Two-dimensional fused targeted ridge regression for health prediction from accelerometer data

Wessel N. van Wieringen
(joint work with Annelinde Lettink & Mai Chinapaw)

Dept. of Epidemiology & Data Science, Amsterdam UMC
Dept. of Mathematics, Vrije Universiteit Amsterdam
Amsterdam, the Netherlands

ISCB44, Milan, 28.08.2023

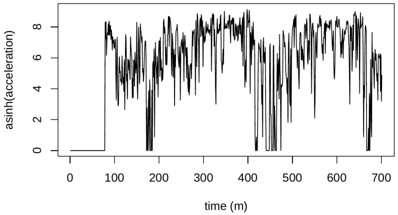
The NHANES study

Epidemiological study into health.

Some NHANES details:

- Two waves, C and D.
- Response: `health`.
- Covariates: `gender`, `age`,
- ... and `accelerometer profile`.

Accelerometer: fitbit that logs acceleration per (m)second.



NHANES: 24×7 acc. per min.

Goal

Does `accelerometer profile` aid in `health` prediction?

Can we underpin:

- platitudes like
  "*Sitting is the new smoking*."
- Dutch health council's recommendation:
  "*Weekly, at least* $2\frac{1}{2}$ *hour of moderate intensive movement.*"

Statistically, find:

$$\texttt{health} \quad = \quad f(\texttt{gender}, \texttt{age}, \ldots, \texttt{accelerometer profile}) + \varepsilon\text{rror}.$$

Accelerometer data: preprocessing

Paradigm
Humans behave like cars:
they operate in gears and shift them aperiodically.

segm.	log(#min.)	asinh(a)
1	5.966147	0.0000
2	4.718499	5.3977
3	5.347108	7.0285
4	3.891820	5.3266
5	3.135494	7.3644
...	...	...

From segments to relative frequencies of (length, acceleration)-boxes.

Model

The linear regression model:

$$\mathbf{Y} = \mathbf{U}\gamma + \mathbf{X}\beta + \varepsilon,$$

or the logistic regression model:

$$Y_i \,|\, \mathbf{U}_{i,*}, \mathbf{X}_{i,*} \sim \text{Bernoulli}\{\exp(\mathbf{U}_{i,*}\gamma + \mathbf{X}_{i,*}\beta)[1 + \exp(\mathbf{U}_{i,*}\gamma + \mathbf{X}_{i,*}\beta)]^{-1}\},$$

where

- $\mathbf{Y} \in \mathbb{R}^n$,
- $\mathbf{U} \in \mathcal{M}^{n,q}$ with $\text{rank}(\mathbf{U}) = q$,
- $\mathbf{X} \in \mathcal{M}^{n,p}$,
- $\gamma \in \mathbb{R}^q$,
- $\beta \in \mathbb{R}^p$, and
- $\varepsilon \sim \mathcal{N}(\mathbf{0}_n, \sigma^2 \mathbf{I}_{nn})$.

The design matrices $\mathbf{U}$ and $\mathbf{X}$ contain low- and high-dimensional covariates, e.g. $\{\texttt{age}, \texttt{gender}\}$ and `accelerometer profile`, respectively.

Estimator

The fused ridge regression estimator then is:

$$\hat{\gamma}(\cdot), \hat{\boldsymbol{\beta}}(\cdot) \;=\; \arg \min_{\boldsymbol{\gamma}\in\mathbb{R}^q, \boldsymbol{\beta}\in\mathbb{R}^p} \|\mathbf{Y} - \mathbf{U}\gamma - \mathbf{X}\boldsymbol{\beta}\|_2^2$$
$$+\lambda\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|_2^2 + \lambda_f(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^\top \boldsymbol{\Delta}(\boldsymbol{\beta} - \boldsymbol{\beta}_0),$$

with penalty parameters $\lambda, \lambda_f \in \mathbb{R}_{>0}$, fusion matrix $\boldsymbol{\Delta} \in \mathcal{S}_+^p$, and target $\boldsymbol{\beta}_0$.

And the fused ridge logistic regression estimator is:

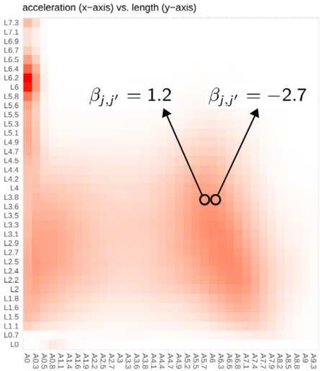
$$\hat{\gamma}(\cdot), \hat{\boldsymbol{\beta}}(\cdot) \;=\; \arg \max_{\boldsymbol{\gamma}\in\mathbb{R}^q, \boldsymbol{\beta}\in\mathbb{R}^p} \sum\nolimits_{i=1}^{n} Y_i(\mathbf{U}_{i,*}\gamma + \mathbf{X}_{i,*}\boldsymbol{\beta}) - \log[1 + \exp(\mathbf{U}_{i,*}\gamma + \mathbf{X}_{i,*}\boldsymbol{\beta})]$$
$$-\lambda\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|_2^2 - \lambda_f(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^\top \boldsymbol{\Delta}(\boldsymbol{\beta} - \boldsymbol{\beta}_0),$$

with $\lambda, \lambda_f, \boldsymbol{\Delta}$, and $\boldsymbol{\beta}$ as above.

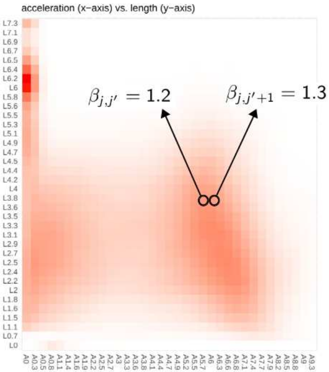
The estimator is found numerically by an IRLS algorithm.

Motivation for fusion

non-sensical



sensical



Axial fusion

Shrinkage of neighboring elements of $\boldsymbol{\beta}$ along axis.

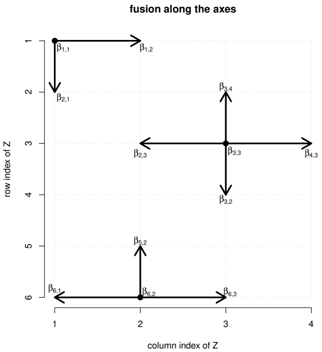
Penalty:

$$(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^\top \boldsymbol{\Delta}_a (\boldsymbol{\beta} - \boldsymbol{\beta}_0)$$

with $\boldsymbol{\Delta}_a$ s.t. penalty equals:

$$\sum_{j_r=1}^{p_r-1} \sum_{j_c=1}^{p_c-1} [(\beta_{j_r,j_c+1} - \beta_{j_r,j_c})^2 + (\beta_{j_r+1,j_c} - \beta_{j_r,j_c})^2] + \ldots,$$

where $\boldsymbol{\beta}_0 = \mathbf{0}_p$.



Also shrinkage of neighboring elements of $\boldsymbol{\beta}$ along *diagonals*.

Cross-validation

Penalty parameters λ and λ_f found by K-fold cross-validation.

Computationally efficient cross-validation.

Constrain the search to models with $\mathrm{df}(\lambda, \lambda_f) \leq \nu$ with $\nu \in \mathbb{N}$, where, e.g.

$$\mathrm{df}(\lambda, \lambda_f) = q + n - \mathrm{tr}\{[\mathbf{I}_{nn} + (\mathbf{I}_{nn} - \mathbf{P}_u)\mathbf{X}(\lambda \mathbf{I}_{pp} + \lambda_f \boldsymbol{\Delta})^{-1}\mathbf{X}^\top(\mathbf{I}_{nn} - \mathbf{P}_u)]^{-1}\},$$

with $\mathbf{P}_u = \mathbf{U}(\mathbf{U}^\top\mathbf{U})^{-1}\mathbf{U}^\top$,

Theoretically,

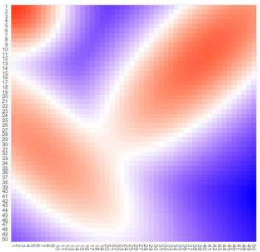
$$\lambda + \lambda_f(\mathbf{D}_\Delta)_{pp} \geq n^{-1}(\nu - q)^{-1}(n - \nu + q)\,\mathrm{tr}[(\mathbf{I}_{nn} - \mathbf{P}_u)\mathbf{X}\mathbf{X}^\top(\mathbf{I}_{nn} - \mathbf{P}_u)],$$

where $(\mathbf{D}_\Delta)_{pp}$ is the smallest eigenvalue of the matrix $\boldsymbol{\Delta}$.

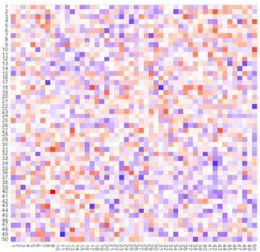
Simulation

- Draw $n = 100$ from $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$.
- Elements of $\boldsymbol{\beta}$ spatially related over 50×50 grid.
- Estimates with cross-validated penalty parameters.

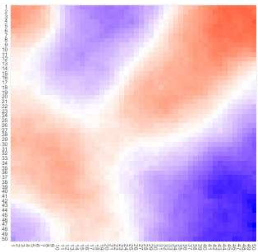
True $\boldsymbol{\beta}$



Reg. ridge $\hat{\boldsymbol{\beta}}$



Fused ridge $\hat{\boldsymbol{\beta}}$



- If $\boldsymbol{\beta}$ is spatially structured, fused ridge est. yields superior performance.

Study + analysis details

More NHANES details:

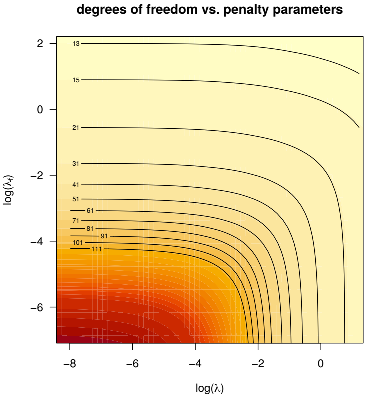
- $n = 2043$ healthy samples of wave C.
- $n = 2282$ healthy samples of wave D.
- $q = 10$ classical covariates.
- $p = 1260$ (pruned from a 41×40 grid).

Analysis details:

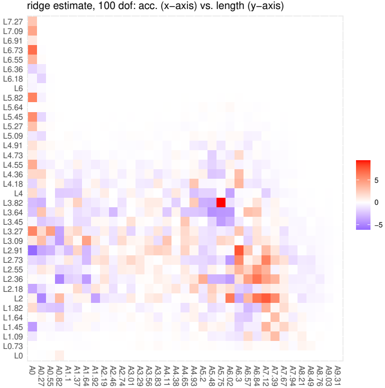
- Fusion matrix $\mathbf{\Delta} = \sqrt{2}\mathbf{\Delta}_a + \mathbf{\Delta}_d$.
- Outer loop split at $2:1$-ratio.
- $K = 10$-fold cross-validation for (λ, λ_f)-tuning.
- Performance: Spearman's ρ, Allen's PRESS.
- 50 outer splits.

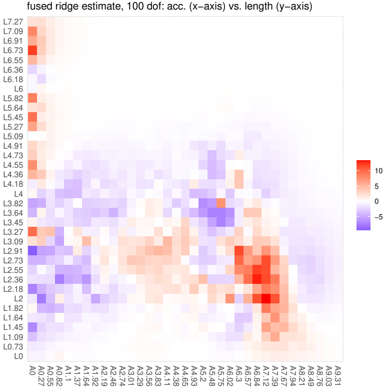
Degrees of freedom

Prior to estimation evaluate $df(\lambda, \lambda_f)$ for a (λ, λ_f)-grid.



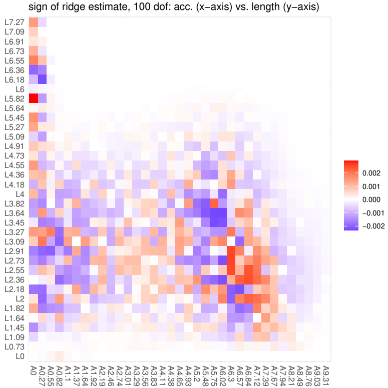
Estimator, regular vs. fused, size

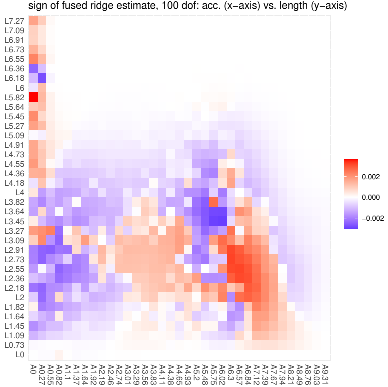




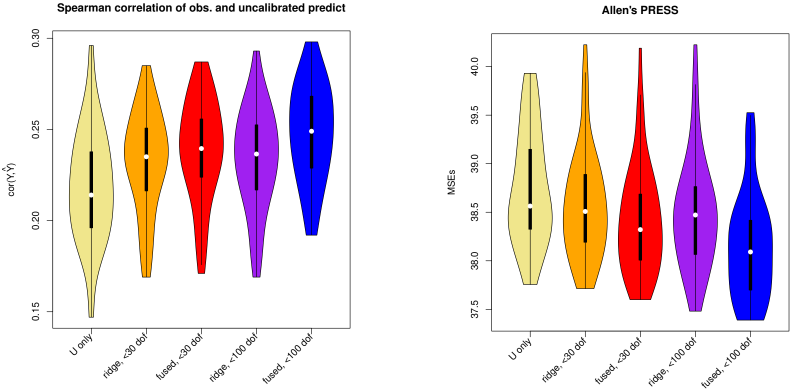
Corroborates with Newton's 2nd law: $F = m \times a$.

Estimator, regular vs. fused, weighted sign

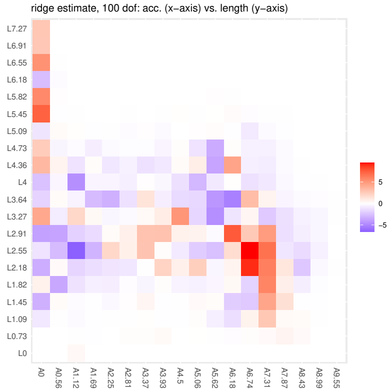


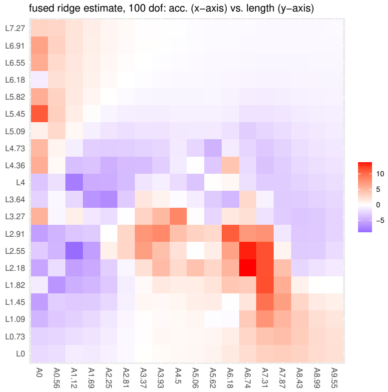


Performance, regular vs. fused



Estimator, regular vs. fused, size, different resolution



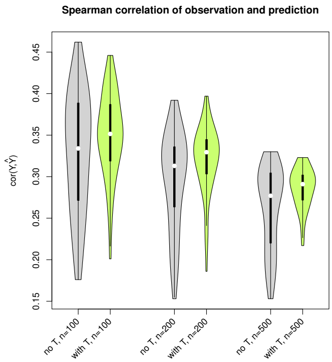


Updating

Use of target $\boldsymbol{\beta}_0$.

Details:

- Prediction of wave D `health`.
- Set $\boldsymbol{\beta}_0 = \hat{\boldsymbol{\beta}}_C(\lambda, \lambda_f)$.
- Training sample size $n = 100, 200$ and 500.
- $K = 10$-fold.
- Remaining samples form test data.
- Repeated 50 times.



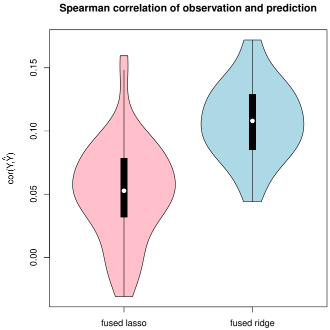
Comparison to 2-dim fused lasso

Details:

- The `genlasso`-package.
- Penalty: $\lambda_1 \|\boldsymbol{\beta}\|_1 + \gamma\lambda_1 \sum_{(u,u') \in \mathcal{E}} |\beta_u - \beta_{u'}|$ with $\mathcal{E}$ 'axial' edge set.
- No unpenalized covariates.
- No tuning of λ_1 and γ implemented.
- Small untuneable ridge penalty added if $p > n$.

Tweaks:

- Set $\gamma = 1 \quad$ (lasso).
- Set $\lambda = \lambda_f$ (ridge).
- Potential performance.



Conclusion

Lessons learned:

- Accelerometer information adds (a bit) in the prediction of `BMI`.
- Meager improvement of fused ridge prediction over regular ridge.
- The interpretation of the effect is enhanced by fused ridge estimation.
- Inclusion of a informed target is beneficial.
- Fused ridge estimation preferable over its lasso counterpart (this context).

Reference:

Lettink, A., Chinapaw, M., van Wieringen, W.N. (2023), "Two-dimensional fused targeted ridge regression for health prediction from accelerometer data", *Journal of the Royal Statistical Society, Series C*.

Software:

porridge-package: `https://CRAN.R-project.org/package=porridge`.