Logistic Regression with a reject option

University of Seoul

Kim, Yoon-Hoe

Contents

Motivation

1

Motivation

- In some cases, classification problems may be **ambiguous or difficult to classify**.
- In this case, rather than making a wrong decision, it would be desirable to hold the decision and make a definite decision after it is supplemented.
- For example, an observation with a conditional probability of 1/2 diagnosed as cancer would be better to consider a better test method than an immediate decision.
- The **reject option** is an option that reports you that it is difficult to make a decision and also holds the decision.

Does 'Logistic Classifier with a reject option' perform better than general Logistic Classifier?

Method

- Algorithm
- Computation

2

Algorithm

Bradley – Terry Model – Bradley and Terry (1952)

$$\Pr(i \text{ beats } j) = \frac{\pi_i}{\pi_i + \pi_j} \qquad \Pr(j \text{ beats } i) = \frac{\pi_j}{\pi_i + \pi_j}$$

Bradley – Terry Model with ties option – Rao and Kupper (1967)

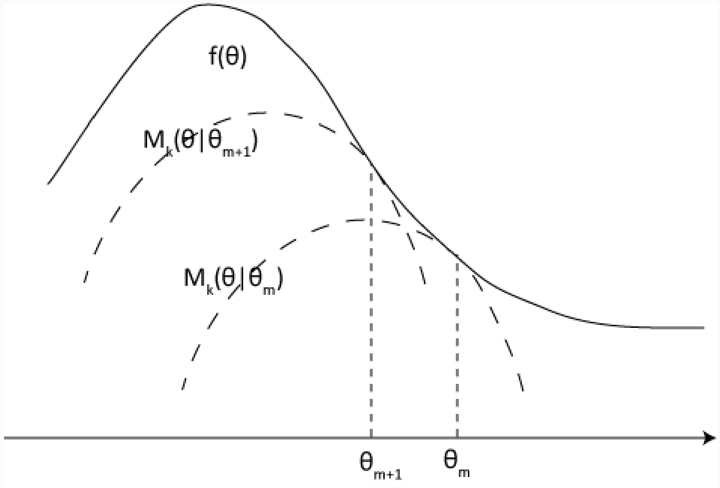
$$\Pr(i \text{ beats } j) = \frac{\pi_i}{\pi_i + \kappa\pi_j} \qquad \Pr(j \text{ beats } i) = \frac{\pi_j}{\kappa\pi_i + \pi_j}$$

$$\Pr(i \text{ ties } j) = \frac{(\kappa^2 - 1)\pi_i\pi_j}{(\pi_i + \kappa\pi_j)(\kappa\pi_i + \pi_j)}$$

Algorithm

MM algorithm for the Bradley – Terry Model – Hunter (2004)

- The MM algorithm works by finding a surrogate function that minorizes or majorizes the objective function.
- Optimizing the surrogate function will drive the objective function upward or downward until a local optimum is reached.



https://upload.wikimedia.org/wikipedia/commons/c/cc/Mmalgorithm.jpg

Algorithm

MM algorithm for the Bradley – Terry Model

The objective function of MM algorithm for the Bradley – Terry Model is loglikelihood.

$$\min_{\mathbf{p},\kappa} l(\mathbf{p},\kappa) = -\sum_{i<j}^{M}\left(r_{ij}\log\frac{\pi_i}{\pi_i+\kappa\pi_j} + r_{ji}\log\frac{\pi_j}{\kappa\pi_i+\pi_j} + r_{ij}^{ties}\log\frac{(\kappa^2-1)\pi_i\pi_j}{(\pi_i+\kappa\pi_j)(\kappa\pi_i+\pi_j)}\right) \quad (1)$$

Algorithm

Logistic Classification Problem with a reject option

$$p_{-1}(\mathbf{x}) = \Pr(Y = -1|\mathbf{x}) = \frac{1}{1 + \kappa\pi(\mathbf{x})} \qquad p_{+1}(\mathbf{x}) = \Pr(Y = +1|\mathbf{x}) = \frac{\pi(\mathbf{x})}{\kappa + \pi(\mathbf{x})}$$

$$p_0(\mathbf{x}) = \Pr(Y = 0|\mathbf{x}) = \frac{(\kappa^2 - 1)\pi(\mathbf{x})}{(1 + \kappa\pi(\mathbf{x}))(\kappa + \pi(\mathbf{x}))}$$

$f(\mathbf{x}) = \beta_0 + \mathbf{x}^T\beta$ and $\pi(\mathbf{x}) = \exp(f(\mathbf{x}))$

Algorithm

The corresponding likelihood is

$$L = \prod_{i=1}^{n} p_{-1}(\mathbf{x})^{I(y_i=-1)} p_{+1}(\mathbf{x})^{I(y_i=+1)} p_0(\mathbf{x})^{I(y_i=0)} \qquad (2)$$

Using censored framework which is similar to case of SVM with a reject option',
the corresponding likelihood can be expressed as follows. $d(0 \leq d < 1)$

$$L = \prod_{i=1}^{n} (p_{-1}(\mathbf{x})^{I(y_i=-1)} p_{+1}(\mathbf{x})^{I(y_i=+1)})^{1-d} (p_0(\mathbf{x})^{I(y_i=0)})^{d} \qquad (3)$$

$p_0(\mathbf{x}) > max(p_{-1}(\mathbf{x}), p_{+1}(\mathbf{x}))$: reject or postpone of the decision

$p_{-1}(\mathbf{x}) > p_{+1}(\mathbf{x})$: allocate class '−1'

$p_{-1}(\mathbf{x}) < p_{+1}(\mathbf{x})$: allocate class '1'

Algorithm

Let the negative loglikelihood of Eq. (3) be,

$$
\begin{aligned}
-l(\beta_+, \kappa) = (1-d)\sum_{i=1}^{n} \big(I(y_i = -1)[-\log(1+\kappa\pi(\mathbf{x}_i))] + I(y_i = +1)[f(\mathbf{x}_i) - \log(\kappa + \pi(\mathbf{x}_i))]\big) \\
+ dI(y_i = 0)\left(\log(\kappa^2 - 1) + f(\mathbf{x}_i) - \log(1+\kappa\pi(\mathbf{x}_i)) = \log(\kappa + \pi(\mathbf{x}_i))\right)
\end{aligned}
$$

let $\beta_+ = (\beta_0, \beta^T)^T$

For some $\lambda > 0$, then, the regularization is

$$\max_{\kappa > 1} \min_{\beta_+} l(\beta_+, \kappa) + \lambda \sum_{j=1}^{p} |\beta_j|$$

Algorithm

Summarized algorithm

- Step 1: Given λ, find inital κ and repeat steps 2-3 until convergence.

- Step 2: $(\boldsymbol{\beta} - \text{step})$ For given κ, find

$$\widehat{\boldsymbol{\beta}}_{+} = \underset{\beta_{+}}{argmin}\ L(\beta_{+}, \kappa) + \lambda \sum_{j=1}^{p} |\beta_j| \,.$$

- Step 3: $(\kappa - \text{step})$ For given β_{+}, find κ.

Computation - Secondary Taylor approximation of objective function

More succulently notation of Eq. (3) is,

$$L(\boldsymbol{\beta}_+) = -\sum_{i=1}^{n} \sum_{k \neq y_i} [f_{y_i}(\mathbf{x}_i) - f_k(\mathbf{x}_i) - \log(\kappa + \exp(f_{y_i}(\mathbf{x}_i) - f_k(\mathbf{x}_i)) \quad (\mathrm{R}^p \to \mathrm{R})$$
$$-d \log(1 + \kappa \exp(f_{y_i}(\mathbf{x}_i) - f_k(\mathbf{x}_i)))] - 2dn \log(\kappa^2 - 1)$$

let $\beta_+ = (\beta_0, \beta^T)^T$

$$k = \{1, 0, -1\}$$

$$f(\mathbf{x}) = \beta_0 + \mathbf{x}^T \beta \qquad (\mathrm{R}^p \to \mathrm{R})$$

$$f_k(\mathbf{x}) = \beta_0 + \mathbf{x}^T \beta^{(k)}$$

$$f_1(\mathbf{x}) = \beta_0 + \mathbf{x}^T \beta^{(1)} \qquad f_0(\mathbf{x}) = \beta_0 + \mathbf{x}^T \beta^{(0)} \qquad f_{-1}(\mathbf{x}) = \beta_0 + \mathbf{x}^T \beta^{(-1)}$$

Computation - Secondary Taylor approximation of objective function

Example

$$\beta = \begin{pmatrix} o \\ o \end{pmatrix} \qquad y = \begin{pmatrix} -1 \\ 1 \\ 0 \\ 1 \end{pmatrix} \qquad \mathbf{x} = \begin{pmatrix} o & o \\ o & o \\ o & o \\ o & o \end{pmatrix}$$

$$\begin{aligned} L(\boldsymbol{\beta}_+) = -\sum_{i=1}^{n} \sum_{k \neq y_i} [f_{y_i}(\mathbf{x}_i) - f_k(\mathbf{x}_i) - \log(\kappa + \exp(f_{y_i}(\mathbf{x}_i) - f_k(\mathbf{x}_i)) \\ -d \log(1 + \kappa \exp(f_{y_i}(\mathbf{x}_i) - f_k(\mathbf{x}_i)))] - 2dn \log(\kappa^2 - 1) \end{aligned}$$

$$i = 1, y_1 = -1 \Rightarrow$$

$$\begin{aligned} &\sum_{k \neq -1} [f_{-1}(\mathbf{x}_1) - f_k(\mathbf{x}_1) - \log(\kappa + \exp(f_{-1}(\mathbf{x}_1) - f_k(\mathbf{x}_1)) - d \log(1 + \kappa \exp(f_{-1}(\mathbf{x}_1) - f_k(\mathbf{x}_1)))] \\ &\quad = f_{-1}(\mathbf{x}_1) - f_1(\mathbf{x}_1) - \log(\kappa + \exp(f_{-1}(\mathbf{x}_1) - f_1(\mathbf{x}_1)) - d \log(1 + \kappa \exp(f_{-1}(\mathbf{x}_1) - f_1(\mathbf{x}_1))) \\ &\quad\ + f_{-1}(\mathbf{x}_1) - f_0(\mathbf{x}_1) - \log(\kappa + \exp(f_{-1}(\mathbf{x}_1) - f_0(\mathbf{x}_1)) - d \log(1 + \kappa \exp(f_{-1}(\mathbf{x}_1) - f_0(\mathbf{x}_1))) \end{aligned}$$

Computation - Secondary Taylor approximation of objective function

The gradient $\nabla(\boldsymbol{\beta}_+)$ $\quad (\mathrm{R}^p \rightarrow \mathrm{R}^p)$

$$k \neq y_1 : \frac{\partial L(\boldsymbol{\beta})}{\partial \beta_1^{(1)}} = -x_{11} + \frac{x_{11}\exp(f_{-1}(\mathbf{x}_1) - f_1(\mathbf{x}_1))}{\kappa + \exp(f_{-1}(\mathbf{x}_1) - f_1(\mathbf{x}_1))} + d\frac{x_{11}\kappa\exp(f_{-1}(\mathbf{x}_1) - f_1(\mathbf{x}_1))}{1 + \kappa\exp(f_{-1}(\mathbf{x}_1) - f_1(\mathbf{x}_1))}$$

$$\frac{\partial L(\boldsymbol{\beta})}{\partial \beta_2^{(1)}} = -x_{12} + \frac{x_{12}\exp(f_{-1}(\mathbf{x}_1) - f_1(\mathbf{x}_1))}{\kappa + \exp(f_{-1}(\mathbf{x}_1) - f_1(\mathbf{x}_1))} + d\frac{x_{12}\kappa\exp(f_{-1}(\mathbf{x}_1) - f_1(\mathbf{x}_1))}{1 + \kappa\exp(f_{-1}(\mathbf{x}_1) - f_1(\mathbf{x}_1))}$$

$$\frac{\partial L(\boldsymbol{\beta})}{\partial \beta_1^{(0)}} = -x_{11} + \frac{x_{11}\exp(f_{-1}(\mathbf{x}_1) - f_0(\mathbf{x}_1))}{\kappa + \exp(f_{-1}(\mathbf{x}_1) - f_0(\mathbf{x}_1))} + d\frac{x_{11}\kappa\exp(f_{-1}(\mathbf{x}_1) - f_0(\mathbf{x}_1))}{1 + \kappa\exp(f_{-1}(\mathbf{x}_1) - f_0(\mathbf{x}_1))}$$

$$\frac{\partial L(\boldsymbol{\beta})}{\partial \beta_2^{(0)}} = -x_{12} + \frac{x_{12}\exp(f_{-1}(\mathbf{x}_1) - f_0(\mathbf{x}_1))}{\kappa + \exp(f_{-1}(\mathbf{x}_1) - f_0(\mathbf{x}_1))} + d\frac{x_{12}\kappa\exp(f_{-1}(\mathbf{x}_1) - f_0(\mathbf{x}_1))}{1 + \kappa\exp(f_{-1}(\mathbf{x}_1) - f_0(\mathbf{x}_1))}$$

$$\begin{aligned} k = y_1 : \frac{\partial L(\boldsymbol{\beta})}{\partial \beta_1^{(-1)}} &= x_{11} + x_{11} - \frac{x_{11}\exp(f_{-1}(\mathbf{x}_1) - f_1(\mathbf{x}_1))}{\kappa + \exp(f_{-1}(\mathbf{x}_1) - f_1(\mathbf{x}_1))} - d\frac{x_{11}\kappa\exp(f_{-1}(\mathbf{x}_1) - f_1(\mathbf{x}_1))}{1 + \kappa\exp(f_{-1}(\mathbf{x}_1) - f_1(\mathbf{x}_1))} \\ &\quad - \frac{x_{11}\exp(f_{-1}(\mathbf{x}_1) - f_0(\mathbf{x}_1))}{\kappa + \exp(f_{-1}(\mathbf{x}_1) - f_0(\mathbf{x}_1))} - d\frac{x_{11}\kappa\exp(f_{-1}(\mathbf{x}_1) - f_0(\mathbf{x}_1))}{1 + \kappa\exp(f_{-1}(\mathbf{x}_1) - f_0(\mathbf{x}_1))} \end{aligned}$$

$$\begin{aligned} \frac{\partial L(\boldsymbol{\beta})}{\partial \beta_2^{(-1)}} &= x_{12} + x_{12} - \frac{x_{12}\exp(f_{-1}(\mathbf{x}_1) - f_1(\mathbf{x}_1))}{\kappa + \exp(f_{-1}(\mathbf{x}_1) - f_1(\mathbf{x}_1))} - d\frac{x_{12}\kappa\exp(f_{-1}(\mathbf{x}_1) - f_1(\mathbf{x}_1))}{1 + \kappa\exp(f_{-1}(\mathbf{x}_1) - f_1(\mathbf{x}_1))} \\ &\quad - \frac{x_{12}\exp(f_{-1}(\mathbf{x}_1) - f_0(\mathbf{x}_1))}{\kappa + \exp(f_{-1}(\mathbf{x}_1) - f_0(\mathbf{x}_1))} - d\frac{x_{12}\kappa\exp(f_{-1}(\mathbf{x}_1) - f_0(\mathbf{x}_1))}{1 + \kappa\exp(f_{-1}(\mathbf{x}_1) - f_0(\mathbf{x}_1))} \end{aligned}$$

Computation - Secondary Taylor approximation of objective function

$$k \neq y_1 : \frac{\partial L(\boldsymbol{\beta})}{\partial \beta_1^{(1)}} = -x_{11} + \frac{x_{11} \exp(f_{-1}(\mathbf{x}_1) - f_1(\mathbf{x}_1))}{\kappa + \exp(f_{-1}(\mathbf{x}_1) - f_1(\mathbf{x}_1))} + d \frac{x_{11} \kappa \exp(f_{-1}(\mathbf{x}_1) - f_1(\mathbf{x}_1))}{1 + \kappa \exp(f_{-1}(\mathbf{x}_1) - f_1(\mathbf{x}_1))}$$

$$k \neq y_1 : \frac{\partial L(\boldsymbol{\beta})}{\partial \beta_1^{(1)}} = x_{11} - \underbrace{x_{11} \cdot \frac{\frac{e^{f_{-1}(\mathbf{x}_1)}}{e^{f_1(\mathbf{x}_1)}}}{\kappa + \frac{e^{f_{-1}(\mathbf{x}_1)}}{e^{f_1(\mathbf{x}_1)}}}}_{1)} - \underbrace{d x_{11} \cdot \frac{\kappa \frac{e^{f_{-1}(\mathbf{x}_1)}}{e^{f_1(\mathbf{x}_1)}}}{1 + \kappa \frac{e^{f_{-1}(\mathbf{x}_1)}}{e^{f_1(\mathbf{x}_1)}}}}_{2)}$$

1) $$x_{11} - x_{11} \cdot \frac{e^{f_{-1}(\mathbf{x}_1)}}{\kappa e^{f_1(\mathbf{x}_1)} + e^{f_{-1}(\mathbf{x}_1)}}$$

$$= x_{11} \cdot \left(1 - \frac{e^{f_{-1}(\mathbf{x}_1)}}{\kappa e^{f_1(\mathbf{x}_1)} + e^{f_{-1}(\mathbf{x}_1)}}\right)$$

$$= x_{11} \cdot \left(\frac{\kappa e^{f_1(\mathbf{x}_1)} + e^{f_{-1}(\mathbf{x}_1)} - e^{f_{-1}(\mathbf{x}_1)}}{\kappa e^{f_1(\mathbf{x}_1)} + e^{f_{-1}(\mathbf{x}_1)}}\right)$$

$$= x_{11} \cdot \frac{\kappa \exp f_1(\mathbf{x}_1)}{\exp f_{-1}(\mathbf{x}_1) + \kappa \exp f_1(\mathbf{x}_1)}$$

2) $$-x_{11} d \cdot \frac{\kappa e^{f_{-1}(\mathbf{x}_1)}}{e^{f_1(\mathbf{x}_1)} + \kappa e^{f_{-1}(\mathbf{x}_1)}}$$

$$= -x_{11} d \cdot \frac{\kappa \exp f_{-1}(\mathbf{x}_1)}{\kappa \exp f_{-1}(\mathbf{x}_1) + \exp f_1(\mathbf{x}_1)}$$

$$k \neq y_1 : \frac{\partial L(\boldsymbol{\beta})}{\partial \beta_j^{(k)}} = x_{1j} \left[\frac{\kappa \exp f_k(\mathbf{x}_1)}{\exp f_{y_1}(\mathbf{x}_1) + \kappa \exp f_k(\mathbf{x}_1)} - d \frac{\kappa \exp f_{y_1}(\mathbf{x}_1)}{\kappa \exp f_{y_1}(\mathbf{x}_1) + \exp f_k(\mathbf{x}_1)}\right]$$

Computation - Secondary Taylor approximation of objective function

$$k = y_1 : \frac{\partial L(\boldsymbol{\beta})}{\partial \beta_1^{(-1)}} = x_{11} + x_{11} - \frac{x_{11}\exp(f_{-1}(\mathbf{x}_1) - f_1(\mathbf{x}_1))}{\kappa + \exp(f_{-1}(\mathbf{x}_1) - f_1(\mathbf{x}_1))} - d\frac{x_{11}\kappa\exp(f_{-1}(\mathbf{x}_1) - f_1(\mathbf{x}_1))}{1 + \kappa\exp(f_{-1}(\mathbf{x}_1) - f_1(\mathbf{x}_1))}$$
$$- \frac{x_{11}\exp(f_{-1}(\mathbf{x}_1) - f_0(\mathbf{x}_1))}{\kappa + \exp(f_{-1}(\mathbf{x}_1) - f_0(\mathbf{x}_1))} - d\frac{x_{11}\kappa\exp(f_{-1}(\mathbf{x}_1) - f_0(\mathbf{x}_1))}{1 + \kappa\exp(f_{-1}(\mathbf{x}_1) - f_0(\mathbf{x}_1))}$$

$$k = y_1 : \frac{\partial L(\boldsymbol{\beta})}{\partial \beta_1^{(-1)}} = \underbrace{-x_{11} + x_{11} \cdot \frac{\frac{e^{f_{-1}(\mathbf{x}_1)}}{e^{f_1(\mathbf{x}_1)}}}{\kappa + \frac{e^{f_{-1}(\mathbf{x}_1)}}{e^{f_1(\mathbf{x}_1)}}}}_{1)} + \underbrace{dx_{11} \cdot \frac{\kappa\frac{e^{f_{-1}(\mathbf{x}_1)}}{e^{f_1(\mathbf{x}_1)}}}{1 + \kappa\frac{e^{f_{-1}(\mathbf{x}_1)}}{e^{f_1(\mathbf{x}_1)}}}}_{2)} - x_{11} + x_{11} \cdot \frac{\frac{e^{f_{-1}(\mathbf{x}_1)}}{e^{f_0(\mathbf{x}_1)}}}{\kappa + \frac{e^{f_{-1}(\mathbf{x}_1)}}{e^{f_0(\mathbf{x}_1)}}} + dx_{11} \cdot \frac{\kappa\frac{e^{f_{-1}(\mathbf{x}_1)}}{e^{f_0(\mathbf{x}_1)}}}{1 + \kappa\frac{e^{f_{-1}(\mathbf{x}_1)}}{e^{f_0(\mathbf{x}_1)}}}$$

1) $$-x_{11} + x_{11} \cdot \frac{e^{f_{-1}(\mathbf{x}_1)}}{\kappa e^{f_1(\mathbf{x}_1)} + e^{f_{-1}(\mathbf{x}_1)}}$$
$$= x_{11} \cdot \left(-1 + \frac{e^{f_{-1}(\mathbf{x}_1)}}{\kappa e^{f_1(\mathbf{x}_1)} + e^{f_{-1}(\mathbf{x}_1)}}\right)$$
$$= x_{11} \cdot \left(\frac{-\kappa e^{f_1(\mathbf{x}_1)} + e^{f_{-1}(\mathbf{x}_1)} - e^{f_{-1}(\mathbf{x}_1)}}{\kappa e^{f_1(\mathbf{x}_1)} + e^{f_{-1}(\mathbf{x}_1)}}\right)$$
$$= x_{11} \cdot \frac{-\kappa\exp f_1(\mathbf{x}_1)}{\exp f_{-1}(\mathbf{x}_1) + \kappa\exp f_1(\mathbf{x}_1)}$$

2) $$x_{11}d \cdot \frac{\kappa e^{f_{-1}(\mathbf{x}_1)}}{e^{f_1(\mathbf{x}_1)} + \kappa e^{f_{-1}(\mathbf{x}_1)}}$$
$$= x_{11}d \cdot \frac{\kappa\exp f_{-1}(\mathbf{x}_1)}{\kappa\exp f_{-1}(\mathbf{x}_1) + \exp f_1(\mathbf{x}_1)}$$

$$k = y_1 : \frac{\partial L(\boldsymbol{\beta})}{\partial \beta_j^{(k)}} = x_{1j}\sum_{g \neq y_1}\left[\frac{-\kappa\exp f_g(\mathbf{x}_1)}{\exp f_{y_1}(\mathbf{x}_1) + \kappa\exp f_g(\mathbf{x}_1)} + d\frac{\kappa\exp f_{y_1}(\mathbf{x}_1)}{\kappa\exp f_{y_1}(\mathbf{x}_1) + \exp f_g(\mathbf{x}_1)}\right]$$

Computation - Secondary Taylor approximation of objective function

$$k \neq y_1 : \frac{\partial L(\boldsymbol{\beta})}{\partial \beta_j^{(k)}} = x_{1j} \left[\frac{\kappa \exp f_k(\mathbf{x}_1)}{\exp f_{y_1}(\mathbf{x}_1) + \kappa \exp f_k(\mathbf{x}_1)} - d \frac{\kappa \exp f_{y_1}(\mathbf{x}_1)}{\kappa \exp f_{y_1}(\mathbf{x}_1) + \exp f_k(\mathbf{x}_1)} \right]$$

$$k \neq y_i : \frac{\partial L(\boldsymbol{\beta})}{\partial \beta_j^{(k)}} = \sum_{i=1}^{n} x_{ij} \left[\frac{\kappa \exp f_k(\mathbf{x}_i)}{\exp f_{y_i}(\mathbf{x}_i) + \kappa \exp f_k(\mathbf{x}_i)} - d \frac{\kappa \exp f_{y_i}(\mathbf{x}_i)}{\kappa \exp f_{y_i}(\mathbf{x}_i) + \exp f_k(\mathbf{x}_i)} \right]$$

$$k = y_1 : \frac{\partial L(\boldsymbol{\beta})}{\partial \beta_j^{(k)}} = x_{1j} \sum_{g \neq y_1} \left[\frac{-\kappa \exp f_g(\mathbf{x}_1)}{\exp f_{y_1}(\mathbf{x}_1) + \kappa \exp f_g(\mathbf{x}_1)} + d \frac{\kappa \exp f_{y_1}(\mathbf{x}_1)}{\kappa \exp f_{y_1}(\mathbf{x}_1) + \exp f_g(\mathbf{x}_1)} \right]$$

$$k = y_i : \frac{\partial L(\boldsymbol{\beta})}{\partial \beta_j^{(k)}} = \sum_{i=1}^{n} x_{ij} \sum_{g \neq y_i} \left[\frac{-\kappa \exp f_g(\mathbf{x}_i)}{\exp f_{y_i}(\mathbf{x}_i) + \kappa \exp f_g(\mathbf{x}_i)} + d \frac{\kappa \exp f_{y_i}(\mathbf{x}_i)}{\kappa \exp f_{y_i}(\mathbf{x}_i) + \exp f_g(\mathbf{x}_i)} \right]$$

Computation - Secondary Taylor approximation of objective function

$$k \neq y_i : \frac{\partial L(\boldsymbol{\beta})}{\partial \beta_j^{(k)}} = \sum_{i=1}^{n} x_{ij} \left[\frac{\kappa \exp f_k(\mathbf{x}_i)}{\exp f_{y_i}(\mathbf{x}_i) + \kappa \exp f_k(\mathbf{x}_i)} - d \frac{\kappa \exp f_{y_i}(\mathbf{x}_i)}{\kappa \exp f_{y_i}(\mathbf{x}_i) + \exp f_k(\mathbf{x}_i)} \right]$$

$$k \neq y_i : \frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_j^{(k)} \partial \beta_l^{(k)}}$$

$$= \sum_{i=1}^{n} x_{ij} \cdot \left[\frac{x_{il} \kappa \exp f_k(\mathbf{x}_i) \times (\exp f_{y_i}(\mathbf{x}_i) + \kappa \exp f_k(\mathbf{x}_i)) - \kappa \exp f_k(\mathbf{x}_i) \kappa \exp f_k(\mathbf{x}_i) \cdot x_{il}}{(\exp f_{y_i}(\mathbf{x}_i) + \kappa \exp f_k(\mathbf{x}_i))^2} + d \cdot \frac{\kappa \exp f_{y_i}(\mathbf{x}_i) \exp f_k(\mathbf{x}_i) \cdot x_{il}}{(\kappa \exp f_{y_i}(\mathbf{x}_i) + \exp f_k(\mathbf{x}_i))^2} \right]$$

Hessian matrix be $\mathbf{H}$ with an element of being the second derivatirve of j and l^{th} variable of k class

$$k \neq y_i : \frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_j^{(k)} \partial \beta_l^{(k)}} = \sum_{i=1}^{n} x_{ij} x_{il} \left(\frac{\kappa \exp f_{y_i}(\mathbf{x}_i) \exp f_k(\mathbf{x}_i)}{(\exp f_{y_i}(\mathbf{x}_i) + \kappa \exp f_k(\mathbf{x}_i))^2} + \frac{d\kappa \exp f_{y_i}(\mathbf{x}_i) \exp f_k(\mathbf{x}_i)}{(\kappa \exp f_{y_i}(\mathbf{x}_i) + \exp f_k(\mathbf{x}_i))^2} \right)$$

Computation - Secondary Taylor approximation of objective function

$$k = y_i : \frac{\partial L(\boldsymbol{\beta})}{\partial \beta_j^{(k)}} = \sum_{i=1}^{n} x_{ij} \sum_{g \neq y_i} \left[\frac{-\kappa \exp f_g(\mathbf{x}_i)}{\exp f_{y_i}(\mathbf{x}_i) + \kappa \exp f_g(\mathbf{x}_i)} + d \frac{\kappa \exp f_{y_i}(\mathbf{x}_i)}{\kappa \exp f_{y_i}(\mathbf{x}_i) + \exp f_g(\mathbf{x}_i)} \right]$$

$$k = y_i : \frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_j^{(k)} \partial \beta_l^{(k)}}$$

$$= \sum_{i=1}^{n} x_{ij} \sum_{g \neq y_i} \left(\frac{\kappa \exp f_g(\mathbf{x}_i) \exp f_{y_i}(\mathbf{x}_i) \cdot x_{il}}{(\exp f_{y_i}(\mathbf{x}_i) + \kappa \exp f_g(\mathbf{x}_i))^2} + d \cdot \frac{x_{il} \cdot \kappa \exp f_{y_i}(\mathbf{x}_i) \times (\kappa \exp f_{y_i}(\mathbf{x}_i) + \exp f_g(\mathbf{x}_i)) - \kappa \exp f_{y_i}(\mathbf{x}_i) \cdot \exp f_{y_i}(\mathbf{x}_i) \cdot x_{il}}{(\kappa \exp f_{y_i}(\mathbf{x}_i) + \exp f_g(\mathbf{x}_i))^2} \right)$$

Hessian matrix be $\mathbf{H}$ with an element of being the second derivatirve of j and l^{th} variable of k class

$$k = y_i : \frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_j^{(k)} \partial \beta_l^{(k)}} = \sum_{i=1}^{n} \sum_{g \neq y_i} x_{ij} x_{il} \left(\frac{\kappa \exp f_{y_i}(\mathbf{x}_i) \exp f_g(\mathbf{x}_i)}{(\exp f_{y_i}(\mathbf{x}_i) + \kappa \exp f_g(\mathbf{x}_i))^2} + \frac{d\kappa \exp f_{y_i}(\mathbf{x}_i) \exp f_g(\mathbf{x}_i)}{(\kappa \exp f_{y_i}(\mathbf{x}_i) + \exp f_g(\mathbf{x}_i))^2} \right)$$

Computation - Secondary Taylor approximation of objective function

$$\text{The gradient } \nabla(\boldsymbol{\beta}_{+}) = \left(\frac{\partial L(\boldsymbol{\beta})}{\partial \beta_j^{(k)}}\right) = \left(\frac{\partial L(\boldsymbol{\beta})}{\partial \beta_1^{(k)}}, \frac{\partial L(\boldsymbol{\beta})}{\partial \beta_2^{(k)}}\right) \qquad (\mathrm{R}^p \to \mathrm{R}^p)$$

$$\mathbf{H}_k = \begin{pmatrix} h_0^{(k)} & \mathbf{h}_0^{(k)\prime} \\ \mathbf{h}_0^{(k)} & H_k \end{pmatrix} = \begin{pmatrix} \frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_1^{(k)} \partial \beta_1^{(k)}} & \frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_1^{(k)} \partial \beta_2^{(k)}} \\ H_0 & 0 \\ \frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_2^{(k)} \partial \beta_1^{(k)}} & \frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_2^{(k)} \partial \beta_2^{(k)}} \end{pmatrix}$$

$$(\mathrm{R}^p \to \mathrm{R}^{p \times p})$$

$$\text{Hessian matrix } \mathbf{H} = \begin{pmatrix} H_1 & 0 & 0 \\ 0 & H_0 & 0 \\ 0 & 0 & H_{-1} \end{pmatrix}$$

By Talyor expansion,

$$L(\boldsymbol{\beta}) \approx L(\boldsymbol{\gamma}) + \sum_{k=1}^{K} \nabla^{(k)}(\boldsymbol{\gamma}_{+})'(\beta_{+}^{(k)} - \gamma_{+}^{(k)}) + \frac{1}{2} \sum_{k=1}^{K} (\beta_{+}^{(k)} - \gamma_{+}^{(k)})' \mathbf{H}_k (\beta_{+}^{(k)} - \gamma_{+}^{(k)})$$

$$+\frac{1}{2}$$

Computation - Secondary Taylor approximation of objective function

We can find optimizer of

$$L(\boldsymbol{\beta}_+) = f(\boldsymbol{\gamma}_+) + \sum_{k=1}^{K} \nabla^{(k)}(\boldsymbol{\gamma}_+)'(\beta_+^{(k)} - \gamma_+^{(k)}) + \frac{1}{2}\sum_{k=1}^{K}(\beta_+^{(k)} - \gamma_+^{(k)})'\mathbf{H}_k(\beta_+^{(k)} - \gamma_+^{(k)})$$

$$\underline{+ \lambda \sum_{k=1}^{K} \|\beta^{(k)}\|_1}$$

with CD algorithm

Or solve regression problem of corresponding Lagrange function

$$L(\boldsymbol{\beta}_+) = f(\boldsymbol{\gamma}_+) + \sum_{k=1}^{K} \nabla^{(k)}(\boldsymbol{\gamma}_+)'(\beta_+^{(k)} - \gamma_+^{(k)}) + \frac{1}{2}\sum_{k=1}^{K}(\beta_+^{(k)} - \gamma_+^{(k)})'\mathbf{H}_k(\beta_+^{(k)} - \gamma_+^{(k)})$$

$$\underline{+ \eta(\sum_{k=1}^{K} \|\beta^{(k)}\|_1 - s)}$$

with LARS algorithm. - Efron (2004)

Simulation

- Simulation Data
- Metrics for comparison
- Numerical result

3

Simulation

Simulation Data

n : sample size

μ^+ : p – dimensional vector whose first q entries are D and the other $p-q$ entries are zero. $(\mu^- = -\mu^+)$

Σ : variance–covariance matrix that (k,l) entry of it be $r^{|k-l|}$, where $r \in [0,1)$

$$\mathcal{X} = \mathcal{X}_1 + \mathcal{X}_2$$

$\mathcal{X}_1 \sim N_p(\mu^+, \Sigma)$, assign $y = 1$ $\mathcal{X}_2 \sim N_p(\mu^-, \Sigma)$, assign $y = -1$

Simulation

Simulation Data – Regularization Parameter Tuning

d : controlling parameter for the reject probability

λ : shrinkage parameter in Lasso penalty term

- Training set : $n = 100$
- Validation set : $n = 100$
- Test set : $n = 2000$

Simulation

Metrics for comparison

- Total MIS
  misclassification error rates obtained based on all observations in a test sample
- Accept MIS
  misclassification error rates obtained based only on accepted observations by a lgs-R
- Reject MIS
  misclassification error rates obtained based only on rejected observations by a lgs-R
- Reject rate
  portion of rejected observations by a lgs-R
- p-value
  obtained by the Wilcoxon signed-rank test with 30 paired Total MIS of lgs and lgs-R

Simulation

Numerical result

Table 1
Comparison between prediction accuracy of the lgs and those of lgs-R : average misclassification errors (standard errors).

p	r	Method	Total MIS	Accept MIS	Reject MIS	Reject rate	p-value
100	0	lgs	0.1589(6e-04)	0.1169(0.0019)	0.3528(0.0034)		
		lgs-R	**0.1571**(6e-04)	**0.1162**(0.0019)	**0.3472**(0.0037)	0.1725(0.0074)	0.0313
	0.3	lgs	0.2236(6e-04)	0.1621(0.0023)	0.3697(0.0019)		
		lgs-R	**0.2202**(6e-04)	**0.1610**(0.0023)	**0.3642**(0.0021)	0.2418(0.0085)	0.0201
	0.6	lgs	0.2655(7e-04)	0.1960(0.0021)	0.3963(0.0018)		
		lgs-R	**0.2601**(7e-04)	**0.1939**(0.0021)	**0.3913**(0.0017)	0.2889(0.009)	0.0083

- lgs-R always has significantly lower misclassification errors than the lgs, and the improvements are statistically significant in all cases.

Conclusion

4

Conclusion

- Analysis of simulated data sets proved that lgs-R has better performance than lgs.
- Analysis of simulated data sets suggested the potential for the application of lgs-R for real data sets.
- The real data to which lgs-R is applied may be limited.
- Multiple regularization parameters for tuning (λ, d) are exists.

Q & A