Superager Elastic Net Regression: Methods

N Green

12/12/2020

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

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

Related work

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

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.