NBA 4920/6921 Lecture 11

Linear Model Stepwise Selection

Murat Unal

Johnson Graduate School of Management

10/05/2021

Agenda

Quiz 9

Best subset selection

Stepwise selection
- Forward stepwise selection
- Backward stepwise selection

Application in R

Model selection

Best subset selection:

The idea is to estimate a model for every possible subset of variables; then compare their performances

Model selection

Best subset selection:

1. Let M_0 denote the null model, which contains no predictors.
2. For k in 1 to p:
   - Fit every possible model with k variables
   - Let M_k denote the **best** model with k variables
   - Here best is defined as having the smallest RSS, or equivalently largest R^2
3. Select the **best** model from $M_0, \cdots, M_p$ using cross-validated prediction error
4. Train the chosen model on the full dataset

Model selection

Best subset selection:

Problem?

Model selection

Best subset selection:

Problem?

- $p = 10 \rightsquigarrow$ fitting 1,024 models
- $p = 25 \rightsquigarrow$ fitting ≈ 33.5 mil models

Stepwise selection

Stepwise selection: reduces the computational burden of best subset selection

The main idea is:

1. Start by fitting an arbitrary model
2. Improve the model by adding/removing variables one-at-a-time
3. Select the best fitting model

Model selection

Two main strategies for stepwise selection are:

1. **Forward**: start with only intercept model M_0, and add variables one-at-a-time
2. **Backward**: start with all variables model M_p, and remove variables one-at-a-time

Forward stepwise selection

Forward stepwise selection:

Start with the null model, and add predictors one-at-a-time, until all p variables are in the model

At each step the variable that give the greatest improvement in model fit is added

Forward stepwise selectionn

Forward stepwise selection:

1. Let M_0 denote the null model, which contains no predictors.
2. For k in 0 to $p-1$:
   - Consider all $p-k$ models that augment the variables in M_k with one additional predictor
   - Choose the **best** among these $p-k$ models and call it M_{k+1}
   - Here best is defined as having the smallest RSS, or equivalently largest R^2
3. Select the **best** model from $M_0, \cdots, M_p$ using cross-validated prediction error
4. Train the chosen model on the full dataset

Forward stepwise selection

Forward stepwise selection:

Problem? It is not guaranteed to find the best possible model out of all 2^p models containing subsets of the p predictors

Model selection

Best subset and forward stepwise selection for Credit dataset

# Variables	Best subset	Forward stepwise
One	rating	rating
Two	rating, income	rating, income
Three	rating, income, student	rating, income, student
Four	cards, income, student, limit	rating, income, student, limit

Source: ISL

Backward stepwise selection

Backward stepwise selection:

Start with the full model containing all p predictors, and then remove the least useful predictor one-at-a-time

Backward stepwise selection:

1. Let M_p denote the full model, which contains all predictors.
2. For k in p to 1:
   - Consider all k models that contain all but one of the predictors in M_k, for a total of $k-1$ predictors
   - Choose the **best** among these k models and call it M_{k-1}
   - Here best is defined as having the smallest RSS, or equivalently largest R^2
3. Select the **best** model from $M_0, \cdots, M_p$ using cross-validated prediction error
4. Train the chosen model on the full dataset

Backward stepwise selection:

Problem? It is not guaranteed to find the best possible model out of all 2^p models containing subsets of the p predictors

References

Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani (2017)
An Introduction to Statistical Learning
Springer.
https://www.statlearning.com/

Ed Rubin (2020)
Economics 524 (424): Prediction and Machine-Learning in Econometrics
Univ, of Oregon.