Regression assignment

1 difference between training and test data set

Training data set is the one we use to train the model and test data set is used to test the model and to predict the output and check with the actual output. Test data should always be aprroxipately 30 percent of the whole data set.

2 evaluation metrics for regression model are:

a.sum of squared error

b. sum of square total

c.R2 score

d. sum of square regression

3. categorical data is unsuitable for regression.

4. difference between mse and mae

Mse: mean squared error is the average or mean of the squared difference btw predicted output and actual output.it is highly sensitive and wen theres an outlier it analyses the model biased toward the outlier, eg dataset=[2,4,3,500] 500 is the outlier which will skew the model performance.

MAE: is the mean absolute error, it is the average of the absolute difference between the predicted output and actual output. It will treat all the errors same because it doesn't amplify them by squaring them like MSE.

5.how to interpret coefficients of regression model: with the help of slope and bias.

6. residual in a regression : it is the difference btw the actual output and predicted output.

7.why is cross validation important in regression: cross validation creates different iternaries with multiple folds of datasets. So the model will get trained on each iternary and prevent overfitting and bias by jus learning a fixed split. So cross validation improves the performance of the model.

8, R2 score of .85 is below 1 so its an average performoing model.

9. to prevent slowing of the model to converge we can use standardisation in SVM algorithm or pruning in random forest to prevent this.

10. if model is under perforforming try:

Data preprocessing

Try using various hyperparameters to find the best model

Try cross validation

We should try all the algorithms to choose the best model.