KU LEUVEN LRISK

Sparse modeling of risk factors in insurance analytics

Sander Devriendt

Joint work with Katrien Antonio, Edward Frees and Roel Verbelen

R in Insurance Conference
Paris, June 8, 2017

Outline

Motivation: actuarial models in insurance pricing

Problem: determine the (pure) premium π_i for insured i with

- number of claims N_i over exposure e_i,
- aggregate loss L_i over exposure e_i.

Decompose the premium in frequency and severity:

$$\pi_i = \mathsf{E}\left[\frac{L_i}{e_i}\right] = \mathsf{E}\left[\frac{N_i}{e_i}\right] \times \mathsf{E}\left[\frac{L_i}{N_i}\right] = \mathsf{E}\left[\mathsf{Freq}_i\right] \times \mathsf{E}\left[\mathsf{Sev}_i\right].$$

Classical assumption of independence allows for separate predictive modeling of $\mathsf{E}\left[\mathsf{Freq}_i\right]$ and $\mathsf{E}\left[\mathsf{Sev}_i\right]$.

Motivation: actuarial models in practice

In practice, insurers often use GLMs with observable risk factors:

- Continuous risk factors: age, experience, car power, ...
- Nominal (multi-level) risk factors: gender, fuel type, coverage type, car brand and model, ...
- Spatial risk factor (postal code), interactions, ...

Goals:

- use of GLM framework;
- data driven risk factor selection;
- data driven risk factor binning;
- transparent, communicable to insurers and insureds.

Motivation: actuarial models in practice

In practice, insurers often use GLMs with observable risk factors:

- Continuous risk factors: age, experience, car power, ...
- Nominal (multi-level) risk factors: gender, fuel type, coverage type, car brand and model, ...
- Spatial risk factor (postal code), interactions, ...

Goals:

- use of GLM framework;
- data driven risk factor selection;
- data driven risk factor binning;
- transparent, communicable to insurers and insureds.

Motivation: beyond current practice

Standard GLM binning algorithm:

1. A priori find the relevant risk factors and their bins.
   (e.g. through professional expertise)
2. Optimize the GLM loglikelihood to obtain the parameter for every bin.

A data driven GLM binning algorithm:

1. Make very small bins.
   (e.g. every age its specific bin)
2. Optimize the GLM loglikelihood while 'regularizing' the parameters to encourage selection and binning/fusion.

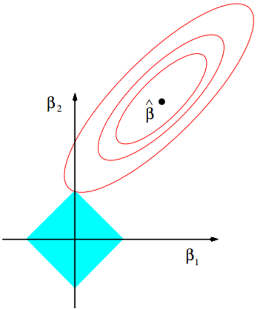
$$\mathcal{O}(\boldsymbol{\beta}) = -\ell(\boldsymbol{\beta}) + \lambda P(\boldsymbol{\beta}).$$

Regularization: the LASSO

2D example

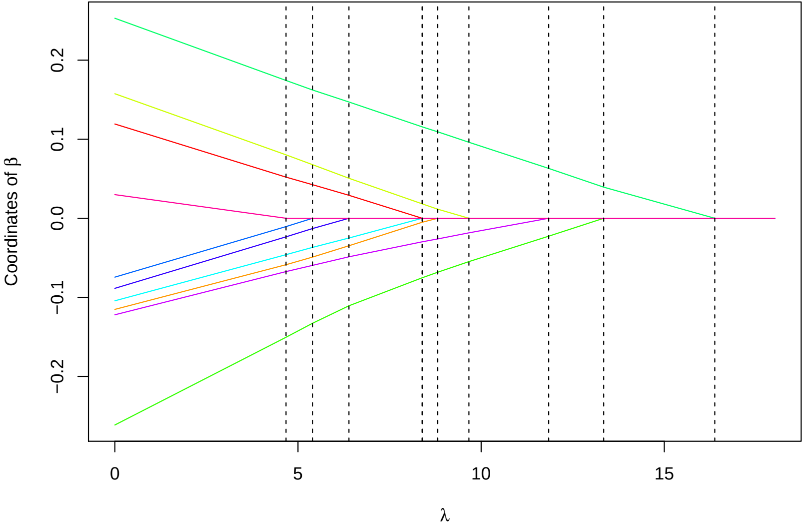
$$\mathcal{O}(\boldsymbol{\beta}) = -\ell(\boldsymbol{\beta}) + \lambda\left(|\beta_1| + |\beta_2|\right).$$

- Constraint is sharp, non-smooth.
- Encourages selection of either β_1 or β_2.
- Extensively studied and efficiently solved.



'The Elements of Statistical Learning' Hastie et al. (2009).

LASSO regularization



Beyond the LASSO

LASSO has been extensively studied and used (but largely unexplored in actuarial literature).

- DNA - gene selection (classical example).
- Portfolio selection: select the most important stocks for a certain strategy.

LASSO regularization is not fit for all types of variables, but can be adjusted to the type of risk factor. E.g. 'age', 'bm-scale'?

1. Determine the type of your risk factor.
2. Allocate a logical penalty to your risk factor.

Beyond the LASSO

LASSO has been extensively studied and used (but largely unexplored in actuarial literature).

- DNA - gene selection (classical example).
- Portfolio selection: select the most important stocks for a certain strategy.

LASSO regularization is not fit for all types of variables, but can be adjusted to the type of risk factor. E.g. ‘age’, ‘bm-scale’?

1. Determine the type of your risk factor.
2. Allocate a logical penalty to your risk factor.

Matching regularization to type of risk factor

- Ordinal risk factors (e.g. age): Fused Lasso

$$\lambda \sum_{i} w_i |\beta_{i+1} - \beta_i|.$$

- Nominal risk factors (e.g. car brand and model): Generalized Fused Lasso

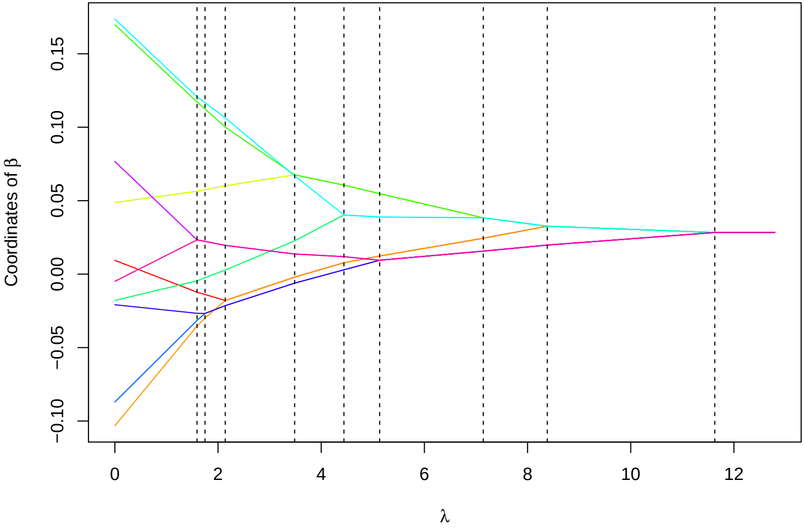
$$\lambda \sum_{i>k} w_{i,k} |\beta_i - \beta_k|.$$

- Spatial risk factors (e.g. postal code): Graph Guided Fused Lasso

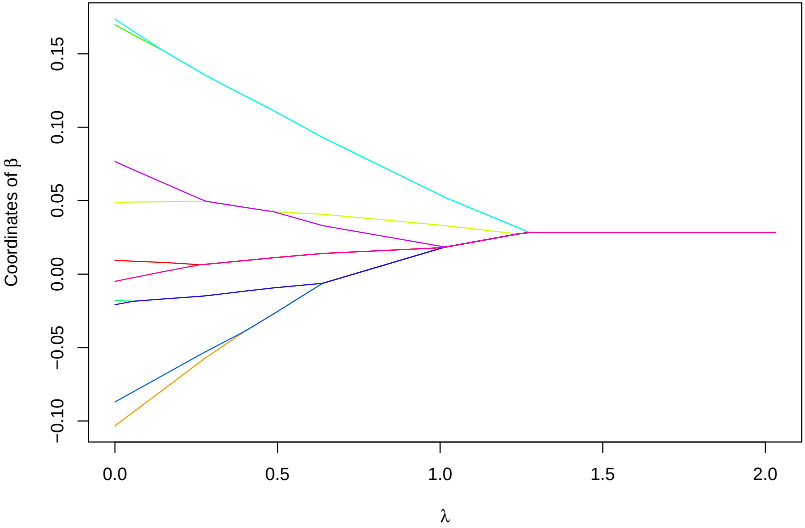
$$\lambda \sum_{(i,k)\in G} w_{i,k} |\beta_i - \beta_k|.$$

- ...

Fused Lasso



Generalized Fused Lasso



Interludium

- Regularization is very popular in machine learning (big data!) and statistics literature BUT only does regularization with one type of risk factor at a time.
- Efficient algorithms and R packages are available in the Gaussian case and for 'one type/penalty'
  - glmnet (Simon): Lasso, ridge en elastic net for GLMs.
  - genlasso (Arnold): 1D and 2D Fused Lasso, signal approximation, trend filtering for Gaussian case.

Need for:

- Extension of literature and algorithms to GLMs.
- Simultaneously work with risk factors of different types.

A unified framework??

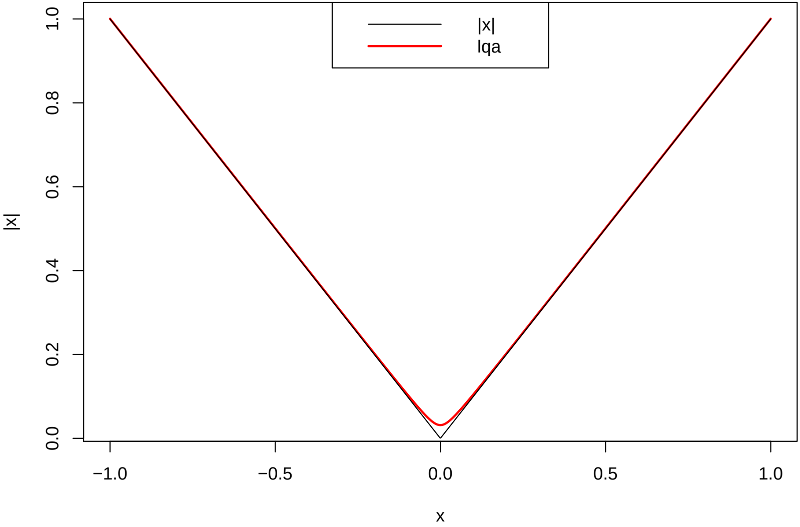
Gertheiss - Tutz - Oelker (2010-2016)
'Sparse modeling of categorial explanatory variables' - Annals of Applied Statistics

- GLM implementation.
- Many different penalties.
- R package available: `gvcm.cat` (not maintained).

But...
Fitting algorithm: 'local quadratic approximation' and subsequent quadratic programming:

- Only approximate clustering.
- How to choose approximation accuracy? Cluster accuracy?
- Computationally intensive.

Local quadratic approximation



A unified framework!!

For J risk factors, each with regularization term $P_j()$, we want to optimize:

$$-\ell\left(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_J\right) + \sum_{j=1}^{J} P_j\left(\boldsymbol{\beta_j}\right),$$

For this we use the theory of proximal operators (PO):

$$\text{Prox}_P(\boldsymbol{v}) = \underset{\boldsymbol{z}}{\text{argmin}}\left(P(\boldsymbol{z}) + \frac{1}{2}||\boldsymbol{z} - \boldsymbol{v}||_2^2\right).$$

Interpretation:

- POs are (generalized) projections. From a starting point $\boldsymbol{v}$, the PO will project this $\boldsymbol{v}$ to the closest point in the constraint associated with penalty $P()$ (remember the diamond surface for LASSO).

Optimization algorithm using proximal operators

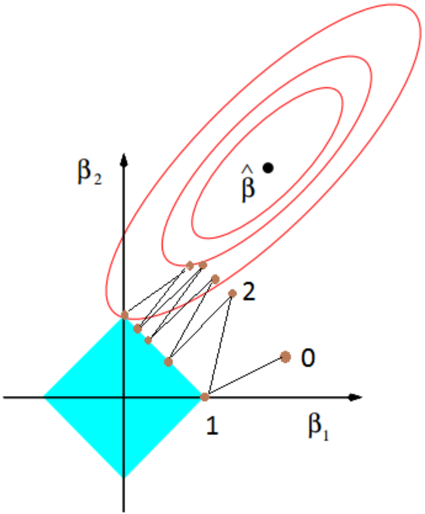
Efficient algorithm to optimize

$$-\ell\left(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_J\right) + \sum_{j=1}^{J} P_j\left(\boldsymbol{\beta}_{\boldsymbol{j}}\right).$$

1. Choose a (good) starting value.
2. Ignore penalties $P_j()$ and move in the direction of optimal point for $\ell()$.
3. Project new point onto the constraint set ($=$ calculate the PO of this new point).
4. Repeat until convergence.

Step 3 is 'easy', because projection splits into projecting the separate components $\boldsymbol{\beta}_j$.
This makes our algorithm efficient and scalable!

Proximal operator as projections



Practical example: MTPL data

Motor third party liability dataset (163 234 observations):

- response is *number of claims*;
- ordered predictors *age*, *bonus malus scale*, *power of car*;
- nominal predictors *type of coverage*, *type of fuel*;
- total of 281 parameters.

Fit GLM with Poisson assumption with weighted regularization terms.

Practical example: MTPL data

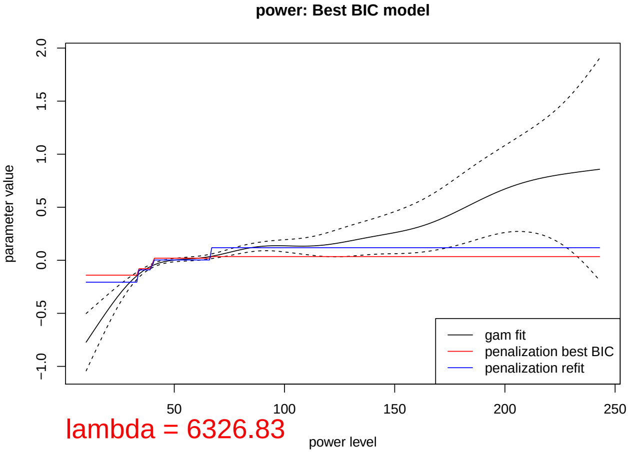


Figure: Comparison of parameter estimates for predictor *power*. GAM fit, penalized fit and re-estimated penalized fit for MTPL dataset. Penalties were weighted using GAM-based weights.

Practical example: MTPL data

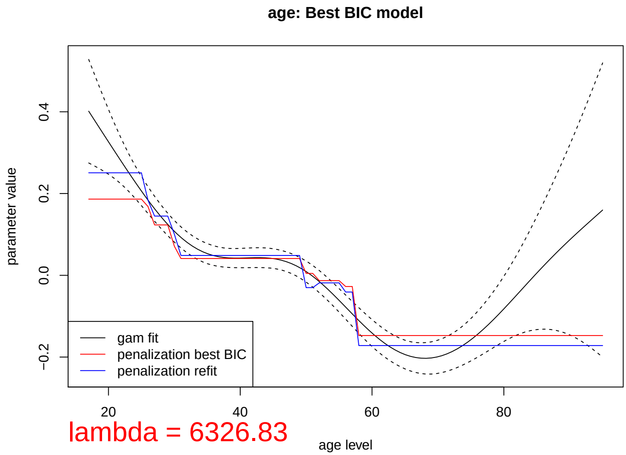
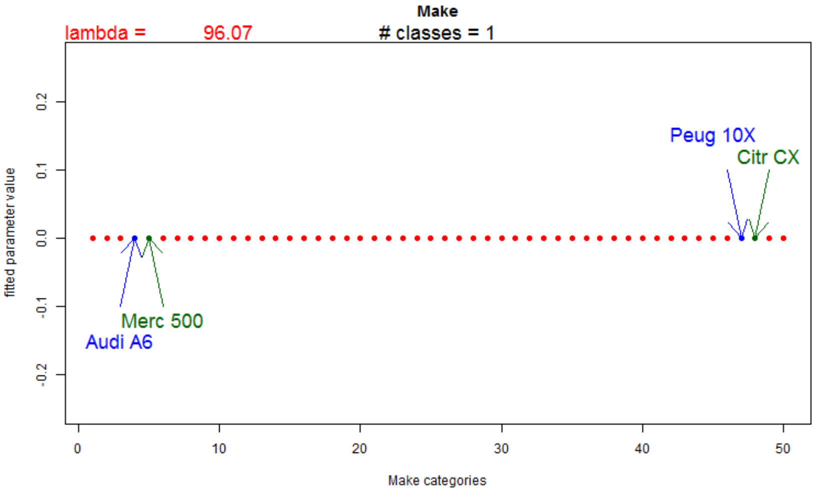


Figure: Comparison of parameter estimates for predictor *age* between GAM fit, penalized fit and re-estimated penalized fit for MTPL dataset. Penalties were weighted using GAM-based weights.

Practical example: MTPL data



Conclusion: our contribution

- Applying machine learning techniques to a classical statistical problem.
- Implementing an efficient algorithm which is scalable and interpretable.
- Flexibility of regularization takes into account type/structure of risk factor.
- Works for all popular penalties.
- Makes use of available penalty-specific literature.

Conclusion: further research

- Further improving algorithm efficiency.
- Implementing new penalties for spatial information, interaction effects...
- R package building in progress.
- Write a paper!

References I

[1] R. Tibshirani (1996)
Regression shrinkage and selection via the lasso.
Journal of the Royal Statistical Society, Series B, vol. 58, no. 1, pp. 267-288.

[2] R. Tibshirani, M. Saunders, J. Zhu, K. Knight (2005)
Sparsity and smoothness via the fused lasso.
Journal of the Royal Statistical Society, Series B, vol. 67, no. 1, pp. 91-108.

[3] J. Lee, B. Recht, R. Salakhutdinov, N. Srebro, J. Tropp (2010),
Practical largescale optimization for max-norm regularization.
Advances in Neural Information Processing Systems, vol. 23, pp. 1297-1305.

References II

[4] K. Toh, S. Yun (2010),
An accelerated proximal gradient algorithm for nuclear norm regularized least squares problems.
Pacific Journal of Optimization, vol. 6, pp. 615-640.

[5] J. Gertheiss, G. Tutz (2010)
Sparse modeling of categorial explanatory variables.
The Annals of Applied Statistics, vol. 4, no. 4, pp. 2150-2180.

[6] N. Parikh, S. Boyd, (2014),
Proximal Algorithms.
Foundations and Trends in Optimization, 1(3), pp. 123-231.

References III

[7] T. Arnold, R. Tibshirani (2016),
Efficient Implementations of the Generalized Lasso Dual Path Algorithm.
Journal of Computational and Graphical Statistics, vol. 25, no.1 pp. 1-27.

[8] Y. Nesterov (2007),
Gradient methods for minimizing composite objective function.
CORE Discussion paper, Catholic University of Louvain, 2007/76.