CS6301: ADVANCED COMPUTATIONAL METHODS FOR DATA SCIENCE

Lecture 11: High Dimensional Problems I
Algorithm 6.1 Best subset selection

1. Let $\mathcal{M}_0$ denote the null model, which contains no predictors. This model simply predicts the sample mean for each observation.

2. For $k = 1, 2, \ldots, p$:
   
   (a) Fit all $\binom{p}{k}$ models that contain exactly $k$ predictors.

   (b) Pick the best among these $\binom{p}{k}$ models, and call it $\mathcal{M}_k$. Here best is defined as having the smallest RSS, or equivalently largest $R^2$.

3. Select a single best model from among $\mathcal{M}_0, \ldots, \mathcal{M}_p$ using cross-validated prediction error, $C_p$ (AIC), BIC, or adjusted $R^2$. 
**Algorithm 6.2 Forward stepwise selection**

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 having smallest RSS or highest $R^2$.

3. Select a single best model from among $M_0, \ldots, M_p$ using cross-validated prediction error, $C_p$ (AIC), BIC, or adjusted $R^2$.  

FORWARD STEPWISE SELECTION
Algorithm 6.3 Backward stepwise selection

1. Let $M_p$ denote the full model, which contains all $p$ predictors.

2. For $k = p, p - 1, \ldots, 1$:
   
   (a) Consider all $k$ models that contain all but one of the predictors in $M_k$, for a total of $k - 1$ predictors.
   
   (b) Choose the best among these $k$ models, and call it $M_{k-1}$. Here best is defined as having smallest RSS or highest $R^2$.

3. Select a single best model from among $M_0, \ldots, M_p$ using cross-validated prediction error, $C_p$ (AIC), BIC, or adjusted $R^2$. 

BACKWARDS STEPWISE SELECTION
\[ C_p = \frac{1}{n} (\text{RSS} + 2d\sigma^2), \]

\[ \text{Adjusted } R^2 = 1 - \frac{\text{RSS}/(n - d - 1)}{\text{TSS}/(n - 1)}. \]

\[ \text{AIC} = \frac{1}{n\sigma^2} (\text{RSS} + 2d\sigma^2), \]

\[ \text{BIC} = \frac{1}{n} (\text{RSS} + \log(n)d\sigma^2). \]
MODEL ASSESSMENT
RIDGE REGRESSION

$$\text{RSS} = \sum_{i=1}^{n} \left( y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2.$$ 

$$\sum_{i=1}^{n} \left( y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 = \text{RSS} + \lambda \sum_{j=1}^{p} \beta_j^2.$$
RIDGE REGRESSION