High-Dimensional Variable Selection for Competing Risks with Cooperative Penalized Regression

«CooPeR»

Lukas Burk[1,2,3,4] Andreas Bender[2,4] Marvin N. Wright[1,3]

[1]Leibniz Institute for Prevention Research and Epidemiology – BIPS
[2]LMU Munich
[3]University of Bremen
[4]Munich Center for Machine Learning

CEN 2023

Introduction

- Settings with two competing events $e \in \{1, 2\}$, e.g.,
  (1) Death from bladder cancer
  (2) Death from other causes

Introduction

- Settings with two competing events $e \in \{1, 2\}$, e.g.,
  (1) Death from bladder cancer
  (2) Death from other causes
- Common approach:
  - Fit cause-specific model for event of interest
  - Treats other event as censored
    $\Rightarrow$ loses information

Introduction

- Settings with two competing events $e \in \{1, 2\}$, e.g.,
  (1) Death from bladder cancer
  (2) Death from other causes
- Common approach:
  - Fit cause-specific model for event of interest
  - Treats other event as censored
    $\Rightarrow$ loses information

Main goal: Fit cause-specific model for event 1 using shared information from event 2

Elastic Net

Objective function with negative log-likelihood contribution for observation i:

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \quad \sum_{i=1}^{n} \ell(y_i, \mathbf{x}_i^\top, \boldsymbol{\beta}) + \lambda \sum_{j=1}^{p} \left(\alpha |\beta_j| + \frac{1-\alpha}{2} \beta_j^2 \right)$$

Elastic Net

Objective function with negative log-likelihood contribution for observation i:

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \quad \sum_{i=1}^{n} \ell(y_i, \mathbf{x}_i^\top, \boldsymbol{\beta}) + \sum_{j=1}^{p} \lambda_j \left(\alpha |\beta_j| + \frac{1-\alpha}{2} \beta_j^2 \right)$$

Feature-Weighted Elastic Net

Feature-weighted elastic net[1] extends objective function:

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \quad \sum_{i=1}^{n} \ell(y_i, \mathbf{x}_i^\top, \boldsymbol{\beta}) + \lambda \sum_{j=1}^{p} w_j(\theta) \left(\alpha |\beta_j| + \frac{1-\alpha}{2} \beta_j^2 \right)$$

[1] Tay et al. (2023)

Feature-Weighted Elastic Net

Feature-weighted elastic net[1] extends objective function:

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \quad \sum_{i=1}^{n} \ell(y_i, \mathbf{x}_i^\top, \boldsymbol{\beta}) + \lambda \sum_{j=1}^{p} w_j(\theta) \left(\alpha |\beta_j| + \frac{1-\alpha}{2} \beta_j^2 \right)$$

$$w_j(\theta) = \frac{\sum_{l=1}^{p} \exp(\mathbf{z}_l^\top \theta)}{p \exp(\mathbf{z}_j^\top \theta)}$$

[1] Tay et al. (2023)

Feature-Weighted Elastic Net

Incorporate prior external information with $\mathbf{Z} \in \mathbb{R}^{p \times K}$

Feature-Weighted Elastic Net

Incorporate prior external information with $\mathbf{Z} \in \mathbb{R}^{p \times K}$

Grouping: $\mathbf{Z} \in \mathbb{R}^{5 \times 2}$

$$\mathbf{Z} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}$$

Feature-Weighted Elastic Net

Incorporate prior external information with $\mathbf{Z} \in \mathbb{R}^{p \times K}$

Grouping: $\mathbf{Z} \in \mathbb{R}^{5 \times 2}$

$$\mathbf{Z} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}$$

Individual weighting: $\mathbf{Z} \in \mathbb{R}^{5 \times 1}$

$$\mathbf{Z} = \begin{pmatrix} 1.5 \\ 1 \\ 1.2 \\ 0.7 \\ 0.3 \end{pmatrix}$$

Feature-Weighted Elastic Net

$$w_j(\theta) = \frac{\sum_{l=1}^{p} \exp(\mathbf{z}_l^\top \theta)}{p \exp(\mathbf{z}_j^\top \theta)}$$

Feature-Weighted Elastic Net

$$w_j(\theta) = \frac{\sum_{l=1}^{p} \exp(\mathbf{z}_l^\top \theta)}{p \exp(\mathbf{z}_j^\top \theta)}$$

- $\mathbf{z}_j^\top \theta$: “Importance score”
- Larger value $\Rightarrow$ lower w_j

Feature-Weighted Elastic Net

$$w_j(\theta) = \frac{\sum_{l=1}^{p} \exp(\mathbf{z}_l^\top \theta)}{p \exp(\mathbf{z}_j^\top \theta)}$$

- $\mathbf{z}_j^\top \theta$: “Importance score”
- Larger value $\Rightarrow$ lower w_j

- $\theta \in \mathbb{R}^{K \times 1}$ fit internally
- $\theta = 0 \Rightarrow w_j = 1$

Individual Feature Weighting

- Simulation from Tay et al.: $\mathbf{Z}$ set to noisy version of true $|\boldsymbol{\beta}|$
- $|\beta_j|$ large $\Rightarrow$ weaker penalization for $\widehat{\beta}_j$
- $|\beta_j|$ small $\Rightarrow$ stronger penalization for $\widehat{\beta}_j$

“Cooperative Penalized Regression” (CooPeR)

Based on multi-task algorithm suggested by Tay et al.

“Cooperative Penalized Regression” (CooPeR)

Based on multi-task algorithm suggested by Tay et al.

1. Set $\widehat{\beta}_1^{(0)}, \widehat{\beta}_2^{(0)}$ to elastic net solution for $(\mathbf{X}, \mathbf{y}_1), (\mathbf{X}, \mathbf{y}_2)$ with $\mathbf{y}_e := (\mathbf{t}_e, \delta_e)$

“Cooperative Penalized Regression” (CooPeR)

Based on multi-task algorithm suggested by Tay et al.

1. Set $\hat{\beta}_1^{(0)}, \hat{\beta}_2^{(0)}$ to elastic net solution for $(\mathbf{X}, \mathbf{y}_1), (\mathbf{X}, \mathbf{y}_2)$ with $\mathbf{y}_e := (\mathbf{t}_e, \delta_e)$
2. For $k = 0, 1, \ldots$ until stopped:

a) Fit $\mathrm{fwelnet}\left(\mathbf{X}, \mathbf{y}_2, \mathbf{Z}_2 = \left|\hat{\beta}_1^{(k)}\right|\right)$ to determine $\left|\hat{\beta}_2^{(k+1)}\right|$

“Cooperative Penalized Regression” (CooPeR)

Based on multi-task algorithm suggested by Tay et al.

1. Set $\hat{\beta}_1^{(0)}, \hat{\beta}_2^{(0)}$ to elastic net solution for $(\mathbf{X}, \mathbf{y}_1), (\mathbf{X}, \mathbf{y}_2)$ with $\mathbf{y}_e := (\mathbf{t}_e, \delta_e)$
2. For $k = 0, 1, \ldots$ until stopped:

a) Fit $\text{fwelnet}\left(\mathbf{X}, \mathbf{y}_2, \mathbf{Z}_2 = \left|\hat{\beta}_1^{(k)}\right|\right)$ to determine $\left|\hat{\beta}_2^{(k+1)}\right|$

b) Fit $\text{fwelnet}\left(\mathbf{X}, \mathbf{y}_1, \mathbf{Z}_1 = \left|\hat{\beta}_2^{(k+1)}\right|\right)$ to determine $\left|\hat{\beta}_1^{(k+1)}\right|$

Simulation Study

- Simulation adapted from Binder et al.[2]
- Mimics gene expression data
- Comparison with CoxBoost[3], Random Survival Forests [4]

[2] Binder, Allignol, et al. (2009)
[3] Binder and Schumacher (2008)
[4] Ishwaran et al. (2014)

Simulation Study

- Simulation adapted from Binder et al.[2]
- Mimics gene expression data
- Comparison with CoxBoost[3], Random Survival Forests [4]
- n = 400, p = 5000,
- 4 covariate blocks
- 4 informative variables each

[2]Binder, Allignol, et al. (2009)
[3]Binder and Schumacher (2008)
[4]Ishwaran et al. (2014)

Assignment of True Effects

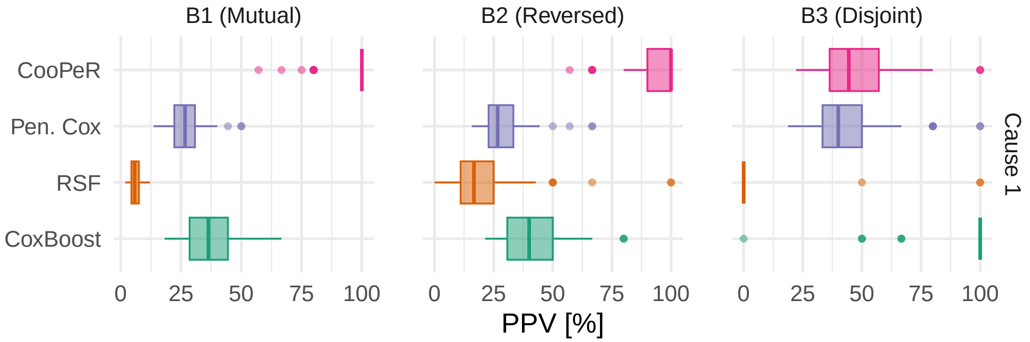
- Block 1 (**Mutual**): Same effect on both cause-specific hazards
- Block 2 (**Reversed**): Cause 1 (+) Cause 2 (-)
- Block 3 (**Disjoint**): Cause 1 or 2

Assignment of True Effects

- Block 1 (**Mutual**): Same effect on both cause-specific hazards
- Block 2 (**Reversed**): Cause 1 (+) Cause 2 (-)
- Block 3 (**Disjoint**): Cause 1 or 2
- Block 4 (**Cor. Noise**)
- Rest: Uncorrelated noise

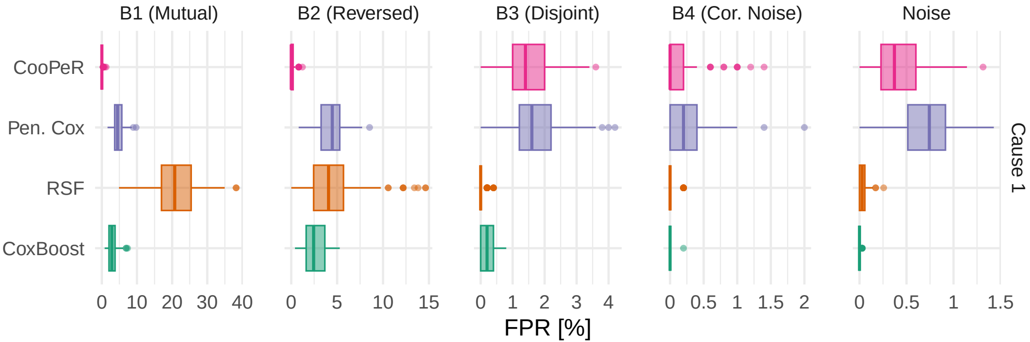
Positive Predictive Value

Probability a selected variable is informative



False Positive Rate

Proportion of noise variables falsely selected



Application on Bladder Cancer Data

- Clinical & gene expression features [5]

[5] Dyrskjøt et al. (2005)

Application on Bladder Cancer Data

- Clinical & gene expression features [5]
- Proxy to estimate variable selection performance:
  1. Apply algorithms for variable selection
  2. Fit standard cause-specific Cox model using only selected variables
  3. Evaluate prediction performance (Brier(t), AUC(t))

[5] Dyrskjøt et al. (2005)

Application on Bladder Cancer Data

- Clinical & gene expression features [5]
- Proxy to estimate variable selection performance:
  1. Apply algorithms for variable selection
  2. Fit standard cause-specific Cox model using only selected variables
  3. Evaluate prediction performance (Brier(t), AUC(t))
- Results:
  - CooPeR-selected variables mostly identical to pen. Cox regression
  - Difference in metrics far from conclusive in either direction

[5] Dyrskjøt et al. (2005)

Application on Bladder Cancer Data

- Clinical & gene expression features [5]
- Proxy to estimate variable selection performance:
    1. Apply algorithms for variable selection
    2. Fit standard cause-specific Cox model using only selected variables
    3. Evaluate prediction performance (Brier(t), AUC(t))
- Results:
    - CooPeR-selected variables mostly identical to pen. Cox regression
    - Difference in metrics far from conclusive in either direction
- No shared effects? Effects too small?

[5] Dyrskjøt et al. (2005)

Conclusion & Outlook

- Promising variable selection behavior in simulations

Conclusion & Outlook

- Promising variable selection behavior in simulations
- So far no promising results on real data

Conclusion & Outlook

- Promising variable selection behavior in simulations
- So far no promising results on real data
- Lack of readily available high-dimensional data with competing risks

Conclusion & Outlook

- Promising variable selection behavior in simulations
- So far no promising results on real data
- Lack of readily available high-dimensional data with competing risks
- Generalization to $e > 2$ events: Unclear

Thank you for your attention!

www.leibniz-bips.de/en

Contact
Lukas Burk
Leibniz Institute for Prevention Research
and Epidemiology – BIPS
Achterstraße 30
D-28359 Bremen
burk@leibniz-bips.de

References I

- Binder, Harald, Arthur Allignol, et al. (Apr. 1, 2009). "Boosting for High-Dimensional Time-to-Event Data with Competing Risks". In: Bioinformatics 25.7, pp. 890–896. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btp088.
- Binder, Harald and Martin Schumacher (2008). "Adapting Prediction Error Estimates for Biased Complexity Selection in High-Dimensional Bootstrap Samples". In: Statistical Applications in Genetics and Molecular Biology 7.1, Article12. ISSN: 1544-6115. DOI: 10.2202/1544-6115.1346. pmid: 18384265.
- Dyrskjøt, Lars et al. (June 1, 2005). "A Molecular Signature in Superficial Bladder Carcinoma Predicts Clinical Outcome". In: Clinical Cancer Research: An Official Journal of the American Association for Cancer Research 11.11, pp. 4029–4036. ISSN: 1078-0432. DOI: 10.1158/1078-0432.CCR-04-2095. pmid: 15930337.

References II

- Ishwaran, Hemant et al. (Oct. 1, 2014). “Random Survival Forests for Competing Risks”. In: Biostatistics 15.4, pp. 757–773. ISSN: 1465-4644. DOI: 10.1093/biostatistics/kxu010.
- Tay, Jingyi Kenneth et al. (2023). “Feature-Weighted Elastic Net: Using ”Features of Features” for Better Prediction”. In: Statistica Sinica. ISSN: 10170405. DOI: 10.5705/ss.202020.0226.