Regression Analysis

Model Selection

Nicoleta Serban, Ph.D.
Professor
Stewart School of Industrial and Systems Engineering

Model Search

Georgia Tech

About This Lesson

Georgia Tech

Bias-Variance Tradeoff

- **Variable Selection:** Bias vs. Variance
  - Many covariates
    - Low bias, high variance
  - Few covariates
    - High bias, low variance
  - Too few covariates
    - High bias, high variance
- **Prediction Risk:** Measure of the Bias-Variance Tradeoff

  $$R(S) = \frac{1}{n}\sum_{i=1}^{n} E(\hat{\boldsymbol{Y}}_i(S) - \boldsymbol{Y}_i^*)^2$$

  with $\hat{\boldsymbol{Y}}_i(S)$ the fitted response for submodel S and $\boldsymbol{Y}_i^*$ the future observation

Given an estimate of the prediction risk for a submodel S, choose the submodel with the smallest prediction risk.

➔ ***How to search over all submodels?***

Georgia Tech

Model Search

- If p is the number of predicting variables, there are $\mathbf{2^p}$ possible submodels
  - If p is small
    - Fit all submodels
  - If p is large
    - Search using heuristics/greedy search
- **Stepwise Regression**
  - Forward
    - Start with no predictors, add one at a time
  - Backward
    - Start with all predictors, drop one at a time
  - Forward-Backward
    - Add and drop one variable at a time iteratively

Georgia Tech

Model Search

- Stepwise regression is a greedy algorithm. It does not guarantee to find the model with the best score.
- Forward stepwise regression is preferable to backward stepwise regression.
- Forward stepwise regression does not necessarily select the same model as the one selected using backward stepwise regression.

Georgia Tech

Forward Stepwise Regression

1. Select criterion for model selection (e.g., AIC)
2. Establish minimum model, and compute its criterion value, C_0
3. Fit p marginal regressions for p predictors, V_j $(j = 1, \cdots, p)$, that are not in minimum model
   - C_j is the criterion value for the model that includes the j-th predictor, V_j
   - If possible, select predictor $P_1 = V_k$ whose inclusion yields the smallest criterion value where $C_k < C_0$
   - If P_1 exists, add it to the minimum model and continue; otherwise, stop
4. Fit p-1 regressions, and use the same method to test if another predictor should be added
   - Regressions will now be based on models with the previous predictors, including P_1, and with each V_j additionally included one at a time, for $j = 1, \cdots, (k-1), (k+1), \cdots, p$
   - If possible, select predictor $P_2 = V_l$ whose inclusion yields the smallest criterion value where $C_l < C_k$
     - C_l is based on the current regressions; C_k is based on the regressions from the previous step
   - If P_2 exists, add it to the model and continue; otherwise, stop
5. Continue adding predictors one at a time until the criterion does not improve

Georgia Tech

Backward Stepwise Regression

1. Select criterion for model selection (e.g., AIC)
2. Establish the minimum model and the predictors that must be included
3. Fit full model with p additional predictors not in the minimum model, V_j $(j = 1, \cdots, p)$, and compute its criterion value, C_{F}
4. Fit p regressions, removing one predictor, V_j $(j = 1, \cdots, p)$, each time
   - C_j is the criterion value for the model that excludes the j-th predictor, V_j
   - If possible, select predictor $P_1 = V_k$ whose removal yields the smallest criterion value where $C_k \leq C_{\mathrm{F}}$
   - If P_1 exists, remove it from the full model and continue; otherwise, stop
5. Fit p-1 regressions, and use the same method to test if another predictor should be removed
   - Regressions will now be based on models with the previous predictors, excluding P_1, and with each remaining V_j removed one at a time, for $j = 1, \cdots, (k-1), (k+1), \cdots, p$
   - If possible, select $P_2 = V_l$ whose removal yields the smallest criterion value where $C_l \leq C_k$
     - C_l is based on the current regressions; C_k is based on the regressions from the previous step
   - If P_2 exists, remove it from the model and continue; otherwise, stop
6. Continue discarding predictors one at a time until the criterion does not improve

Georgia Tech

Forward vs Backward Stepwise Regression

Backward stepwise regression:

- Cannot be performed if there are more predictors than the sample size $(\mathrm{p} > \mathrm{n})$
- Is more computationally expensive than forward stepwise regression
- Will select larger models if p is large

Georgia Tech

Summary