Midterm 1 Review

Wells

9/25/2020

The Data

The `mtcars` data set was extracted from the 1974 Motor Trend US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973–74 models).

You can load and view the data using the following code

```
data(mtcars)
View(mtcars)
```

Exploration

Perform some preliminary data exploration and analysis. What types of variables are present (quantitative vs. categorical)? Which variables seem correlated? Do any of the variables seem to have a linear relationship with `mpg`? Non-linear relationship?

Model Building

Suppose you are interested in learning more about which features increase fuel efficiency of cars in the 1970s.

Build a variety of models for `mpg` as functions of the remaining variables, using tools we've discussed in class. Consider interaction terms, variable selection, and transformations.

Which model seems to have the most accuracy on the training data?

What effect might some of your modeling choices have on accuracy on a test data set? Consider the Bias-Variance tradeoff.

Model Diagnostics

Perfom diagnostics on several of the models you explored previously. Which seem to fit the assumptions for MLR? For those that do not satisfy assumptions, are there any fixes you can implement so that the models/data do satisfy the assumptions? What effect does this have on training accuracy? What effect might this have on test accuracy?