<u>MACHINE LEARNING-6</u>

1-

2-B

3-C

4-C

5-C

6-A, D

7-A, D

8-D

9-D

10-Regression analysis is a predictive modelling technique that assesses the relationship between a dependent (i.e., the goal/target variable) and independent factors. Using regression analysis, forecasting, time series modelling, determining the relationship between variables, and predicting continuous values can all be done.

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)

Ridge Regression:

Ridge regression is a technique used to analyze multi-linear regression (multicollinear), also known as L2 regularization. It is Applied when predicted values are greater than the observed values.

$$\sum_{i=1}^{n}\left(y_i-\beta_0-\sum_{j=1}^{p}\beta_j x_{ij}\right)^2+\lambda\sum_{j=1}^{p}\beta_j^2=\mathrm{RSS}+\lambda\sum_{j=1}^{p}\beta_j^2$$

Above equation represents the formula for Ridge Regression! where, Lambda (λ) in the equation is tuning parameter which is selected using cross-validation technique which makes the fit small by making squares small (β2) by adding shrinkage factor.

The shrinkage factor is lambda times the sum of squares of regression coefficients (The last element in the above equation).

Lasso Regression:

Lasso stands for – Least Absolute Shrinkage and Selection Operator. It is a technique where data points are shrunk towards a central point, like the mean. Lasso is also known as L1 regularization.

It is applied when the model is overfitted or facing computational challenges.

$$\sum_{i=1}^{n}\left(y_i-\beta_0-\sum_{j=1}^{p}\beta_j x_{ij}\right)^2+\lambda\sum_{j=1}^{p}|\beta_j| \;=\mathrm{RSS}+\lambda\sum_{j=1}^{p}|\beta_j|.$$

The above equation represents the formula for Lasso Regression! where, Lambda (λ) is a tuning parameter selected using the before Cross-validation technique.

Unlike Ridge Regression, Lasso uses |β| to penalize the high coefficients.

The shrinkage factor is lambda times the sum of Regression coefficients (The last factor in the above equation).

13-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. Thus, the variance inflation factor can estimate how much the variance of a regression coefficient is inflated due to multicollinearity.

The formula for VIF is:

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

Where R_i^2 represents the unadjusted coefficient of determination for regressing the i^{th} independent variable on the remaining ones.

When R_i^2 is equal to 0, and therefore, when VIF or tolerance is equal to 1, the i^{th} independent variable is not correlated to the remaining ones, meaning that multicollinearity does not exist.1

In general terms,

- VIF equal to 1 = variables are not correlated
- VIF between 1 and 5 = variables are moderately correlated
- VIF greater than 5 = variables are highly correlated2

The higher the VIF, the higher the possibility that multicollinearity exists, and further research is required. When VIF is higher than 10, there is significant multicollinearity that needs to be corrected

13-Feature scaling is a method used to normalize the range of independent variables or features of data. In data processing, it is also known as data normalization and is generally performed during the data preprocessing step

There are two reasons which supports the need for scaling.

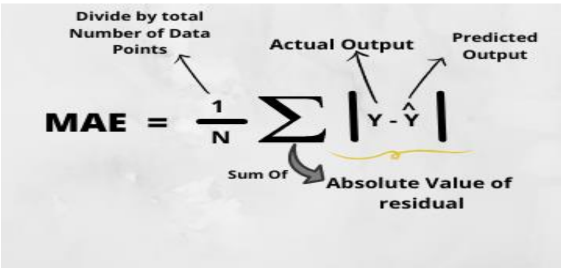
- Scaling the features in a machine learning model can improve the optimization process by making the flow of gradient descent smoother and helping algorithms reach the minimum of the cost function more quickly.
- Without scaling features, the algorithm may be biased toward the feature with values higher in magnitude. Hence we scale features that bring every feature in the same range, and the model uses every feature wisely.

14-Regression is a type of Machine learning which helps in finding the relationship between independent and dependent variable

Different type of metrices of linear regression are as follows.

1) Mean Absolute Error(MAE)

MAE is a very simple metric which calculates the absolute difference between actual and predicted values.



2) Mean Squared Error(MSE)

- MSE is a most used and very simple metric with a little bit of change in mean absolute error. Mean squared error states that finding the squared difference between actual and predicted value.

$$MSE = \frac{1}{n} \Sigma \underbrace{\left(y - \widehat{y} \right)^2}_{\text{The square of the difference between actual and predicted}}$$

It represents the squared distance between actual and predicted values. we perform squared to avoid the cancellation of negative terms and it is the benefit of MSE.

3) Root Mean Squared Error(RMSE)

As RMSE is clear by the name itself, that it is a simple square root of mean squared error.

$$\text{RMSE} = \sqrt{\text{MSE}}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)}$$

4) Root Mean Squared Log Error(RMSLE)

Taking the log of the RMSE metric slows down the scale of error. The metric is very helpful when you are developing a model without calling the inputs. In that case, the output will vary on a large scale.

5) R Squared (R2)

R2 score is a metric that tells the performance of your model, not the loss in an absolute sense that how many wells did your model perform.

So, with help of R squared we have a baseline model to compare a model which none of the other metrics provides. The same we have in classification problems which we call a threshold which is fixed at 0.5. So basically R2 squared calculates how must regression line is better than a mean line.

Hence, R2 squared is also known as Coefficient of Determination or sometimes also known as Goodness of fit.

$$\text{R2 Squared} = 1 - \frac{SSr}{SSm}$$

SSr = Squared sum error of regression line

SSm = Squared sum error of mean line

6) Adjusted R Squared

The disadvantage of the R2 score is while adding new features in data the R2 score starts increasing or remains constant but it never decreases because It assumes that while adding more data variance of data increases.

But the problem is when we add an irrelevant feature in the dataset then at that time R2 sometimes starts increasing which is incorrect.

Hence, To control this situation Adjusted R Squared came into existence.

$$R_a^2 = 1 - \left[\left(\frac{n-1}{n-k-1}\right) \times (1 - R^2)\right]$$

where:
n = number of observations
k = number of independent variables
R_a^2 = adjusted R^2

Now as K increases by adding some features so the denominator will decrease, n-1 will remain constant. R2 score will remain constant or will increase slightly so the complete answer will increase and when we subtract this from one then the resultant score will decrease. so this is the case when we add an irrelevant feature in the dataset.

And if we add a relevant feature then the R2 score will increase and 1-R2 will decrease heavily and the denominator will also decrease so the complete term decreases, and on subtracting from one the score increases.

15-The precision for this model is calculated as:

$$Precision = \frac{TP}{TP + FP}$$

$$Precision = \frac{1000}{1000 + 50}$$

$\boldsymbol{Precision}$ =0.95

Accuracy is calculated for this models-

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

$$Accuracy = \frac{1000 + 1200}{1000 + 50 + 1200 + 250}$$

$\boldsymbol{Accuracy}$ =88%

Recall is calculated for this models-

$$Recall = \frac{TP}{TP + FN}$$

$$Recall = \frac{1000}{1000 + 250}$$

$\boldsymbol{Recall}$ =0.8 %