ML ASSIGNMENT – LAB02

QUESTION 2:

1. What are the assumptions of linear regression?

* There should be a linear relationship.
* Multivariate normality.
* There should not be auto correlation.

1. How can we evaluate a Regression model? Define each metric and its interpretation.

Evaluation metrics are a measure of how good a model performs and how well it approximates the relationship.

There are 3 error metrics that are commonly used. They are:

* Mean squared error (MSE):

While R square is a relative measure of how well the model fits dependent variables, MSE is an absolute measure of the goodness for the fit. It 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.

* 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.

* Mean absolute error (MAE) :

It 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. Compare to MSE or RMSE, MAE is a more direct representation of sum of error terms.

1. Can R squared be negative?

Yes, if it do not contain a constant term and if regression is worse fit.

1. What is dummy variable trap?

The dummy variable trap occurs when two or more dummy variables created by one-hot encoding are highly correlated. One variable predicts the value of others.

1. Is one hot encoding different from dummy variables?

No difference actually. One hot encoding is the thing you do to create dummy variables.

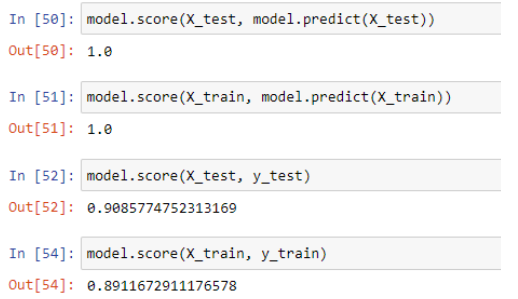
A dummy variable just takes the value 0 or 1 to indicate the exclusion or inclusion of a category.

In one-hot encoding, for example “red” color is encoded as [1 0 0] vector of size 3. “green” color is encoded as [0 1 0] vector of size 3.

1. How is polynomial regression different from linear regression?

Polynomial regression is a form of linear regression where only due to the non-linear relationship between dependent and independent variables we add some polynomial terms to linear regression to convert it into polynomial regression. Polynomial basically fits a wide range of curvature.





Model.score() is used to get the accuracy of a model whereas model.predict is used to get predict variable.

In the first two lines of code, we can see that the r squares value is 1 which is the best fit. The first, code shows the variation of x\_test is explained by the predicted value of x\_test. the In second code, the variation of x\_train is explained by the predicted x\_train variable.

In the third and fourth code, the accuracy of x\_test and y\_test is 0.90 i.e 90% and the accuracy of X\_train and y\_train variables are 0.89 i.e 89%