ridge regression is also referred to as L2 Regularization.Ridge and lasso regression are two common machine learning approaches for constraining model parameters. Both methods try to get the coefficient estimates as close to zero as possible because

minimizing (or shrinking) coefficients can reduce variance dramatically (i.e., overfitting). Lasso will eliminate many features, and reduce overfitting in your linear model. Ridge will reduce the impact of features that are not important in predicting your y values. Elastic Net combines feature elimination from Lasso and feature coefficient reduction from the Ridge model to improve your model's predictions.

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

Answer= A 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. This can adversely affect the regression results. VIF starts at 1 and has no upper limit. VIF = 1, no correlation between the independent variable and the other variables. VIF exceeding 5 or 10 indicates high multicollinearity between this independent variable and the others. The variance inflation factor (VIF) quantifies the extent of correlation between one predictor and the other predictors in a model. It is used for diagnosing collinearity/multicollinearity. Higher values signify that it is difficult to impossible to assess accurately the contribution of predictors to a model.

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

Answer= Scaling. This means that you're transforming your data so that it fits within a specific scale, like 0-100 or 0-1. You want to scale data when you're using methods based on measures of how far apart data points, like support vector machines, or SVM or k-nearest neighbors, or KNN. 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, we scale the data before feeding it to the model**.** Scaling the target value is a good idea in regression modelling; 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. What are the different metrics which are used to check the goodness of fit in linear regression?

Answer = There are 3 main metrics for model evaluation in regression:

1. R Square/Adjusted R Square
2. Mean Square Error(MSE)/Root Mean Square Error(RMSE)
3. Mean Absolute Error(MAE)

a) R Square/Adjusted R Square R Square measures how much variability in dependent variable can be explained by the model. It is the square of the Correlation Coefficient(R) and that is why it is called R Square. R square formula R Square is calculated by the sum of squared of prediction error divided by the total sum of the square which replaces the calculated prediction with mean. R Square value is between 0 to 1 and a bigger value indicates a better fit between prediction and actual value. R Square is a good measure to determine how well the model fits the dependent variables. However, it does not take into consideration of overfitting problem.

b) Mean Square Error(MSE)/Root Mean Square Error(RMSE) While R Square is a relative measure of how well the model fits dependent variables, Mean Square Error is an absolute measure of the goodness for the fit. Mean Square Error formula MSE is calculated by the sum of square of prediction error which is real output minus predicted output and then divide by the number of data points. It gives you an absolute number on how much your predicted results deviate from the actual number. You cannot interpret many insights from one single result but it gives you a real number to compare against other model results and help you select the best regression model.

c) Mean Absolute Error(MAE) is similar to Mean Square Error(MSE). However, instead of the sum of square of error in MSE, MAE is taking the sum of the absolute value of error. Mean Absolute Error formula Compare to MSE or RMSE, MAE is a more direct representation of sum of error terms. MSE gives larger penalization to big prediction error by square it while MAE treats all errors the same.

15. From the following confusion matrix calculate sensitivity, specificity, precision, recall and accuracy?

Actual/predicted	**True**	**False**
True	**1000**	**50**
False	**250**	**1200**

Answer=

a) Sensitivity = TP/TP+FN = 0.8
b) Specificity = TN/TN+FP = 0.8
c) Precision = TP/TP+FP = 0.952
d) Recall = TP/TP+FN = 0.8
e) Accuracy = TP+TN/TP+TN+FP+FN = 0.88