

# Model Search Methods

## Best Subset Selection

1. Let  $M_0$  denote null model which contains no predictors. This model simply predicts the of the response for each observation.
2. For  $k = 1, 2, \dots, p$ 
  - Fit all  $\binom{p}{k}$  models that contain exactly  $p$  predictors
  - Pick the best among these  $\binom{p}{k}$  models and call it  $M_k$ .

Here, *best* is defined as having the smallest RSS or largest  $R^2$
3. Select a single best model from among  $M_0, M_1, \dots, M_p$  using cross validated prediction error,  $C_p$  (*AIC*), *BIC*, or Adjusted- $R^2$

requires training  $2^p$  models

### Example

$p = 3$

$M_0$ : intercept only (null)

$C_1$ :  $X_1$   $X_2$   $X_3$



lowest training RSS within  $C_1$

$\Rightarrow M_1$

$C_2$ :  $X_1, X_2$   $X_1, X_3$   $X_2, X_3$



lowest training RSS within  $C_2$

$\Rightarrow M_2$

$M_3$ : full model with

$X_1$   $X_2$   $X_3$

# Model Search Methods

## Forward Stepwise Selection

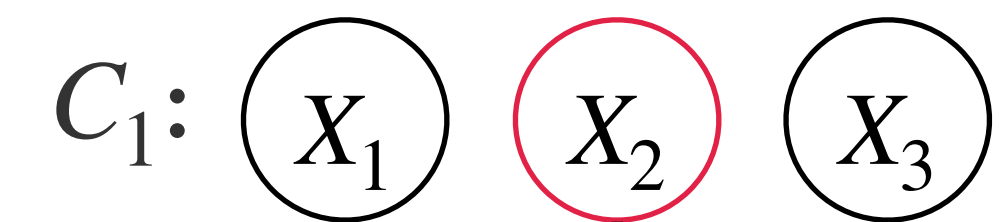
1. Let  $M_0$  denote null model which contains no predictors.
2. For  $k = 1, 2, \dots, p - 1$ 
  - ▶ Consider all  $p - k$  models that augment the predictors in  $M_k$  with one additional predictor
  - ▶ Choose the best among these  $p - k$  models and call it  $M_{k+1}$ .  
Here, *best* is defined as having the smallest RSS or largest  $R^2$
3. Select a single best model from among  $M_0, M_1, \dots, M_p$  using cross validated prediction error,  $C_p$  (AIC), BIC, or Adjusted- $R^2$

requires training  $1 + \frac{p(p+1)}{2}$  models

### Example

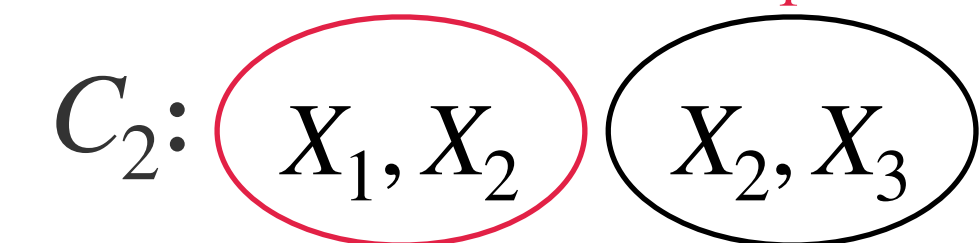
$$p = 3$$

$M_0$ : intercept only (null)



lowest training RSS within  $C_1$

$\Rightarrow M_1$



lowest training RSS within  $C_2$

$\Rightarrow M_2$

$M_3$ : full model with

