STAT 213
Model Selection II

Colin Reimer Dawson

Oberlin College

March 30, 2018

Outline

Outline

So many models...

- How to decide among all these models?
  1. Understand the subject area! Build sensible models.
  2. Nested F-tests
  3. Model quality measures

What Makes a Good Model?

Fit

High R^2
Small SSE
Large F

Validity

Strong evidence for predictors
Simple (Parsimonious)
Generalizes outside sample

Why Does Parsimony Matter?

Don't we just care about good predictions?

Not exclusively...

- We also use models to *understand* the world (harder with more complexity)

And even so...

- We really care about making predictions for data we *haven't seen yet*.

Criteria to "score" models

1. high R^2/low SSE/low $\hat{\sigma}_\varepsilon^2$: always prefers more complex models
2. Adj. R^2: balances fit and complexity
3. Mallow's C_p / Akaike Information Criterion (AIC): estimates mean squared prediction error based on $\hat{\sigma}_\varepsilon^2$ from a "full" model
4. Out-of-sample predictive accuracy (next time)

Mallow's C_p / AIC

Two measures that reduce to the same thing in the case of MLR with independent, equal variance, Normal residuals. For a "reduced" model with $p_{reduced}$ total parameters (including the intercept) which is nested in a "full" model with p_{full} parameters, both fit using n observations:

$$C_p = \frac{SSE_{\text{reduced}}}{MSE_{\text{full}}} + 2p_{\text{reduced}} - n \tag{1}$$

$$= p_{\text{reduced}} + \frac{SSE_{diff}}{MSE_{full}} \tag{2}$$

where smaller values indicate a simpler model (smaller $p_{reduced}$) and/or a better fit (smaller SSE_{diff})

Outline

Model Selection

Five predictor-selection methods:

1. Domain knowledge (+ a few F-tests)
2. Best subset
3. Forward selection
4. Backward selection
5. Stepwise selection

Automated exploration of predictor subsets

1. Best subset: consider all possible combinations (2^K)
2. Forward selection: start with null model, and consider adding one predictor at a time
3. Backward elimination: start with full model and consider removing one predictor at a time
4. Stepwise regression: consider both additions and subtractions at each iteration

Note: Choose best step based on adj-R^2 or C_p/AIC, *not* based on P-values

Model Selection

		"Scoring"		
		$R^2_{adj.}$	C_p	CV Error (next time)
"Search"	Domain Knowledge			
	Best Subset			
	Forward Selection			
	Backward Selection			
	Stepwise Selection			

Example: Baseball Win %

Demo