Stepwise Selection

DS 301

Iowa State University

Stepwise Selection (large p)

- For computational reasons, best subset selection cannot be applied when p is large. ($p > 30$)
- We need an alternative: **stepwise selection**
- These methods are based on *greedy algorithms*.
- Greedy here means we make the choice of 'best' model based on our current information, but this is not necessarily the 'best' global choice.
- That means, unlike subset selection, we do not consider all possible models.
- Therefore, greedy decisions may lead to a suboptimal choice of model.

local decision

Forward stepwise selection

Algorithm:

Y~1.

1. Let M_0 denote the null model, which contains no predictors.
2. For $k = 0, \ldots, p-1$
   a. Consider all $p - k$ models that augment the predictors in M_k with one additional predictor.
   b. Choose the best among these $p - k$ models and call it M_{k+1}. Here best is defined as the model with the smallest RSS.
3. Select a best model $M_1, \ldots, M_p$ using cross-validated error, AIC, BIC, Mallow's C_p, or adjusted R^2.

Example

X_1, X_2, X_3, X_4

$M_0: Y \sim 1$ (Step 1).

Step 2: $Y \sim X_1$

$Y \sim X_2 \longrightarrow M_1$

$Y \sim X_3$

$Y \sim X_4$

Step 3:

$Y \sim X_2 + X_1$

$Y \sim X_2 + X_3 \longrightarrow M_2$

$Y \sim X_2 + X_4$

Step 4:

$Y \sim X_2 + X_3 + X_1$

$Y \sim X_2 + X_3 + X_4 \longrightarrow M_3$

Step 5:

$Y \sim X_2 + X_3 + X_4 + X_1 \longrightarrow M_4$

M_1, M_2, M_3, M_4

select a final model among these.

Backward stepwise selection

- Same algorithm as forward selection, but now direction moves backward.
- Initial model: full model
- Iteratively remove each predictor one at a time.

Forward vs. backward

- If your goal is prediction and a more complex model is acceptable: backward selection.
- If, all things considered, you would prefer a simpler model: forward selection.
- High-dimensional data: backward selection.

$p > n$. breaks down

in this class: $p < n$.

$n = 200$, $p = 12$.

Hybrid approach (stepwise selection)

- Best subset, forward, and backward stepwise will generally give similar but not identical models.
- An alternative is a hybrid version of forward/backward selection: stepwise selection.
- At each step, you are not restricted to one direction. You can add and remove predictors.
- Stepwise selection attempts to mimic subset selection while still retaining computational advantage.

hybrid approach

See `R` script: `stepwise.R`