MACHINE LEARNING

Q.1] C

Q.2] B

Q.3] C

Q.4] B

Q.5] B

Q.6] A & B

Q.7] B & C

Q.8] D

Q.9] A & B

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

ANS. = Adjusted R-squared is a modified version of R-squared that adjusts for the number of predictors in a model. The traditional R-squared value simply compares the variance of the predicted values to the total variance of the response variable. However, as more predictors are added to a model, the R-squared value will always increase, even if the new predictors do not improve the model's ability to predict the response.

The adjusted R-squared is calculated by taking the R-squared value and subtracting the proportion of variance that is attributed to each additional predictor. The adjustment is based on the number of predictors in the model and the sample size. As a result, as the number of predictors increases, the adjusted R-squared value will decrease if the additional predictors do not improve the model's ability to predict the response.

In this way, the adjusted R-squared penalizes the presence of unnecessary predictors in the model by decreasing the R-squared value if the additional predictors do not improve the model's ability to predict the response. This makes it a better measure of goodness of fit of a model as it accounts for the number of predictors, which R-squared does not. It's a way to adjust for the possible overfitting problem caused by adding multiple predictors in the model.

Q.11] Differentiate between Ridge and Lasso Regression ?

ANS. = Ridge and Lasso Regression are both regularization techniques used to prevent overfitting in Linear Regression. Both techniques add a penalty term to the cost function to shrink the coefficients of the model towards zero. However, the way the penalty term is added to the cost function is different for Ridge and Lasso Regression.

Ridge Regression, also known as L2 regularization, adds a penalty term equal to the square of the magnitude of the coefficients. The regularization term is the sum of the squares of the coefficients multiplied by a hyperparameter alpha. This has the effect of shrinking the coefficients, but it also keeps all the coefficients non-zero.

Lasso Regression, also known as L1 regularization, adds a penalty term equal to the absolute value of the magnitude of the coefficients. The regularization term is the sum of the absolute values of the coefficients multiplied by a hyperparameter alpha. This has the effect of shrinking the coefficients and it also makes some coefficients zero, which is known as feature selection.

In Ridge Regression, all the coefficients are shrunk by the same factor, whereas in Lasso Regression, the coefficients that are less important are shrunk more than the important coefficients. Lasso Regression is particularly useful when the number of predictors is high and some of them are not important.

In summary, Ridge Regression is used to prevent overfitting by shrinking the coefficients of the model towards zero, while Lasso Regression is used to select the most important predictors by shrinking the less important predictors to zero.

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

ANS. = VIF, or Variance Inflation Factor, is a statistical measure used to assess the level of multicollinearity in a multiple regression model. Multicollinearity occurs when two or more predictor variables in a model are highly correlated with each other. This can cause problems in interpreting the individual coefficients of the predictor variables, as well as in estimating the model's standard errors and confidence intervals.

VIF is calculated for each predictor variable in a multiple regression model, and is defined as the ratio of the variance of the predicted value of the variable to the variance of the variable when it is used alone. A VIF of 1 indicates that there is no multicollinearity between the predictor variable and the other predictor variables in the model. A VIF greater than 1 indicates that there is multicollinearity, with the degree of multicollinearity increasing as the VIF increases.

The general rule of thumb is that a VIF of 1 is suitable for a feature to be included in a regression modeling, and it is considered high when it's greater than 5 and it should be removed. But some researcher consider VIF of 10 as high. However, it also depends on the context of the problem, and in some cases, a VIF of greater than 1 may be acceptable. In order to avoid VIF, the best practice is to use principal component analysis (PCA) or independent component analysis (ICA) to decorrelate the features.

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

ANS. = Scaling the data is an important step in preprocessing the data before training a model for several reasons:

1. Many machine learning algorithms are sensitive to the scale of the input features. For example, algorithms based on gradient descent, such as linear regression and neural networks, can converge much faster when the features are on a similar scale.

2. Scaling the data can also help improve the performance of some algorithms, such as k-nearest neighbors and k-means clustering, by reducing the impact of outliers and ensuring that all features are given equal weight.

3. Scaling the data can also help to prevent overfitting, by reducing the magnitude of the larger-scale features and giving a more equal weight to all features in the model.

4. Some models like SVM, Neural Networks and KNN are sensitive to the magnitude of the features, so it is important to bring all features to the same scale to get better results.

5. Standardization is a common preprocessing method to scale the data. It is a technique of transforming the data into a standard normal distribution (mean = 0, standard deviation = 1)

6. Avoiding dominance of one variable over other variable: Scaling the data can also help to avoid the dominance of one variable over the other variable, which can be a problem in some machine learning models.

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

ANS. = The different metrics used to check the goodness of fit in linear regression include:

1. R-squared: It measures the proportion of variation in the response variable that is explained by the predictor variables. It ranges between 0 and 1. The higher the value, the better the model fit.

2. Adjusted R-squared: It is similar to R-squared but it penalizes the presence of unnecessary predictors in the model.

3. Mean Squared Error (MSE): It measures the average of the square of the errors (prediction errors). It is used to estimate the variance of the errors.

4. Root Mean Squared Error (RMSE): It is the square root of the mean squared error. It is used to estimate the standard deviation of the errors.

5. Mean Absolute Error (MAE): It measures the average of the absolute errors. It is used to estimate the errors' magnitude.

6. Median Absolute Error (MedAE): It measures the median of the absolute errors. It is used to estimate the errors' magnitude.

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

ACTUAL/PREDICTED	TRUE	FALSE
TRUE	1000	50
FALSE	250	1200

ANS. = 1. SENSITIVITY = 1000/(1000+250)= 0.8

2. SPECIFICITY= 1200/(1200+50)= 0.96

3. PRECISION= 1000/(1000+50)= 0.95

4. RECALL= 1000/(1000+250) = 0.8

5. ACCURACY= (1000+1200)/(1000+50+250+1200)= 0.86