Bagging and Metropolis Elastic Net (BagMEN)

Alencar Xavier
Department of Agronomy, Purdue University
xaviera@purdue.edu

Statistical model

For the linear model

$$\mathbf{y} = \mu + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

The elastic net loss function (Friedman et al. 2010) is defined by

$$l_{EN} = \sum\varepsilon^2 + \lambda(\alpha\sum\beta^2 + 0.5 \times (1-\alpha)\sum|\beta|)$$

such that

$$EN(\alpha = 0) = \text{LASSO}$$

and

$$EN(\alpha = 1) = \text{Ridge regression.}$$

The regression coefficient is updated via coordinate descent, analogous to GSRU (Legarra and Misztal 2008), as

$$\beta^{t+1} = \delta\,\psi\,\phi$$

Sign δ can assume value {-1,1} and is defined by

$$\delta = \text{sign}\left(\frac{\mathbf{x}'\mathbf{y} + \beta^t\mathbf{x}'\mathbf{x}}{\mathbf{x}'\mathbf{x}}\right),$$

the coefficient ψ must be non-negative and is defined by

$$\psi = \text{abs}\left(\frac{\mathbf{x}'\mathbf{y} + \beta^t\mathbf{x}'\mathbf{x}}{\mathbf{x}'\mathbf{x}}\right) - \frac{\lambda(1-\alpha)}{2\,\mathbf{x}'\mathbf{x}}$$

with ridge shrinkage parameter

$$\phi = \left(\frac{\mathbf{x}'\mathbf{x}}{\mathbf{x}'\mathbf{x} + \lambda\alpha}\right)$$

The regularization parameters are

$$\lambda = \frac{\sigma_e^2}{\sigma_b^2}$$

and

$$\alpha = U[0,1] \text{ (Metropolis)}$$

Variance components can be sampled from inversed-chi squared distributions as

$$\sigma_e^2 = \chi^{-2}(S = \sum\varepsilon + S_e, \nu = n + \nu_e)$$

and

$$\sigma_b^2 = \chi^{-2}(S = \sum\beta + S_b, \nu = p + \nu_b)$$

with hyper-parameters

$$S_e = S_b = U[0,1]$$

and

$$\nu_e = \nu_b = \text{Poi}(\lambda = 4)$$

When the model is being fitted through bagging (Xavier et al. 2016), the loss function can be enhanced as

$$l_{EN} = \sum\varepsilon^2 + \sum\omega^2 + \lambda(\alpha\sum\beta^2 + 0.5 \times (1-\alpha)\sum|\beta|)$$

Where ω is the out-of-bag prediction error.

For the model fitting, the loss function is used for the Metropolis algorithm to find an ideal value of alpha. In this case, bagging EN is also beneficial by ensuring a robust choice of alpha for prediction purposes. For this framework, a new value of alpha is accepted with probability

$$\Pr(\alpha \to \alpha^*) = \frac{l_{EN}(\alpha)}{l_{EN}(\alpha) + l_{EN}(\alpha^*)}.$$

The only changes in the algorithm required for bagging EN are replacing x'x by γ(x'x) and n by γn, where gamma is the fraction of samples used to fit the model in each iteration.

References

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. Journal of statistical software, 33(1), 1.

Legarra, A., & Misztal, I. (2008). Technical note: Computing strategies in genome-wide selection. Journal of Dairy Science, 91(1), 360-366.

Xavier, A., Muir, W.M. & Rainey, K.M. (2016). Bagging Bayesian Learners in Genomic Prediction. Plant and Animal Genome XXIV Conference. https://pag.confex.com/pag/xxiv/webprogram/Paper18478.html.