Introduction to Knock-off Filters

"Controlling False Discovery Rates via Knock-offs "
Authors: Rina Barber and Emmanuel Candes

Nandana Sengupta

October 18, 2015

Introduction

"All models are wrong, but some models are useful" – George Box

Introduction

"All models are wrong, but some models are useful" – George Box

- Consider the simple linear regession model:

$$y = X\beta + \varepsilon; \qquad y \in \mathbb{R}^n, \quad X \in \mathbb{R}^{n \times p}, \quad \varepsilon \sim N(0, \sigma^2)$$

$$\Rightarrow \hat{\beta}_{ols} = \underset{\beta}{\operatorname{argmin}} \, (y - X\beta)' \, (y - X\beta)$$

Introduction

"All models are wrong, but some models are useful" – George Box

- Consider the simple linear regession model:

$$y = X\beta + \varepsilon; \qquad y \in \mathbb{R}^n, \quad X \in \mathbb{R}^{n \times p}, \quad \varepsilon \sim N(0, \sigma^2)$$

$$\Rightarrow \hat{\beta}_{ols} = \underset{\beta}{\operatorname{argmin}} \, (y - X\beta)' \, (y - X\beta)$$

  - Large number of X: low bias but larger variance

Introduction

"All models are wrong, but some models are useful" – George Box

- Consider the simple linear regession model:

$$y = X\beta + \varepsilon; \qquad y \in \mathbb{R}^n, \quad X \in \mathbb{R}^{n \times p}, \quad \varepsilon \sim N(0, \sigma^2)$$

$$\Rightarrow \hat{\beta}_{ols} = \underset{\beta}{\operatorname{argmin}} \, (y - X\beta)' \, (y - X\beta)$$

  - Large number of X: low bias but larger variance
  - Small number of X: higher bias but small variance

Introduction

"All models are wrong, but some models are useful" – George Box

- Consider the simple linear regession model:

$$y = X\beta + \varepsilon; \qquad y \in \mathbb{R}^n, \quad X \in \mathbb{R}^{n\times p}, \quad \varepsilon \sim N(0, \sigma^2)$$

$$\Rightarrow \hat{\beta}_{ols} = \underset{\beta}{\operatorname{argmin}}\, (y - X\beta)'\,(y - X\beta)$$

  - Large number of X: low bias but larger variance
  - Small number of X: higher bias but small variance
  - Ideal: Small set of X truly associated with y

Introduction

"All models are wrong, but some models are useful" – George Box

- Consider the simple linear regession model:

$$y = X\beta + \varepsilon; \qquad y \in \mathbb{R}^n, \quad X \in \mathbb{R}^{n\times p}, \quad \varepsilon \sim N(0, \sigma^2)$$

$$\Rightarrow \hat{\beta}_{ols} = \underset{\beta}{\operatorname{argmin}} \, (y - X\beta)' \, (y - X\beta)$$

  - Large number of X: low bias but larger variance
  - Small number of X: higher bias but small variance
  - Ideal: Small set of X truly associated with y

Introduction

"All models are wrong, but some models are useful" – George Box

- Consider the simple linear regession model:

$$y = X\beta + \varepsilon; \qquad y \in \mathbb{R}^n, \quad X \in \mathbb{R}^{n \times p}, \quad \varepsilon \sim N(0, \sigma^2)$$

$$\Rightarrow \hat{\beta}_{ols} = \underset{\beta}{\operatorname{argmin}} \, (y - X\beta)' \, (y - X\beta)$$

  - Large number of X: low bias but larger variance
  - Small number of X: higher bias but small variance
  - Ideal: Small set of X truly associated with y

- Motivating example from genetics:

Introduction

"All models are wrong, but some models are useful" – George Box

- Consider the simple linear regession model:

$$y = X\beta + \varepsilon; \qquad y \in \mathbb{R}^n, \quad X \in \mathbb{R}^{n\times p}, \quad \varepsilon \sim N(0, \sigma^2)$$

$$\Rightarrow \hat{\beta}_{ols} = \underset{\beta}{\operatorname{argmin}} \, (y - X\beta)' \, (y - X\beta)$$

  - Large number of X: low bias but larger variance
  - Small number of X: higher bias but small variance
  - Ideal: Small set of X truly associated with y

- Motivating example from genetics:
  - y: phenotype (observable traits eg: eye color, height)

Introduction

"All models are wrong, but some models are useful" – George Box

- Consider the simple linear regession model:

$$y = X\beta + \varepsilon; \qquad y \in \mathbb{R}^n, \quad X \in \mathbb{R}^{n \times p}, \quad \varepsilon \sim N(0, \sigma^2)$$

$$\Rightarrow \hat{\beta}_{ols} = \underset{\beta}{\operatorname{argmin}} \, (y - X\beta)' \, (y - X\beta)$$

  - Large number of X: low bias but larger variance
  - Small number of X: higher bias but small variance
  - Ideal: Small set of X truly associated with y

- Motivating example from genetics:
  - y: phenotype (observable traits eg: eye color, height)
  - X: genes

Introduction

"All models are wrong, but some models are useful" – George Box

- Consider the simple linear regession model:

$$y = X\beta + \varepsilon; \qquad y \in \mathbb{R}^n, \quad X \in \mathbb{R}^{n\times p}, \quad \varepsilon \sim N(0, \sigma^2)$$

$$\Rightarrow \hat{\beta}_{ols} = \underset{\beta}{\operatorname{argmin}} \, (y - X\beta)' \, (y - X\beta)$$

  - Large number of X: low bias but larger variance
  - Small number of X: higher bias but small variance
  - **Ideal: Small set of X truly associated with y $\Longleftarrow$**

- Motivating example from genetics:
  - y: phenotype (observable traits eg: eye color, height)
  - X: genes

Preliminary Concept 1: Variable Selection Techniques

"Selecting a Small Subset of Variables"

Preliminary Concept 1: Variable Selection Techniques

"Selecting a Small Subset of Variables"

- Forward Stepwise Regression

Preliminary Concept 1: Variable Selection Techniques

"Selecting a Small Subset of Variables"

- Forward Stepwise Regression
- Backward Stepwise Regression

Preliminary Concept 1: Variable Selection Techniques

"Selecting a Small Subset of Variables"

- Forward Stepwise Regression
- Backward Stepwise Regression
- $\Rightarrow$**LASSO**

Preliminary Concept 1: Variable Selection Techniques

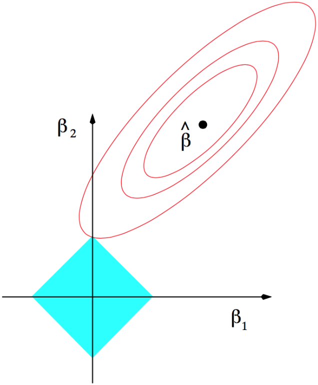
"Selecting a Small Subset of Variables"

- Forward Stepwise Regression
- Backward Stepwise Regression
- ⇒**LASSO**
  - $\hat{\beta}_{lasso} = \underset{\beta}{\operatorname{argmin}} \, (y - X\beta)' \, (y - X\beta) + \lambda \sum_{j=1}^{p} |\beta_j|$

Preliminary Concept 1: Variable Selection Techniques

"Selecting a Small Subset of Variables"

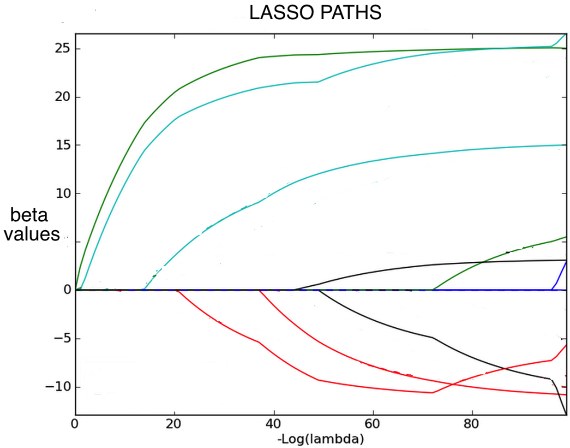
- Forward Stepwise Regression
- Backward Stepwise Regression
- ⇒**LASSO**
  - $\hat{\beta}_{lasso} = \underset{\beta}{\operatorname{argmin}} \, (y - X\beta)' \, (y - X\beta) + \lambda \sum_{j=1}^{p} |\beta_j|$



-

Preliminary Concept 1: Variable Selection Techniques

"Selecting a Small Subset of Variables"



Preliminary Concept 2: False Discovery Rate

"Selecting Variables **truly associated** with y. "

Preliminary Concept 2: False Discovery Rate

"Selecting Variables **truly associated** with y. "

- Null hypothesis $\mathcal{H}_0 : \beta_j = 0$

Preliminary Concept 2: False Discovery Rate

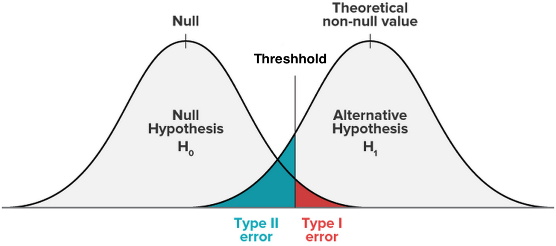
"Selecting Variables **truly associated** with y. "

- Null hypothesis $\mathcal{H}_0 : \beta_j = 0$
- Set of selected covariates: $\mathcal{S}$

Preliminary Concept 2: False Discovery Rate

"Selecting Variables **truly associated** with y. "

- Null hypothesis $\mathcal{H}_0 : \beta_j = 0$
- Set of selected covariates: $\mathcal{S}$

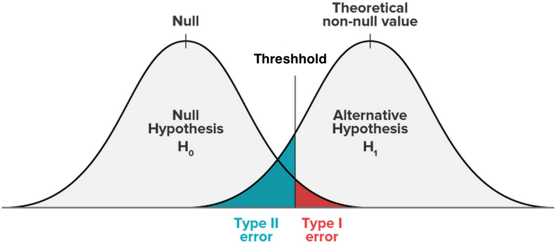


-

Preliminary Concept 2: False Discovery Rate

"Selecting Variables **truly associated** with y. "

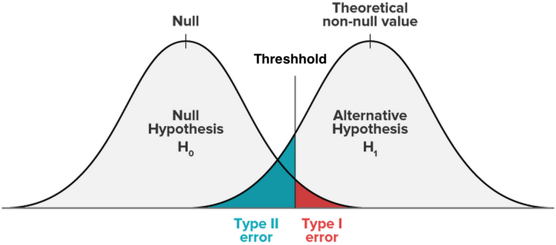
- Null hypothesis $\mathcal{H}_0 : \beta_j = 0$
- Set of selected covariates: $\mathcal{S}$



- Want to control the proportion of Type I error

Preliminary Concept 2: False Discovery Rate

"Selecting Variables **truly associated** with y. "

- Null hypothesis $\mathcal{H}_0 : \beta_j = 0$
- Set of selected covariates: $\mathcal{S}$
- 

- Want to control the proportion of Type I error

$$\underset{\substack{\uparrow \\ \text{False discovery rate}}}{\text{FDR}} = \mathbb{E}\left[\underbrace{\frac{\#\text{ false positives}}{\text{total \# of features selected}}}_{\text{False discovery proportion}}\right] = \mathbb{E}\left[\frac{|S \cap \mathcal{H}_0|}{|S|}\right].$$

-

Knock-off Filters: Algorithm

Knock-off Filters: Algorithm

- Construct Knock-offs $\tilde{X}$ such that

Knock-off Filters: Algorithm

- Construct Knock-offs $\tilde{X}$ such that
  - $\tilde{X} = X(I - \Sigma^{-1} diag(s)) + \tilde{U}C$

Knock-off Filters: Algorithm

- Construct Knock-offs $\tilde{X}$ such that
  - $\tilde{X} = X(I - \Sigma^{-1} diag(s)) + \tilde{U}C$
  - Correlation structure: $\tilde{X}'\tilde{X} = X'X, \qquad \forall j, k$

Knock-off Filters: Algorithm

- Construct Knock-offs $\tilde{X}$ such that
  - $\tilde{X} = X(I - \Sigma^{-1} diag(s)) + \tilde{U}C$
  - Correlation structure: $\tilde{X}'\tilde{X} = X'X, \qquad \forall j, k$
  - Correlation structure: $\tilde{X}'X = X'X, \qquad \forall j = k$

Knock-off Filters: Algorithm

- Construct Knock-offs $\tilde{X}$ such that
  - $\tilde{X} = X(I - \Sigma^{-1} diag(s)) + \tilde{U}C$
  - Correlation structure: $\tilde{X}'\tilde{X} = X'X, \qquad \forall j, k$
  - Correlation structure: $\tilde{X}'X = X'X, \qquad \forall j = k$
  - Augmented Matrix: $[X \quad \tilde{X}]$

Knock-off Filters: Algorithm

- Construct Knock-offs $\tilde{X}$ such that
  - $\tilde{X} = X(I - \Sigma^{-1} diag(s)) + \tilde{U}C$
  - Correlation structure: $\tilde{X}'\tilde{X} = X'X, \qquad \forall j,k$
  - Correlation structure: $\tilde{X}'X = X'X, \qquad \forall j = k$
  - Augmented Matrix: $[X \quad \tilde{X}]$
- Compute Lasso with Augmented Matrix

$$\beta_\lambda = \arg\min_{\beta \in \mathbb{R}^{2p}} \left\{ \frac{1}{2} \left\| y - [X \ \ \widetilde{X}] \cdot \beta \right\|_2^2 + \lambda \left\| \beta \right\|_1 \right\}$$

Knock-off Filters: Algorithm

- Construct Knock-offs $\tilde{X}$ such that
  - $\tilde{X} = X(I - \Sigma^{-1} diag(s)) + \tilde{U}C$
  - Correlation structure: $\tilde{X}'\tilde{X} = X'X, \qquad \forall j, k$
  - Correlation structure: $\tilde{X}'X = X'X, \qquad \forall j = k$
  - Augmented Matrix: $[X \quad \tilde{X}]$
- Compute Lasso with Augmented Matrix

$$\beta_\lambda = \arg\min_{\beta \in \mathbb{R}^{2p}} \left\{ \frac{1}{2} \left\| y - [X \ \widetilde{X}] \cdot \beta \right\|_2^2 + \lambda \left\| \beta \right\|_1 \right\}$$

- For each pair of knock-off and original variables, calculate

Knock-off Filters: Algorithm

- Construct Knock-offs $\tilde{X}$ such that
  - $\tilde{X} = X(I - \Sigma^{-1} diag(s)) + \tilde{U}C$
  - Correlation structure: $\tilde{X}'\tilde{X} = X'X, \qquad \forall j, k$
  - Correlation structure: $\tilde{X}'X = X'X, \qquad \forall j = k$
  - Augmented Matrix: $[X \quad \tilde{X}]$
- Compute Lasso with Augmented Matrix

$$\beta_\lambda = \underset{\beta \in \mathbb{R}^{2p}}{\arg\min} \left\{ \frac{1}{2} \left\| y - \left[X \ \widetilde{X}\right] \cdot \beta \right\|_2^2 + \lambda \left\| \beta \right\|_1 \right\}$$

- For each pair of knock-off and original variables, calculate

$$\lambda_j = \sup \left\{ \lambda \ : \ \beta_j^\lambda \neq 0 \right\} = \text{ first time } X_j \text{ enters Lasso path}$$

$$\widetilde{\lambda}_j = \sup \left\{ \lambda \ : \ \widetilde{\beta}_j^\lambda \neq 0 \right\} = \text{ first time } \widetilde{X}_j \text{ enters Lasso path}$$

- $$W_j = \max\{\lambda_j, \widetilde{\lambda}_j\} \cdot \text{sign}(\lambda_j - \widetilde{\lambda}_j)$$

Knock-off Filters: Algorithm

Knock-off Filters: Algorithm

- For each λ value calculate

Knock-off Filters: Algorithm

- For each λ value calculate
-

$$\begin{aligned} \text{Selected variables:} \quad & S_\lambda = \{j : W_j \geq +\lambda\} \\ \text{Control group:} \quad & \widetilde{S}_\lambda = \{j : W_j \leq -\lambda\} \end{aligned} \quad \rightsquigarrow \quad \widehat{\mathsf{FDP}}(S_\lambda) := \frac{|\widetilde{S}_\lambda|}{|S_\lambda|}$$

Knock-off Filters: Algorithm

- For each λ value calculate
- $$\begin{aligned} \text{Selected variables:} \quad & S_\lambda = \{j : W_j \geq +\lambda\} \\ \text{Control group:} \quad & \widetilde{S}_\lambda = \{j : W_j \leq -\lambda\} \end{aligned} \quad \rightsquigarrow \quad \widehat{\mathsf{FDP}}(S_\lambda) := \frac{|\widetilde{S}_\lambda|}{|S_\lambda|}$$
- Choose threshold level q

Knock-off Filters: Algorithm

- For each λ value calculate
-

$$\begin{aligned} \text{Selected variables:} \quad & S_\lambda = \{j : W_j \geq +\lambda\} \\ \text{Control group:} \quad & \widetilde{S}_\lambda = \{j : W_j \leq -\lambda\} \end{aligned} \quad \rightsquigarrow \quad \widehat{\mathsf{FDP}}(S_\lambda) := \frac{|\widetilde{S}_\lambda|}{|S_\lambda|}$$

- Choose threshold level q
- Select the variables based on

Knock-off Filters: Algorithm

- For each λ value calculate
- $$\begin{aligned} \text{Selected variables:} \quad & S_\lambda = \{j : W_j \geq +\lambda\} \\ \text{Control group:} \quad & \widetilde{S}_\lambda = \{j : W_j \leq -\lambda\} \end{aligned} \quad \rightsquigarrow \quad \widehat{\mathsf{FDP}}(S_\lambda) := \frac{|\widetilde{S}_\lambda|}{|S_\lambda|}$$
- Choose threshold level q
- Select the variables based on
- $$\Lambda = min\{\lambda : \widehat{FDP}(S_\lambda) \leq q\}$$
$$S_\Lambda = \{j : W_j \geq \Lambda\}$$

Knock-off Filters: Algorithm

- For each λ value calculate
-

$$\begin{array}{rl} \text{Selected variables:} & S_\lambda = \{j : W_j \geq +\lambda\} \\ \text{Control group:} & \widetilde{S}_\lambda = \{j : W_j \leq -\lambda\} \end{array} \quad \rightsquigarrow \quad \widehat{\mathrm{FDP}}(S_\lambda) := \frac{|\widetilde{S}_\lambda|}{|S_\lambda|}$$

- Choose threshold level q
- Select the variables based on
-

$$\Lambda = min\{\lambda : \widehat{FDP}(S_\lambda) \leq q\}$$

$$S_\Lambda = \{j : W_j \geq \Lambda\}$$

- Knock-off+ filter:

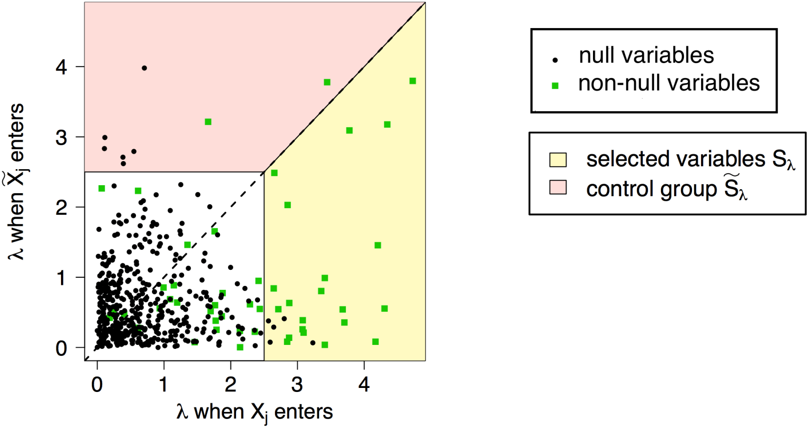
Knock-off Filters: Algorithm

- For each λ value calculate
- $$\begin{aligned} \text{Selected variables:} \quad & S_\lambda = \{j : W_j \geq +\lambda\} \\ \text{Control group:} \quad & \widetilde{S}_\lambda = \{j : W_j \leq -\lambda\} \end{aligned} \quad \rightsquigarrow \quad \widehat{\mathsf{FDP}}(S_\lambda) := \frac{|\widetilde{S}_\lambda|}{|S_\lambda|}$$
- Choose threshold level q
- Select the variables based on
- $$\Lambda = min\{\lambda : \widehat{FDP}(S_\lambda) \leq q\}$$
  $$S_\Lambda = \{j : W_j \geq \Lambda\}$$
- Knock-off+ filter:

$$\widehat{\mathsf{FDP}}_+(S_\lambda) := \frac{|\widetilde{S}_\lambda| + 1}{|S_\lambda|}$$

-

Knock-off Filters: Intuition



Rest of the paper

Rest of the paper

- Theoretical Guarantees

Rest of the paper

- Theoretical Guarantees
  -

Rest of the paper

- Theoretical Guarantees
  -
  -

Rest of the paper

- Theoretical Guarantees
  -
  -
- Simulation Results

Rest of the paper

- Theoretical Guarantees
  -
  -
- Simulation Results
  -

Rest of the paper

- Theoretical Guarantees
  -
  -
- Simulation Results
  -
  -

Rest of the paper

- Theoretical Guarantees
  -
  -
- Simulation Results
  -
  -
- Empirical Application

Rest of the paper

- Theoretical Guarantees
    -
    -
- Simulation Results
    -
    -
- Empirical Application
    -

Going Further: Issues and Possible Applications

Going Further: Issues and Possible Applications

-

Going Further: Issues and Possible Applications

-
-

Going Further: Issues and Possible Applications

-
-

Going Further: Issues and Possible Applications

Going Further: Issues and Possible Applications

Going Further: Issues and Possible Applications

Going Further: Issues and Possible Applications

Going Further: Issues and Possible Applications

Going Further: Issues and Possible Applications

Going Further: Issues and Possible Applications

-
-
-
  -
  -
  -
-
-
  -
  -

Going Further: Issues and Possible Applications

Thanks!