Start-Tech Academy

Other Linear Regression

Subset Selection

This approach involves identifying a subset of the *p* predictors that we believe to be related to the response.

We will be discussing these three techniques of Subset selection

1. Best Subset Selection
2. Forward Step-wise selection
3. Backward Step-wise selection

Other Linear Regression

Best Subset Selection

In Best Subset selection technique, we will be training the model on all possible subsets of p variables. For p variables the total number of subsets are 2^p

1. Let *M0* denote the *null model* , which contains no predictors. This model simply predicts the sample mean for each observation.
2. For *k* = 1, 2, . . .*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 *Mk*. Here *best* is defined as having the smallest RSS, or equivalently largest *R*^2.
3. Select a single best model from among *M0*, . . . ,*Mp* using adjusted *R*^2.

Other Linear Regression

Forward Step-wise selection

In Forward stepwise selection technique, we start training the model with no predictors and continue to add predictors one by one till we have added all of them.

For p variables the total number of models are $1 + p(p+1)/2$

1. Let *M*0 denote the *null* model, which contains no predictors.
2. For *k* = 0*, . . . , p* − 1:
   A. Consider all *p* − *k* models that augment the predictors in *Mk* with one additional predictor.
   B. Choose the *best* among these *p* − *k* models, and call it *Mk*+1. Here *best* is defined as having smallest RSS or highest *R*^2.
3. Select a single best model from among *M*0*, . . . ,Mp* using adjusted *R*^2.

Other Linear Regression

Backward Step-wise selection

In Backward stepwise selection technique, we start training the model with all the predictors and continue to remove them one by one till we have removed all of them.

For p variables the total number of models are $1+p(p+1)/2$

1. Let *Mp* denote the *full* model, which contains all *p* predictors.
2. For *k* = *p, p* − 1*, . . . ,* 1:
   1. Consider all *k* models that contain all but one of the predictors in *Mk*, for a total of *k* − 1 predictors.
   2. Choose the *best* among these *k* models, and call it *Mk*−1. Here *best* is defined as having smallest RSS or highest *R*^2.
3. Select a single best model from among *M*0*, . . . ,Mp* using adjusted *R*^2.

If n is less than p backward stepwise selection will not work , onlt forward selection technique can work in this case