P8160 - Less is More: Comparing Logistic and Lasso-Logistic Regression in Breast Cancer Diagnosis

Amy Pitts, Hun Lee, Jimmy Kelliher,
Tucker Morgan, and Waveley Qiu

2022-03-28

Full Model: Newton-Raphson

Consider the following log-likelihood, gradient, and hessian matrix. First, let

$$\pi_i = P(Y_i = 1|x_{i,1}, \ldots x_{i,p}) = \frac{e^{\beta_0 + \sum_{j=1}^{p} \beta_j x_{i,j}}}{1 + e^{\beta_0 + \sum_{j=1}^{p} \beta_j x_{i,j}}}.$$

The log-likelihood:

$$l(\mathbf{X}|\vec{\beta}) = \sum_{i=1}^{n} \left[y_i \left(\beta_0 + \sum_{j=1}^{p} \beta_j x_{i,j} \right) - \log \left(1 + \exp \left(\beta_0 + \sum_{j=1}^{p} \beta_j x_{i,j} \right) \right) \right]$$

The gradient:

$$\nabla l(\mathbf{X}|\vec{\beta}) = \left[\sum^n y_i - \pi_i \quad \sum^n x_{i,1}(y_i - \pi_i) \quad \ldots \quad \sum^n x_{i,p}(y_i - \pi_i) \right]^T_{1 \times (p+1)}$$

The hessian: produces a matrix $(p + 1 \times p + 1)$

$$\nabla^2 l(\mathbf{X}|\vec{\beta}) = -\sum_{i=1}^{n} \begin{pmatrix} 1 \\ X \end{pmatrix} \begin{pmatrix} 1 & X \end{pmatrix} \pi_i (1 - \pi_i)$$

Optimal Model: Logistic LASSO

For vector $\boldsymbol{\alpha} \in \mathbb{R}^{p+1}$, define $g : \mathbb{R}^{p+1} \to \mathbb{R}$ to be

$$g(\beta) \equiv -\frac{1}{2n} \sum_{i=1}^{n} w_i (z_i - \mathbf{X}_i^t \beta)^2 + O(\boldsymbol{\alpha}),$$

the Taylor expansion of our log-likelihood centered around $\boldsymbol{\alpha}$, where

$$z_i \equiv \mathbf{X}_i^t \boldsymbol{\alpha} + \frac{y_i - \pi_i}{w_i}, \qquad \text{(effective response)}$$

$$w_i \equiv \pi_i (1 - \pi_i), \text{ and} \qquad \text{(effective weights)}$$

$$\pi_i \equiv \frac{e^{\mathbf{X}_i^t \boldsymbol{\alpha}}}{1 + e^{\mathbf{X}_i^t \boldsymbol{\alpha}}}$$

for $i \in \{1, \ldots, n\}$.

Optimal Model: Logistic LASSO

It follows that for any $\lambda \in \mathbb{R}_+$,

$$\underset{\beta_k \in \mathbb{R}}{\arg\min} \left\{ g(\boldsymbol{\beta}) + \lambda \sum_{j=1}^{p} |\beta_j| \right\} = S\left(\hat{\beta}_k, \lambda_k\right), \text{ where}$$

$$\hat{\beta}_k \equiv \left(\sum_{i=1}^{n} w_i x_{ik}^2 \right)^{-1} \sum_{i=1}^{n} w_i x_{ik} \left(z_i - \sum_{j \neq k} \beta_j x_{ij} \right),$$

$$\lambda_k \equiv \left(\frac{1}{n} \sum_{i=1}^{n} w_i x_{ik}^2 \right)^{-1} \lambda,$$

and S is the soft-thresholding (or *shrinkage*) function. This is analogous to a penalized, weighted Gaussian regression.

Optimal Model: Logistic LASSO

Our coordinate descent algorithm proceeds as follows.

- Outer Loop: Decrement over $\lambda \in (\lambda_{\text{max}}, \dots, \lambda_{\text{min}})$
- Middle Loop: Update $\boldsymbol{\alpha} = \boldsymbol{\beta}$ and Taylor expand g around $\boldsymbol{\alpha}$.
- Inner Loop: Update $\beta_k = S\left(\hat{\beta}_k, \lambda_k\right)$ sequentially for $k \in \{0, 1, \dots, p, 0, 1, \dots, p, 0, 1, \dots\}$ until convergence.

Note: the middle loop terminates when a given Taylor expansion no longer yields updates (within the specified tolerance) to $\boldsymbol{\beta}$ in the inner loop.