Back Next Page 1

7.1: Introduction to the Lesson

Print view Index of pages

Topic 7.1: Introduction to the Lesson

In building a regression model, we are faced with two conflicting objectives: (1) to include more predictors into the model so as to improve the predictive ability of the model and (2) to not include predictors that are unnecessary, which will lead to more uncertainty in our predictions and make our model needlessly complicated.

For example, recall the gas mileage example that you have seen and worked on in the previous modules. In the data set, you are presented with 11 potential predictors to predict the gas mileage of vehicles. You are faced with the question of which and how many of the predictors you will need to include in the model. Do we start by using just horsepower and weight of the vehicle, or should we start by also including its transmission type? Or why not just use all of the predictors right away? While using all the predictors may seem like an easy choice, needlessly adding predictors actually adds additional uncertainty in our predictions and makes the model needlessly complicated. Typically, a reasonable model is built first, and is then assessed in terms of meeting the regression assumptions. In this lesson, you will learn the various ways to aid in selecting this first model.

Chapter 2.1.1 from this book, Introduction to Statistical Learning (James et. al.), provides a nice discussion about the two conflicting objectives. There should be a link to download the pdf of the book for free.

Back

Next