Print

Model Assessment and Selection

During the model fitting phase of predictive modeling, you use the training data to create a variety of different models, and possibly use several model selection methods. In this example, we'll use the forward selection method to generate several candidate models. In forward selection, variables are added into the model as long as they meet the criterion for inclusion. For example, with the AICC criterion, variables will be added as long as the criterion value continues to decrease. When the value can no longer be reduced, indicating that the model can no longer be improved, the process stops.

Here five possible models of increasing complexity were generated. For simplicity, the complexity of each model is indicated by a number from 1 to 5. Next, the validation data is used to assess the performance of each model. These performance measures are then used as the criteria for selecting the best model. In this simplified example, a star rating indicates how well the model fits the validation data. The validation assessment rating ranges from one star to three stars. Model 1 received a one-star rating, models 2 and 5 received two stars, and models 3 and 4 received three stars, indicating that they are the best models. But what if you need to choose only one?

When there's a tie or a near tie, the most parsimonious model is typically chosen in other words, the simplest model with the highest validation assessment. Recall that, in this example, the models are listed in order of increasing complexity.This means that model 3 is simpler than model 4, so you choose model 3.

Close