Multiple Linear Regression

1. Download the “BodyFat” dataset from the assignment page.

Fit a regression model that explains the “PercentBodyFat” as response variable. Use all the rest of variables that are in the data set as explanatory variables(except for the “BodyDensity”). What is the R^2 of this regression?

Optimize the model based on proper model selection procedures explained in the class step by step. Justify your optimization of the problem.

Check for any violations of the regression assumptions (homoscedasticity, Normality, linearity) by using appropriate plots. Identify possible outliers, if exists.

Use a 70-30 splits to validate your model with proper metrics.

All the steps need to be submitted with explanations and justifications.

2. Download the “auto _mpg” data set from the assignments page. In the “Origin” column, the numbers signify: 1: USA, 2: Europe and 3: Asia.

Fit a regression model that explains the “mpg” in common properties of a car. Use all the variables that are in the data set (except for the “car name”). What is the R^2 of this regression? Be aware of the categorical variables(s).

Optimize the model based on proper model selection procedures explained in the class step by step. Justify your optimization of the problem.

Check for any violations of the regression assumptions by using appropriate plots. Identify possible outliers, if exists.

Use a 70-30 splits to validate your model with proper metrics.

All the steps need to be submitted with explanations.