

Reflection

As a summary understanding for the Linear model Homework, the data advertising.csv is use for the prediction of sale, which will become y, independent variable, in the code.

After the data X and Y are defined, the data is spitted using train test split with 30% on the testing data with random state 42. The StandardScaler is use to standardize the training features, which are TV, radio and newspaper. Linear Regression Model is trained on a scale training data and the coefficient and intercept is printed. The test data is also transformed into standardize data for the prediction of Y(sale data) and do the prediction for the ytest. The performance is evaluated on MAE (mean absolute error), MSE (mean square error) and R square for both training and testing dataset.

As an additional learning, another learning is that as the features are scaled, the predicted result will be also in scaled units and converting the model back to unscaling is needed to have real data.

Challenges : While looking more about scaling and unscaling, the coefficient, a lot of math related information are surfaced and it take time to understand the concept and the mathematic. Although I am not fully understanding about the whole concept, I have much clearer understanding about the linear regression and scaling and model interpretation works.