

Prism: Multiple spline regression with regularization, dimensionality reduction, and feature selection

Christopher R Madan

Boston College

27 June 2016

Paper DOI: <http://dx.doi.org/10.21105/joss.00031>

Software Repository: <https://github.com/cMadan/prism>

Software Archive: <http://dx.doi.org/10.5281/zenodo.56821>

Summary

Prism uses a combination of statistical methods to conduct spline-based multiple regression. Prism conducts this regression using regularization, dimensionality reduction, and feature selection, through a combination of smoothing spline regression, PCA, and RVR/LASSO. Smoothing splines can be used to model non-parametric relationships using piece-wise cubic functions (Wahba and Wold 1975; Fox 2000). Relevance vector regression (RVR) refers to application of a relevance vector machine (RVM) to a regression problem; broadly, RVM is similar to multiple linear regression with regularization, using automatic relevance determination for feature selection (Tipping 2000). RVM shares many commonalities with SVM, and is implemented as a special case of a Sparse Bayesian framework (Tipping 2001; Tipping and Faul 2003).

Prism has been tested in MATLAB 2015b and requires three first-party toolboxes: (1) Curve Fitting Toolbox; (2) Statistics and Machine Learning Toolbox; (3) Signal Processing Toolbox. Relevance vector regression requires the SparseBayes V2 toolbox, which can be obtained from <http://www.relevancevector.com>.

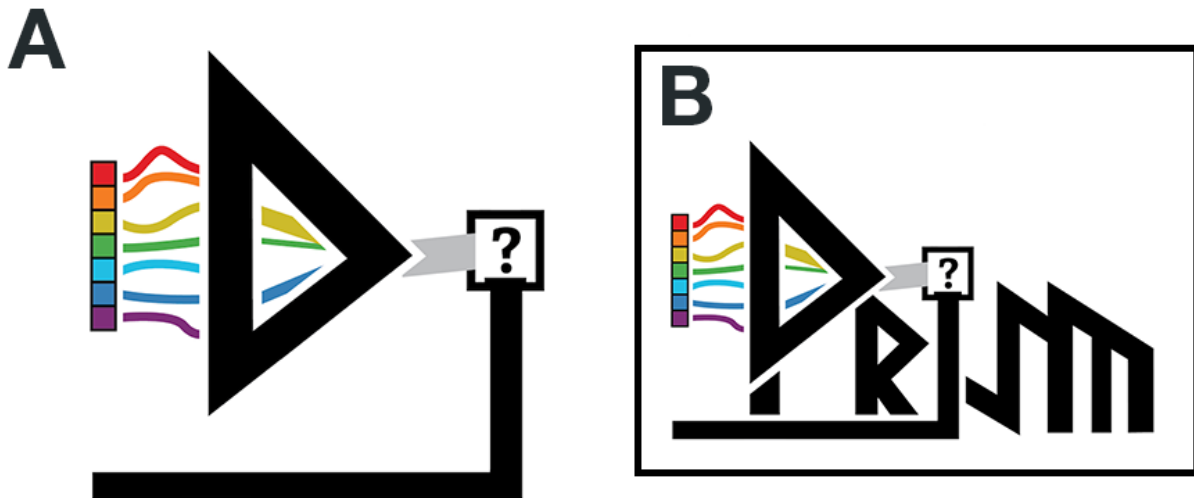


Figure 1: Illustration of Prism regression procedure, first conducting spline regression for each predictor,

followed by dimensionality reduction and feature selection (panel A). The logo for Prism is shown in panel B.

Prism was implemented for conducting multiple regression investigating age-related differences in brain morphology. While linear and quadratic regression are often used (e.g., Sowell et al. (2003); Hogstrom et al. (2013); Madan and Kensinger (2016)), it has been shown that non-linear (spline) regression is more appropriate (Fjell et al. 2010; Fjell et al. 2013).

Example implementation

To provide an example of Prism’s implementation, see the ‘demo’ folder. `demo.mat` was generated using brain morphology data, where each column corresponds to a brain morphology measure. The test and training data are derived from independent MR sites, and the dependent variable is age. The data is from an unpublished study and the pre-processing pipeline is not finalized (e.g., correcting for site effects), so further information of the dataset is not provided at this time.

Figure 2: Output scatter plots for code in `demo.m`.

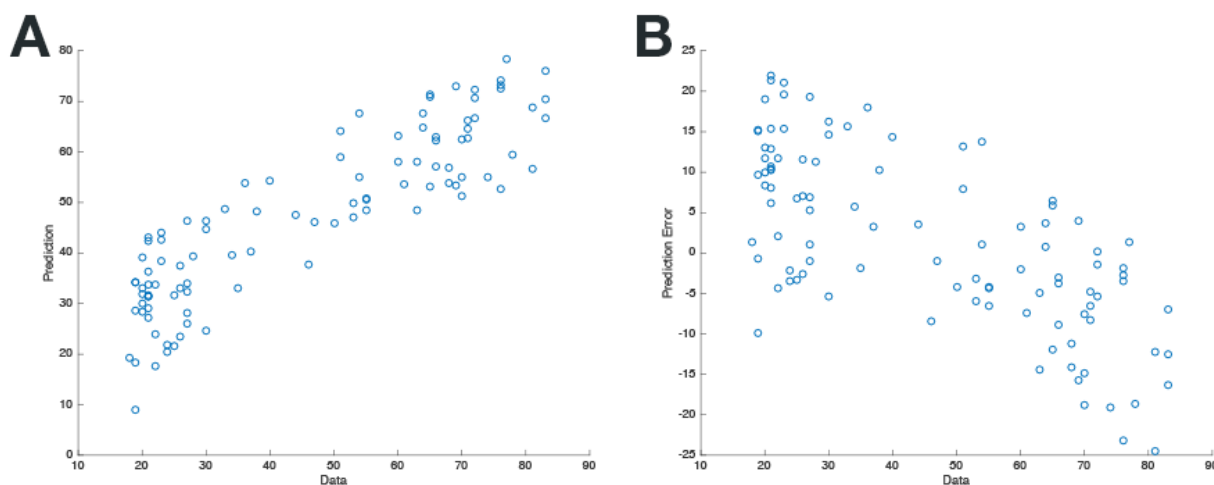


Figure 1: Example output.

See README.md for further details.

References

- Fjell, A M, K B Walhovd, L T Westlye, Y Østby, C K Tamnes, T L Jernigan, A Gamst, and A M Dale. 2010. “When Does Brain Aging Accelerate? Dangers of Quadratic Fits in Cross-Sectional Studies.” *NeuroImage* 50: 1376–83. doi:10.1016/j.neuroimage.2010.01.061.
- Fjell, A M, L T Westlye, H Grydeland, I Amlien, T Espeseth, I Reinvang, N Raz, D Holland, A M.Dale, and K B Walhovd. 2013. “Critical Ages in the Life Course of the Adult Brain: Nonlinear Subcortical Aging.” *Neurobiology of Aging* 34: 2239–47. doi:10.1016/j.neurobiolaging.2013.04.006.
- Fox, J. 2000. *Nonparametric Simple Regression: Smoothing Scatterplots*. Sage.
- Hogstrom, L J, L T Westlye, K B Walhovd, and A M Fjell. 2013. “The Structure of the Cerebral Cortex Across Adult Life: Age-Related Patterns of Surface Area, Thickness, and Gyrification.” *Cerebral Cortex* 23:

2521–30. doi:10.1093/cercor/bhs231.

Madan, C R, and E A Kensinger. 2016. “Cortical Complexity as a Measure of Age-Related Brain Atrophy.” *NeuroImage* 134: 617–29. doi:10.1016/j.neuroimage.2016.04.029.

Sowell, E R, B S Peterson, P M Thompson, S E Welcome, A L Henkenius, and A W Toga. 2003. “Mapping Cortical Change Across the Human Life Span.” *Nature Neuroscience* 6: 309–15. doi:10.1038/nn1008.

Tipping, M E. 2000. “The Relevance Vector Machine.” *Advances in Neural Information Processing Systems (NIPS)*, 652–58.

———. 2001. “Sparse Bayesian Learning and the Relevance Vector Machine.” *Journal of Machine Learning Research* 1: 211–44.

Tipping, M E, and A C Faul. 2003. “Fast Marginal Likelihood Maximisation for Sparse Bayesian Models.” *Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics (AISTATS)*.

Wahba, G, and S Wold. 1975. “A Completely Automatic French Curve: Fitting Spline Functions by Cross Validation.” *Communications in Statistics* 4: 1–17. doi:10.1080/03610927508827223.