Review: Forward stagewise regression and the monotone lasso

Jieun Shin

May 24, 2023

1 Contribution

- LARS and forward stagewise algorithm for solving penalized least square regression problems
- study a condition under which the coefficient paths of lasso are monotone

2 Background

- *least angle regression*: the piecewise linear nature of the lasso profiles. simultaneously solving the entire set of lasso problem
  1. Standardize the predictors to have mean zero and variance 1. Start with the residual $\boldsymbol{r} = \boldsymbol{y} - \bar{\boldsymbol{y}}, \beta_1, \beta_2, \ldots, \beta_p = 0$.
  2. Find the predictor $\boldsymbol{x}_j$ most correlated with $\boldsymbol{r}$
  3. Move β_j from 0 towards its least-squares coefficient $< \boldsymbol{x}_j, \boldsymbol{r} >$, until some other competitor $\boldsymbol{x}_k$ has as much correlation with the current residual as does $\boldsymbol{x}_j$
  4. Move (β_j, β_k) in the direction defined by their joint least squares coefficient of the current residual on $(\boldsymbol{x}_j, \boldsymbol{x}_k)$, until some other competitor $\boldsymbol{x}_l$ has as much correlation with the current residual.
  5. Continue in this way until all p predictors have been entered. After p steps, we arrive at the full least-squares solution.

  현재 residual과 가장 상관성이 높은 변수의 계수값을 더 큰 상관을 가지는 변수가 나타날 때까지 이동해간다. 모든 변수가 다 들어올 때까지 이 과정을 반복한다.

- *incremental forward stagewise* algorithm (FS_ϵ): the lasso coefficient profile produced by a version of boosting for linear models
  1. Start with $\boldsymbol{r} = \boldsymbol{y} - \bar{\boldsymbol{y}}, \beta_1, \beta_2, \ldots, \beta_p = 0$.
  2. Find the predictor $\boldsymbol{x}_j$ most correlated with $\boldsymbol{r}$.
  3. Update $\beta_j \leftarrow \beta_j + \delta_j$, where $\delta_j = \epsilon \cdot \text{sign}[\text{corr}(\boldsymbol{r}, \boldsymbol{x}_j]$.
  4. Update $\boldsymbol{r} \leftarrow \boldsymbol{r} - \delta_j \boldsymbol{x}_j$, and repeat steps 2 and 3 until no predictor has any correlation with $\mathbf{r}$

  현재 residual과 가장 상관성이 높은 계수에 ϵ을 더한다. $\epsilon \to 0$로 제한한 버전이 forward stagewise 알고리즘이며, 어떤 조건 하에서 LASSO path와 같아진다.

3 Forward Stagewise and the Monotone Lasso

we create an expanded data matrix $\tilde{\boldsymbol{X}} = [\boldsymbol{X} : -\boldsymbol{X}]$. The lasso problem becomes

$$\min_{\beta_0, \beta_j^+, \beta_j^-} \sum_{i=1}^{n} \left(y_i - \beta_0 - \left[\sum_{j=1}^{p} x_{ij}\beta_j^+ - \sum_{j=1}^{p} x_{ij}\beta_j^- \right] \right)^2 \tag{1}$$

$$\text{subject to } \beta_j^+, \beta_j^- \geq 0, \forall j \text{ and } \sum_{j=1}^{p} (\beta_j^+ + \beta_j^-) \leq s \tag{2}$$

forward-stagewise을 더 부드러운 계수 프로파일을 가진 lasso로 푸는 방법이다. 독립변수를 확장한 것은 부스팅기법에서 tree의 binary search와 유사한 기법이다.

1. Start with $\boldsymbol{r} = \boldsymbol{y} - \bar{\boldsymbol{y}}, \beta_1, \beta_2, \ldots, \beta_p = 0$.
2. Find the predictor $\boldsymbol{x}_j$ most correlated with $\boldsymbol{r}$.
3. Update $\beta_j \leftarrow \beta_j + \epsilon$.
4. Update $\boldsymbol{r} \leftarrow \boldsymbol{r} - \epsilon \boldsymbol{x}_j$, and repeat steps 2 and 3 until no predictor has any correlation with $\mathbf{r}$