`glmnet`'s Standardization

Peng Xu

August 28, 2020

The standardization process in `glmnet` is complicated and not well documented. To begin with, `glmnet` scales the elastic net loss function by a factor of $1/n$. Furthermore, for linear problems, `glmnet` implicitly "standardizes y to have unit variance before computing its λ sequence (and then unstandardizes the resulting coefficients)". In other words, `glmnet` is in fact optimizing the following problem for the linear elastic net (assuming $\boldsymbol{X}$ is already standardized):

$$\min_{\boldsymbol{\beta}} \frac{1}{2n} \sum_{j=1}^{n} \left(\frac{\boldsymbol{x}_j^\top \boldsymbol{\beta}}{\hat{\sigma}_y} - \frac{y_j}{\hat{\sigma}_y} \right)^2 + \frac{\lambda}{\hat{\sigma}_y} \alpha \|\boldsymbol{\beta}\|_1 + \frac{\lambda}{\hat{\sigma}_y^2} \frac{1-\alpha}{2} \|\boldsymbol{\beta}\|_2^2. \tag{1}$$

Further complications present when the option `standardization = T` is set, in which case `glmnet` will first standardize the data $\boldsymbol{X}$ using $\hat{\sigma}_{\boldsymbol{X}}$:

- If `intercept = F`, standardization is $\boldsymbol{X}^* = \operatorname{diag}[\hat{\sigma}_y \hat{\sigma}_{\boldsymbol{X}}]^{-1} \boldsymbol{X}$.
- If `intercept = T`, standardization is $\boldsymbol{X}^* = \operatorname{diag}[\hat{\sigma}_y \hat{\sigma}_{\boldsymbol{X}}]^{-1} (\boldsymbol{X} - \bar{\boldsymbol{X}} \mathbf{1}\mathbf{1}^\top)$.

Afterwards, the the coefficients are returned *unstandardized*, i.e. if $(\beta_0, \boldsymbol{\beta})$ are the original intercept and coefficients, `glmnet` reports

$$\boldsymbol{\beta}^* = \hat{\sigma}_y \operatorname{diag}[\hat{\sigma}_{\boldsymbol{X}}]^{-1} \boldsymbol{\beta}, \qquad \beta_0^* = \beta_0 - \bar{\boldsymbol{X}} \boldsymbol{\beta}^*.$$

For logistics and Poisson regression, the standardization procedure is basically the same, except `glmnet` no longer standardize by $\hat{\sigma}_y$, which make sense since $\boldsymbol{y}$ is now either categorical or count data.