

MG 222: Assignment II 2018

Due Date: March 12, 2018

The text file in http://www.mgmt.iisc.ernet.in/CM/MG222/Data_Files/auto.data contains information on miles per gallon (**mpg**), number of engine **cylinders**, engine **displacement** in cubic inches, engine **horsepower**, body **weight** in lbs., **acceleration**, last two digits of the **year** of the model, country of **origin** (coded as 1 for American, 2 for European and 3 for Japanese), and the **name** or make of 397 passenger cars. The dependent variable of interest *i.e.* the model sought, is that for a car's mileage (**mpg**), relating it to the remaining independent variables (how many are there?). Among these independent variables, we have no interest in **year** and **name** of the cars. Answer the following:

Visualization It might be impossible to graphically display the simultaneous effects of all the independent variables on the dependent variable in a single image. With the aid of a minimum number of graphical images, displaying the simultaneous effects of as many of the independent variables as possible (think colors, plotting symbols, line types etc.), briefly explain how the mileage of a car (**mpg**) is related to which of the independent variables. [10]

Model Building After accounting for the non-linear effects (if any) and the interactions (if any), present among the independent variables, as revealed through the preliminary visualization step above (or even otherwise, if your prior engineering knowledge compels you to do so), build the “best possible” regression model you can, according to the AIC criterion. You should automate this model building process with the aid of the R function **stepAIC** (residing in the library **MASS**), a successful deployment of which requires preliminary experimentation with a few “small” and a few “large” initial models, between which the “best” model is chosen; but must write one paragraph clearly describing the underlying algorithm, and more importantly its statistical logic, for selecting the “best possible” model. [10+5=15]

Diagnostics Carry out a detailed diagnostic check through residual analysis of the fitted model, ensuring that the data fits the model well, conforming to all the underlying model assumptions. [8]

Conclusion Based on the fitted model and your graphical (as well as subjective engineering) understanding, explain how the different independent variables considered in this study, are “affecting” the gas mileage of different cars. [7]