Accounting for established predictors with the multi-step elastic net

Elizabeth Chase and Philip Boonstra, University of Michigan Department of Biostatistics

JSM, July 29, 2019

Motivation and Goal

Many prediction models are based upon a mixture of novel covariates and covariates that have been evaluated in earlier studies. However, the estimation process rarely takes that prior knowledge into account when fitting the model.

We offer a simple way to incorporate prior research about the relationship between some of the predictors and the outcome in a penalized regression setting.

Method

Let **y** be the outcome vector and **X** be the nxp design vector. We adapt the elastic net, which minimizes the criterion:

$$L(\lambda,\alpha,\beta)=\|\mathbf{y}-X\beta\|_2^2+(1-\alpha)\frac{\lambda}{2}\|\beta\|_2^2+\alpha\lambda\|\beta\|_1$$

where λ>0 is a tuning parameter for the amount of penalization and α in [0,1] tunes the mixture of ridge and lasso penalization.

In the multi-step elastic net, we instead minimize:

$$L(\lambda,\alpha,\phi,\beta)=\|\mathbf{y}-X\beta\|_2^2+(1-\alpha)\frac{\lambda}{2}(\phi\|\beta_1\|_2^2+\|\beta_2\|_2^2)+\alpha\lambda(\phi\|\beta_1\|_1+\|\beta_2\|_1)$$

where ϕ is a tuning parameter to control the amount of penalization on the established covariates.

Using cross-validation, we choose between four options for ϕ over a fixed grid of α, λ:

1. ϕ = 0 (no penalty on established)
2. ϕ = 1/16 (small penalty on established)
3. ϕ = ½ (half penalty on established)
4. ϕ = 1 (standard elastic net)

Simulation Study

Scen.	p_{est}	β_{est}	p_{unest}	β_{unest}
1A	10	All 0.26	30	0
1B	10	All 0.2	30	1 x 0.6
1C	10	All 0.25	30	5 x 0.05
2A	10	All 0.26	90	0
2B	10	All 0.2	90	1 x 0.6
2C	10	All 0.25	90	5 x 0.05
3A	20	Half 0.26	480	0
3B	20	Half 0.2	480	1 x 0.6
3C	20	Half 0.25	480	5 x 0.05
4A	20	All 0.13	480	0
4B	20	All 0.1	480	1 x 0.6
4C	20	All 0.13	480	5 x 0.05

All scenarios were performed at n = 200 or n = 1000, with predictors sampled from a MVN(0,1) distribution with compound symmetric ρ = 0.2 and intercept corresponding to 20% prevalence. We simulated 500 replicates for each scenario.

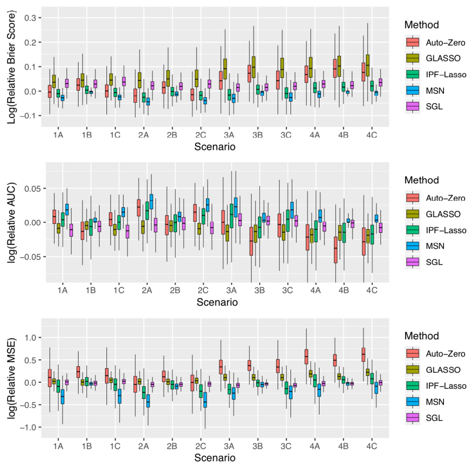

Figure 1. Simulation results for 500 replicates comparing performance of MSN to the Auto-Zero, GLASSO, IPF-Lasso, and SGL. All results are given relative to the elastic net's performance and then log-transformed.

Application

We apply the MSN to build a prediction model of mortality among pediatric ECMO patients. In 2016, Barbaro et al. built an initial prediction model using data from 1,611 patients and selected eleven predictors of mortality. In 2019, they sought to update their model with eleven novel biometric predictors, collected on a non-overlapping cohort of 178 patients.

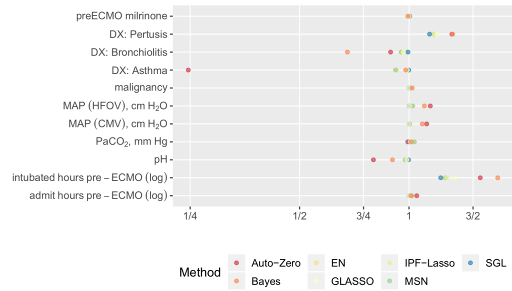

Figure 2. Established coefficient estimates for n = 178 pediatric ECMO patients, estimated using 7 different methods: Auto-Zero, Bayesian historical priors, EN, GLASSO, IPF-Lasso, MSN, and SGL.

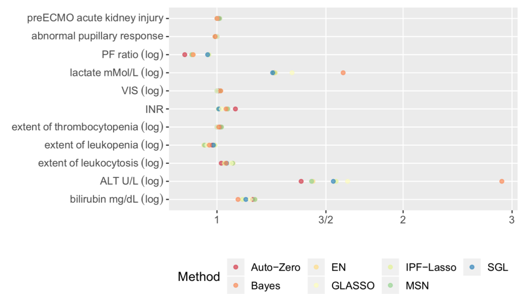

Figure 3. Unestablished coefficient estimates for n = 178 pediatric ECMO patients, estimated using 7 different methods: Auto-Zero, Bayesian historical priors, EN, GLASSO, IPF-Lasso, MSN, and SGL.

Discussion

- The MSN can be used when a subset of the predictors has already been evaluated in previous models.
- Including this extra information improves upon the elastic net's predictive and estimating performance.
- The MSN provides a simple way to use prior knowledge and improve model performance.

References

1. Tibshirani R. Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society: Series B.* 1996;58:267-288.
2. Yuan M, Lin Y. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B.*2006;68:49-67.
3. Zou H, Hastie T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B.* 2005;67:301-320.
4. Boulesteix AL, Bin R De, Jiang X, Fuchs M. IPF-LASSO: Integrative L1-penalized regression with penalty factors for prediction based on multi-omics data. *Computational and Mathematical Methods in Medicine.* 2017;Article ID 7691937.
5. Simon N, Friedman J, Hastie T, Tibshirani R. A sparse-group lasso. *Journal of Computational and Graphical Statistics.*2013;22:231-245.
6. Hoerl AE, Kennard RW. Ridge regression: biased estimation for nonorthogonal problems. *Technometrics.*1970;12:55-67.
7. Barbaro R, Boonstra P, Paden M, et al. Development and validation of the pediatric risk estimates core for children using extracorporeal respiratory support (PED-RESCUERS). *Intensive Care Medicine.* 2016;42:879-888.
8. Barbaro R, Boonstra P, Kuo K, et al. Evaluating mortality risk adjustment among children receiving extracorporeal support for respiratory failure. *ASAIO Journal.* 2019;65:277-284.
9. Boonstra PS, Barbaro RP. Incorporating historical models with adaptive Bayesian updates. *Biostatistics.* 2018;kxy053.

Acknowledgements

This work was supported by the National Institutes of Health (P30 CA046592; R01 CA129102; T32 CA083654) and the National Science Foundation (DGE-1256260).

If you have further questions, please contact Elizabeth Chase (ecchase@umich.edu). The code used to create the simulation, run all methods, and process results is accessible at https://github.com/psboonstra/MSEN.