Group Project 2 & 3

Xinlei Chen, Guojing Wu, Yujing Yao

Introduction

- **Group project 2:** optimization algorithms on a breast cancer diagnosis dataset
- **Group project 3:** bootstrapping on developing classification model

Breast Cancer Data

- **Variable Selection:** reduce multicollinearity based on both correlation coefficient and eigenvalue of correlation matrix
- 18 variables remain

Logistic Model with Newton-Raphson

Logistic Regression:

y: the vector of n response random variable

X: the $n \times p$ design matrix (X_i denote the ith row)

β: the $p \times 1$ coefficient

- The logistic regression model:

$$\log(\frac{\eta}{1-\eta}) = X\beta$$

- The likelihood function:

$$L(\beta; X, y) = \prod_{i=1}^{n} \{ (\frac{\exp(X_i\beta)}{1+\exp(X_i\beta)})^{y_i} (\frac{1}{1+\exp(X_i\beta)})^{1-y_i} \}$$

- The log likelihood:

$$l(\beta) = \sum_{i=1}^{n} \{y_i(X_i\beta) - \log(1 + \exp(X_i\beta))\}$$

- The gradient:

$$\nabla l(\beta) = X^T(y - p)$$

  where $p = \frac{\exp(X\beta)}{1+\exp(X\beta)}$
- The Hessian:

$$\nabla^2 l(\beta) = -X^T WX$$

  where $W = diag(p_i(1 - p_i)), i = 1, \cdots, n$. The Hessian is negative definite.

Newton-Raphson

Update coefficients

$$\beta_{i+1} = \beta_i - [\nabla^2 l(\beta_i)]^{-1} \nabla l(\beta_i)$$

Step-halving

$$\beta_{i+1}(\gamma) = \beta_i - \gamma[\nabla^2 l(\beta_i)]^{-1} \nabla l(\beta_i)$$

- Set $\gamma = 1$
- If $f(\theta_{i+1}(1)) \geq f(\theta_i)$, then set $\theta_{i+1} = \theta_{i+1}(1)$
- If $f(\theta_{i+1}(1)) \leq f(\theta_i)$, search for a value $\gamma \in (0, 1)$ for which $f(\theta_{i+1}(\gamma)) \geq f(\theta_i)$, set $\theta_{i+1} = \theta_{i+1}(\gamma)$

Newton-Raphson: gradient decent

For Newton's method with a large p, the computational burden in calculating the inverse of the Hessian Matrix $[\nabla^2 f(\beta_i)]^{-1}$ increases quickly with p. One can update

$$\beta_{i+1} = \beta_i + H_i \nabla f(\beta_i)$$

where $H_i = (X^T X)^{-1}$ for every i. This is easy to compute, but could be slow in convergence.

The steps are:

- get the objective (loglik,grad,Hess) function
- use the principle of newton raphson to update the estimate, if the step size too large, step-halving step
- stop searching until the convergences of the estimates.

Logistic-LASSO Model with Pathwise Coordinate Descent

- Applied coordinate-wise descent with weighted update:

$$\tilde{\beta}_j^{lasso}(\lambda) \leftarrow \frac{S(\sum_{i=1}^n \omega_i x_{i,j}(y_i - \tilde{y}_i^{(-j)}), \lambda)}{\sum_{i=1}^n \omega_i x_{i,j}^2}$$

  where $\tilde{y}_i^{(-j)} = \sum_{k \neq j} x_{i,k}\tilde{\beta}_k$ and $S(\hat{\beta}, \lambda) = sign(\hat{\beta})(|\hat{\beta}| - \lambda)_+$

- In the context of logistic regression, we are aiming to maximize the penalized log likelihood:

$$\max_{\beta \in \mathbb{R}^{p+1}} \frac{1}{n} \sum_{i=1}^n \{y_i(X_i\beta) - \log(1 + \exp(X_i\beta))\} - \lambda \sum_{j=0}^p |\beta_j|$$

  for some $\lambda \geq 0$

Estimation Path

Cross validation for LASSO

Model Comparison

Conclusion

Down Syndrome Data