Superagers penalised regression with elastic net

Nathan Green

11 March, 2021

Introduction

Rish et al. (2012) Pornpattananangkul (2020) Cui and Gong (2018) Carroll et al. (2009)

Data

Methods

Let $X_1, \dots, X_N$ be a set of N predictors and let Y be the response variable. We consider the problem of estimating the coefficients β_i in the following linear regression model

$$\hat{y} = x_1\beta_1 + \cdots x_N\beta_N = \mathbf{X}\beta,$$

where $\hat{y}$ is an approximation of y. The Ordinary Least Squares (OLS) regression finds a set of β_i that minimize the sum-squared approximation error $(y - x\beta)^2$.

In general, OLS solutions are often unsatisfactory, since there is no unique solution when $p \gg n$ and it is difficult to pinpoint which predictors are most relevant to the response. Various regularization approaches have been proposed in order to handle large-p, small-n datasets, and to avoid the overfitting. Particularly, recently proposed sparse regularization methods such as Lasso, ridge regression and Elastic Net. Lasso and EN address both of the OLS shortcomings, since variable selection is embedded into their model-fitting process. Sparse regularization methods include the l1-norm regularization on the coefficients, which is known to produce sparse solutions, i.e. solutions with many zeros, thus eliminating predictors that are not essential. In this paper, we use the Elastic Net (EN) regression that finds an optimal solution to the OLS problem objective, augmented with additional regularization terms that include the sparsity-enforcing.

l1-norm constraint on the regression coefficients that "shrinks" some coefficients to zero, and a "grouping" l2-norm constraint that enforces similar coefficients on predictors that are highly correlated with each other which l1-constraint alone do not provide. Formally, EN regression optimizes the following function

$$L(\lambda_1, \lambda_2; \beta) = (y - x\beta)^2 + \lambda_1 \|\beta\|_1 + \lambda_2 \|\beta\|_2$$

For each network i, let Y be a binary outcome of either superager or control and $\mathbf{X}$ consist of 832 covariate measurements. This is modelled as

$$logit(p^i) = \mathbf{X}^i \beta^i, \quad i = 1, 2, \dots, 11$$

We can then obtain the odds-ratios using the fitted models to give an average comparison between individuals with or without a unit increase in a particular covariate j

$$OR_j/OR = \frac{p_j/(1-p_j)}{p/(1-p)} = \exp(\beta_j)$$

Results

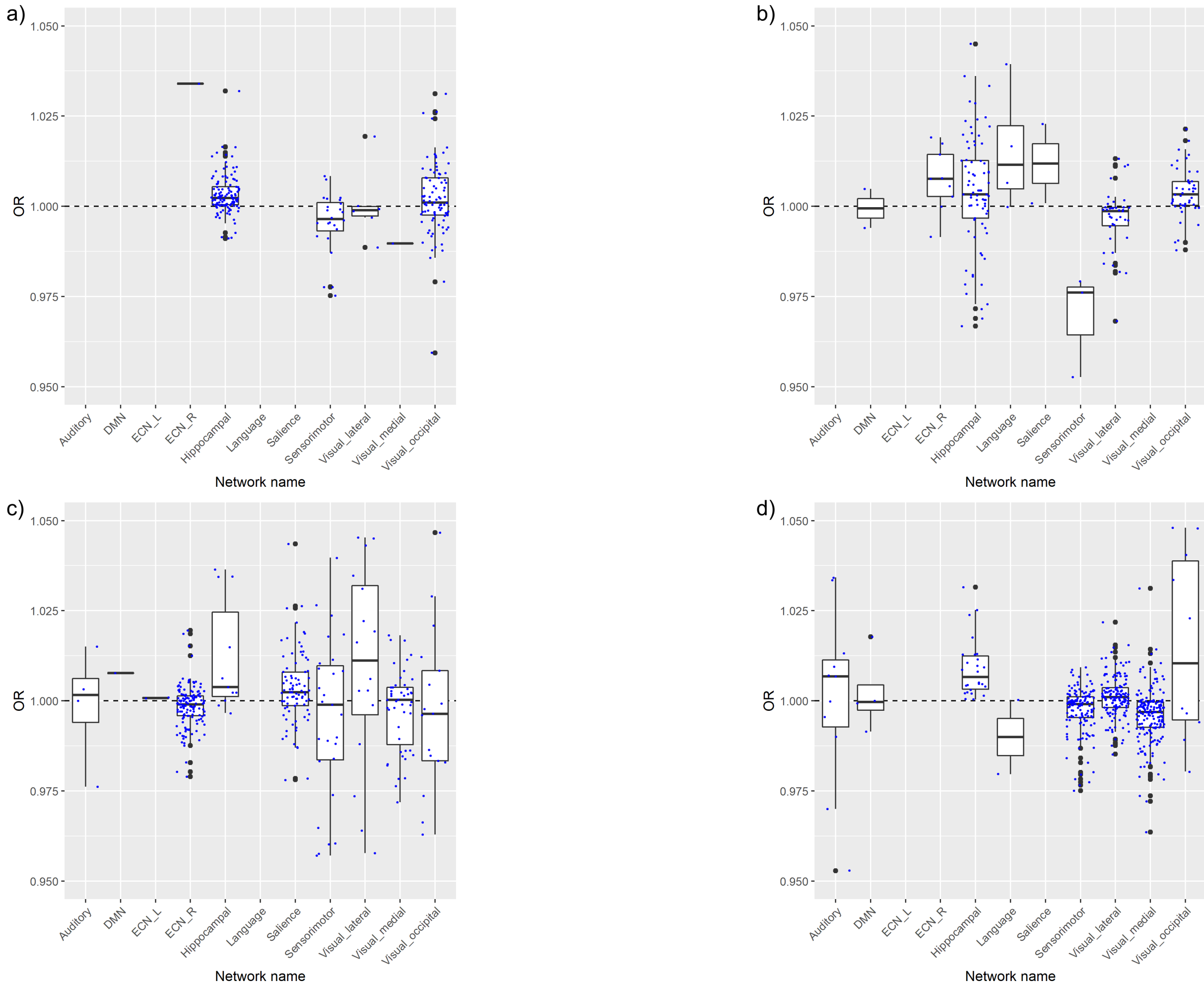

Figure 1: Box plots showing estimated coefficient values using an elastic net model fitted to different data sets and stratified by brain network. Blue points are the individual values and the black box and points show median, 25th and 75th percentile and whiskers to largest and smallest values. a) 3T b) 3T cross-over cohort c) 7T d) 3T quality cohort.

Tables

Table 1: Summary table for 3T.

X	network_name	OR	L95	U95	min	max	count	prop
1	ECN_R	1.034	1.034	1.034	1.034	1.034	1	1
2	Hippocampal	1.003	0.993	1.015	0.991	1.032	118	1
3	Sensorimotor	0.995	0.977	1.008	0.975	1.008	24	1
4	Visual_lateral	1.000	0.990	1.017	0.989	1.019	6	1
5	Visual_medial	0.990	0.990	0.990	0.990	0.990	1	1
6	Visual_occipital	1.002	0.986	1.026	0.959	1.031	89	1

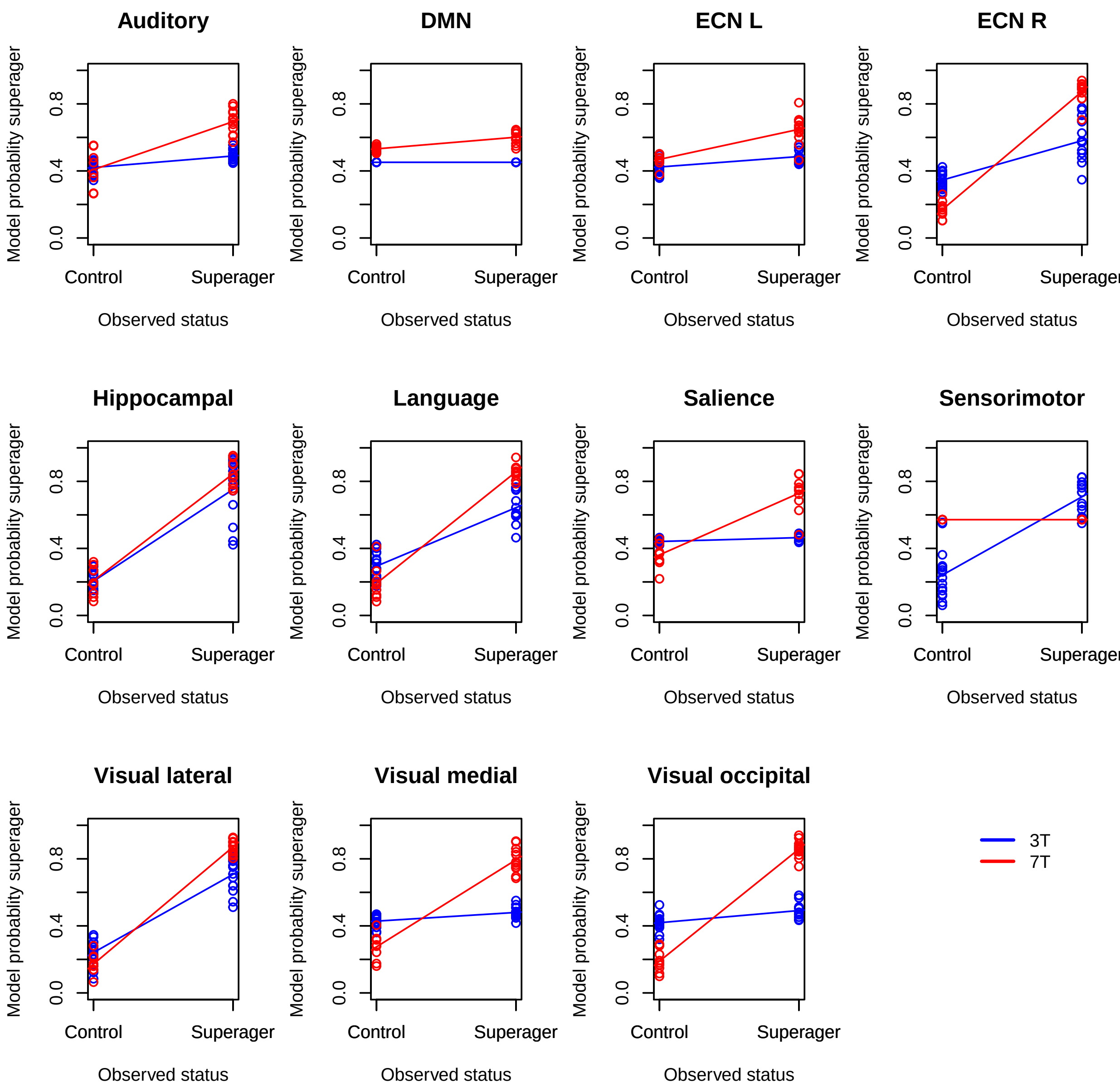

Figure 2: Scatterplot of model fits vs observed data and fitted regression for 3T and 7T datasets.

Table 2: Summary table for 7T.

X	network_name	OR	L95	U95	min	max	count	prop
1	Auditory	0.999	0.978	1.014	0.976	1.015	4	1
2	DMN	1.008	1.008	1.008	1.008	1.008	1	1
3	ECN_L	1.044	1.001	1.115	1.001	1.119	4	1
4	ECN_R	0.999	0.988	1.010	0.979	1.020	137	1
5	Hippocampal	1.016	0.953	1.089	0.932	1.102	14	1
6	Salience	1.004	0.987	1.026	0.978	1.086	86	1
7	Sensorimotor	0.988	0.936	1.030	0.908	1.040	31	1
8	Visual_lateral	1.009	0.960	1.045	0.958	1.045	16	1
9	Visual_medial	0.998	0.977	1.017	0.972	1.018	47	1
10	Visual_occipital	1.000	0.933	1.074	0.906	1.092	20	1

Performance

How well do the model predict the data they were fit with?

The controls-superager classifications for each model and network.

Model fit statistics for each network. RMSE = 0 and $R^2 = 1$ are best.

Table 3: Predictions tables for 3T and 7T.

Network	RMSE_3T	Rsquare_3T	RMSE_7T	Rsquare_7T
Auditory	0.57	0.18	0.31	0.67
DMN	0.67		0.65	
ECN_L	0.57	0.18	0.31	0.65
ECN_R	0.31	0.67	0.00	1
Hippocampal	0.25	0.77	0.00	1
Language	0.18	0.88	0.00	1
Salience	0.67		0.22	0.82
Sensorimotor	0.25	0.77	0.65	
Visual_lateral	0.00	1	0.00	1
Visual_medial	0.57	0.18	0.00	1
Visual_occipital	0.54	0.19	0.00	1

Table 4: Predictions tables for 3T quality and cross-over.

Network	RMSE_3Tqual	Rsquare_3Tqual	RMSE_3Tsame	Rsquare_3Tsame
Auditory	0.21	0.84	0.00	1
DMN	0.36	0.55	0.65	
ECN_L	0.69		0.65	
ECN_R	0.00	1	0.31	0.65
Hippocampal	0.21	0.84	0.00	1
Language	0.36	0.58	0.00	1
Salience	0.69		0.62	0.07
Sensorimotor	0.00	1	0.38	0.53
Visual_lateral	0.00	1	0.00	1
Visual_medial	0.00	1	0.44	0.42
Visual_occipital	0.00	1	0.00	1

Conclusions

References

Carroll, Melissa K., Guillermo A. Cecchi, Irina Rish, Rahul Garg, and A. Ravishankar Rao. 2009. "Prediction and interpretation of distributed neural activity with sparse models." *NeuroImage* 44 (1): 112–22. https://doi.org/10.1016/j.neuroimage.2008.08.020.

Cui, Zaixu, and Gaolang Gong. 2018. "The effect of machine learning regression algorithms and sample size on individualized behavioral prediction with functional connectivity features." *NeuroImage* 178 (February): 622–37. https://doi.org/10.1016/j.neuroimage.2018.06.001.

Pornpattananangkul. 2020. "An Omics-Inspired Elastic Net Approach Drastically Improves Out-of-Sample Prediction and Regional Inference of Task-Based fMRI Narun Pornpattananangkul," 0–22.

Rish, Irina, Guillermo A. Cecchi, Kyle Heuton, Marwan N. Baliki, and A. Vania Apkarian. 2012. "Sparse regression analysis of task-relevant information distribution in the brain." *Medical Imaging 2012: Image Processing* 8314 (February 2012): 831412. https://doi.org/10.1117/12.911318.