Group Elastic Net Regression: A Framework for High-Dimensional Regression and Classification of Psychological Data

Nathaniel E. Helwig

Associate Professor of Psychology and Statistics
Core Member of the Data Science Initiative
University of Minnesota

International Meeting of the Psychometric Society
University of Minnesota, Twin Cities
July 16, 2025

Table of Contents

Table of Contents

Multiple and Generalized Nonparametric Regression

Suppose that Y follows an exponential family distribution.

Given predictors $(X_1, \ldots, X_p)$, consider a MGNR model of the form

$$g(\mu) = f(X_1, \ldots, X_p)$$

where

- μ is the (conditional) expectation of Y (given $X_1, \ldots, X_p$)
- $g(\cdot)$ is a user-specified link function (monotonic and invertible)
- $f(\cdot)$ is an unknown "smooth" function of $q \leq p$ predictors

Goal: estimate the unknown function $f(\cdot)$ from training data.

Challenge: Model Structure is Completely Unknown

Variable Selection Problem: which predictors should be included?

- Not all of the candidate predictors may be needed
- Want to determine which predictors are relevant (active)

Term Selection Problem: how should variables be included?

- Do the predictors combine in an additive fashion? Or interactive?
- Want to learn which main and/or interaction effects matter

Smoothing Problem: how smooth should each term be?

- Predictors may (and likely will) enter model in nonlinear fashion
- Want to avoid under-fitting (bias) and over-fitting (variability)

Three-Pronged Approach

Step 1: The Theory (Tensor Product Reproducing Kernels)

- Existing tensor product RKHS theory is not ideal for selection
- Develop improved representer theorem for tensor products

Step 2: The Algorithm (Adaptively Bounded Gradient Descent)

- Existing algorithms are not easily adaptable (e.g., to different g)
- Develop versatile group elastic net algorithm for GLMs/GAMs

Step 3: The Practice (Smoothing and Subspace Selection)

- Existing implementations only consider additive models
- Develop intuitive software for selecting main and interaction effects

Group Elastic Net Regularized GLM/GAM

I consider a groupwise extension of the classic elastic net estimator:

$$\hat{\boldsymbol{\beta}}_{\lambda,\alpha} = \underset{\boldsymbol{\beta} \in \mathbb{R}^p}{\arg\min} \left\{ \frac{1}{n} L(\boldsymbol{\beta}|\mathbf{D}) + \lambda P_\alpha(\boldsymbol{\beta}) \right\} \quad (1)$$

where the group elastic net penalty function has the form

$$P_\alpha(\boldsymbol{\beta}) = \sum_{k=1}^{K} \omega_k \left(\alpha \|\boldsymbol{\beta}_k\| + \frac{1}{2}(1-\alpha)\|\boldsymbol{\beta}_k\|^2 \right) \quad (2)$$

which is a natural groupwise extension of the elastic net penalty.

- Helwig (2024) Stats paper for smoothing theory
- Helwig (2025) JCGS paper for computational theory
- Helwig (2025) **grpnet** R package for application

Table of Contents

ABGD Algorithm for GLMs with Group Elastic Net

Algorithm 1 Adaptively bounded gradient descent algorithm for generalized linear models with group elastic net penalties.

Calculate $\gamma_k = \text{maxeig}\left(\frac{1}{n}\sum_{i=1}^n w_i \mathbf{x}_{ik}\mathbf{x}_{ik}^\top\right)$ for $k = 1, \ldots, K$
Initialize $\boldsymbol{\beta}_k = \mathbf{0}_{p_k}$ for all $k = 1, \ldots, K$
while $\max_j \frac{|\beta_j - \beta_{j(\text{old})}|}{1+|\beta_j|} > \epsilon$ **do**
    $\boldsymbol{\beta}_{\text{old}} \leftarrow \boldsymbol{\beta}$
    $\delta = \min_i(d_i^2 v_i) = \max_i(1/d_i^2 v_i)$
    for $k = 1, \ldots, K$ **do**
        $\mathbf{g}_k = \frac{1}{n}\sum_{i=1}^n \frac{w_i}{d_i v_i}(y_i - \mu_i)\mathbf{x}_{ik}$
        $\mathbf{b}_k = \boldsymbol{\beta}_k + (\gamma_k \delta)^{-1}\mathbf{g}_k$
        $\boldsymbol{\beta}_k = \frac{1}{1+(1-\alpha)\lambda\omega_k(\gamma_k\delta)^{-1}}\left(1 - \frac{\alpha\lambda\omega_k(\gamma_k\delta)^{-1}}{\|\mathbf{b}_k\|}\right)_+ \mathbf{b}_k$
    end for
end while

Scope of Packages: Loss Functions

Table 1: Loss functions available in each package.

$L(\boldsymbol{\beta}\|\mathbf{D})$	**gglasso**	**grplasso**	**grpnet**	**grpreg**
Gaussian	✓	✓	✓	✓
Multivariate Gaussian			✓	
Binomial / Logistic	✓	✓	✓	✓
Multinomial			✓	
Poisson		✓	✓	✓
Negative Binomial			✓	
Gamma			✓	
Inverse Gaussian			✓	
Squared Hinge	✓		✓	
Smoothed Hinge	✓		✓	
Cox PH				✓

Scope of Packages: Usability Features

Table 2: Usability features of each package.

Features	gglasso	grplasso	grpnet	grpreg
α and γ tuning			✓	
Parallelization			✓	
Default method	✓		✓	✓
Formula method		✓	✓	
Orthogonalization		✓	✓	✓
Standardization			✓	
Additive splines			✓	✓
Tensor product splines			✓	
MCP and SCAD			✓	✓

Table of Contents

Overview of Student Performance Dataset

Student Performance dataset from UCI Machine Learning Repository

- Published by Cortez (2008) from Cortez & Silva (2008)
- Obligatory UCI citation (Kelly et al., 2025)

Focused on the math performance dataset

- $n = 395$ Portuguese secondary school students
- $m = 3$ math exam scores (G1, G2, G3)
- $p = 30$ sociodemograhic features

Goal: predict the exams scores from the sociodemographic features

Response and Predictor Variables

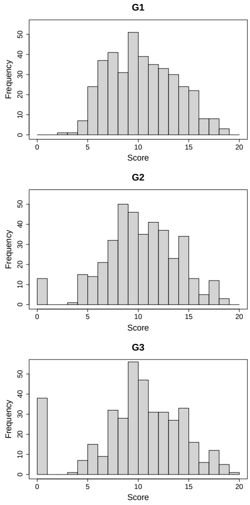

Figure 1: Response variables.

Table 3: Predictor variables.

Variable	R Class	Spline Type	Range/Levels
school	character	Nominal	GP = Gabriel Pereira, MS = Mousinho da Silveira
sex	character	Nominal	F = female, M = male
age	integer	Cubic	[15, 22]
address	integer	Nominal	U = urban, R = rural
famsize	character	Nominal	LE3 = 3 or less, GT3 = 4 or more
Pstatus	character	Nominal	T = together, A = apart
Medu	integer	Ordinal	0, 1, 2, 3, 4
Fedu	integer	Ordinal	0, 1, 2, 3, 4
Mjob	character	Nominal	teacher, health, services, home, other
Fjob	character	Nominal	teacher, health, services, home, other
reason	character	Nominal	home, reputation, course, other
guardian	character	Nominal	mother, father, other
traveltime	integer	Ordinal	1 = <15m, 2 = 15-30m, 3 = 30-60m, 4 = >60m
studytime	integer	Ordinal	1 = <2hr, 2 = 2-5hr, 3 = 5-10hr, 4 = >10hr
failures	integer	Ordinal	0, 1, 2, 3
schoolsup	character	Nominal	no, yes
famsup	character	Nominal	no, yes
paid	character	Nominal	no, yes
activities	character	Nominal	no, yes
nursery	character	Nominal	no, yes
higher	character	Nominal	no, yes
internet	character	Nominal	no, yes
romantic	character	Nominal	no, yes
famrel	integer	Ordinal	1 = very bad, . . . , 5 = very good
freetime	integer	Ordinal	1 = very low, . . . , 5 = very high
goout	integer	Ordinal	1 = very low, . . . , 5 = very high
Dalc	integer	Ordinal	1 = very low, . . . , 5 = very high
Walc	integer	Ordinal	1 = very low, . . . , 5 = very high
health	integer	Ordinal	1 = very bad, . . . , 5 = very good
absences	integer	Cubic	[0, 75]

Note. Medu and Fedu: 0 = none, 1 = primary, 2 = 5th-9th, 3 = secondary, 4 = higher

Overview of Data Splitting and Analysis

Multivariate regression predicting exams (G1, G2, G3) from features

- Consider all main effects (30) and two-way interactions (435)
- Use tensor product (spectral) smoothing splines (Helwig, 2024)
- Note: design matrix has 2837 columns after expansion

Use 80/20 (training/testing) data splitting procedure:

- Fit/tune model on training data; evaluate on testing data
- Repeat process 10 times to investigate stability of solution

For each sample of training data:

- 10-fold CV to tune λ and α (using `lambda.1se` for prediction)
- Compare LASSO, MCP, and SCAD

Results: Prediction Accuracy

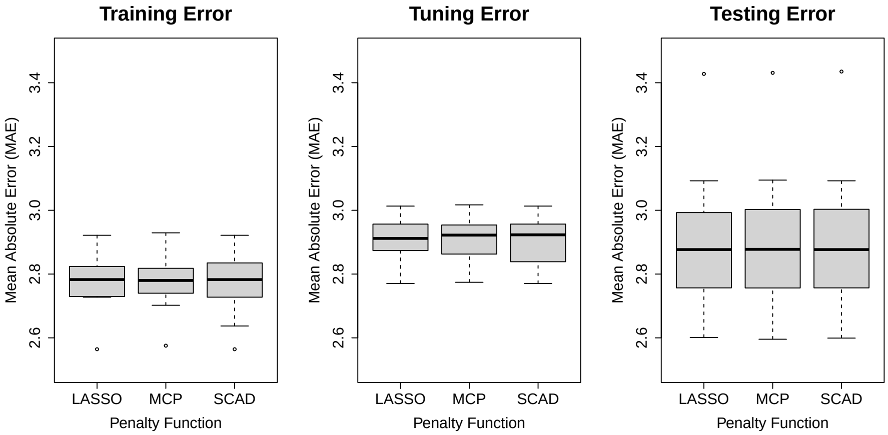

Figure 2: Prediction accuracy.

Results: Variable Selection

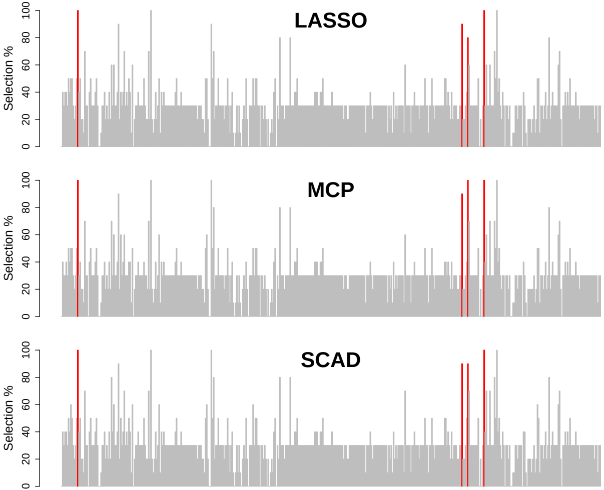

Figure 3: Variable selection.

Results: Variable Importance

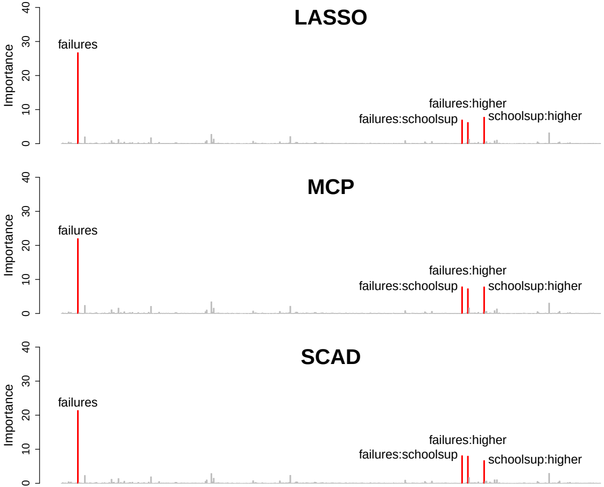

Figure 4: Variable importance.

Results: Model Predictions

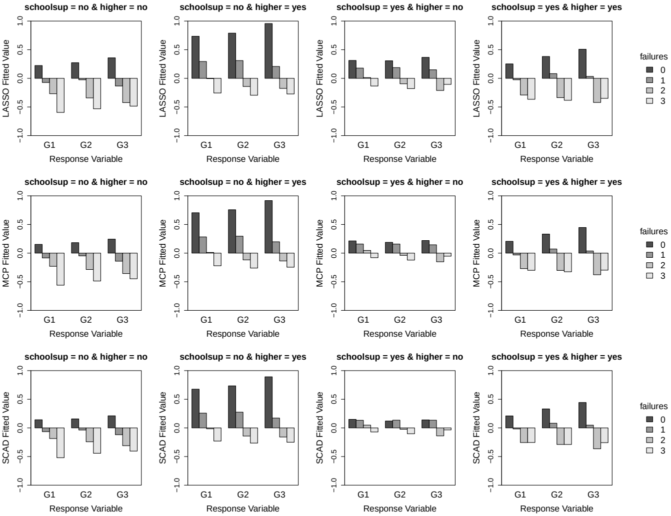

Figure 5: Model predictions.

Table of Contents

Take-Home Points and Extensions

Important main and two-way interaction effects for math performance

- `failures`, `failures:schoolsup`, `failures:higher`, `schoolsup:higher`
- Many small effects are reliably included in prediction rule too

GRPNET provides interpretable machine learning for HD regression

- Finds main and interaction effects without prior assumptions
- Blends strengths of parametric and nonparametric regression

Extensions for ordinal and multivariate outcomes:

- Ordinal outcomes are coming soon (using group penalized POLR)
- Multivariate outcomes with user-specified components is TBD…

References and Funding

References:

Cortez, P. (2008). *Student Performance.* UCI Machine Learning Repository. (DOI: https://doi.org/10.24432/C5TG7T)

Cortez, P., & Silva, A. (2008). Using data mining to predict secondary school student performance. In A. Brito & J. Teixeira (Eds.), *Proceedings of 5th FUture BUsiness TEChnology conference (FUBUTEC 2008)* (p. 5-12). Porto, Portugal: EUROSIS.

Helwig, N. E. (2024). Precise tensor product smoothing via spectral splines. *Stats*, *7*(1), 34-53. doi: 10.3390/stats7010003

Helwig, N. E. (2025). grpnet: Group elastic net regularized GLMs and GAMs [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=grpnet` (R package version 1.0)

Helwig, N. E. (2025). Versatile descent algorithms for group regularization and variable selection in generalized linear models. *Journal of Computational and Graphical Statistics*, *34*(1), 239-252. doi: 10.1080/10618600.2024.2362232

Kelly, M., Longjohn, R., & Nottingham, K. (2025). *The UCI Machine Learning Repository.* `https://archive.ics.uci.edu`.

Funded by a sabbatical award from Minnesota and NIH grants:
R01EY030890, R01MH115046, U01DA046413