MACHINE LEARNING ASSIGNMENT-6

ANS. NO.1 (B)

ANS.NO.2 (B)

ANS.NO.3 (C)

ANS.NO.4 (C)

ANS.NO.5 (C)

ANS.NO.6 (A) (D)

ANS.NO.7 (B) (C)

ANS.NO.8 (A)(C)

ANS.NO.9 (B)(C)

ANS.NO.10

R-Squared (R-Squared is a statistical measure of fit that indicates how much variation of a dependent variable is explained by the independent variable(s) in a regression model.)

The adjusted R-squared penalize the presence of unnecessary predictors in the model?

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.

ANS.NO.11

Ridge Regression: Ridge regression is a technique used to analyze multi-linear regression also known as L2 regularization. It is applied when predicted values are greater than the observed values.

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 over fitted or facing computational challenges.

ANS.NO.12

Variance Inflation Factor (VIF): Variance Inflation Factor (VIF) is used to detect the presence of multicollinearity. Variance inflation factors (VIF) measure how much the variance of the estimated regression coefficients are inflated as compared to when the predictor variables are not linearly related.

Small VIF values, VIF < 3, indicate low correlation among variables under ideal conditions. The default VIF cut off value is 5; only variables with a VIF less than 5 will be included in the model. However, note that many sources say that a VIF of less than 10 is acceptable.

ANS.NO.13

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.

ANS.NO.14

There are multiple types of goodness-of-fit tests, but the most common is the chi-square test.

Chi-square (The Chi-squared test can be used to measure the goodness-of-fit of your trained regression model on the training, validation, or test data sets.)

ANS.NO.15

1. **Sensitivity : TP/TP+FN (1000/(1000+250) = 0.80**
2. **Specificity : TN/TN+FP (1200/(1200+50) = 0.96**
3. **Precision : TP/TP+FP (1000/(1000+50) = 0.95**
4. **Accuracy : TP+TN/TP+TN+FP+FN (1000+1200)/(1000+1200+50+250)=0.88**
5. **Recall : TP/TP+FN (1000/(1000+250) = 0.80**