P8160 Group Project 2: Breast Cancer Diagnosis and Optimizations

Wenhan Bao | Tianchuan Gao | Jialiang Hua | Qihang Wu | Paula Wu

3/25/2022

Objective

Background & Statistical Methods

Background

Statistical Methods

Modified Newton-Raphson Algorithm

Logistic-LASSO

Coordinate-wise Descent Algorithm Usually, to estimate coefficients through Newton-Raphson method, it is necessary to compute the corresponding inverse of Hessian Matrix $[\nabla^2 f(\theta_{i-1})]^{-1}$. However, the computational burden of calculation will increase as the dimension of predictors increases and the collinearity will also be a problem. Therefore, we use a regularization method, (Least Absolute Shrinkage and Selection Operator) LASSO, to shrink coefficients and perform variable selections. For regression, LASSO tries to minimize the following objective function:

$$f(\beta) = \frac{1}{2}\sum_{i=1}^{n}(y_i - \sum_{j=1}^{p} x_{i,j}\beta_j)^2 + \gamma\sum_{j=1}^{p}|\beta_j|,$$

where the first term is residual sum of squares (RSS) and the second term is a $l1$ penalty from LASSO. Noted that the $x_{i,j}$ needs to be standardized before LASSO to ensure $\sum_{i=1}^{n}\frac{x_{i,j}}{n} = 0$ and $\sum_{i=1}^{n} x_{i,j}^2 = 1$ so that the penalty will be equally applied to all predictors. For each single predictor, the LASSO solution is like:

$$\hat{\beta}^{lasso}(\gamma) = S(\hat{\beta}, \gamma) = \begin{cases} \hat{\beta} - \gamma, \text{ if } \hat{\beta} > 0 \text{ and } \gamma < |\hat{\beta}| \\ \hat{\beta} + \gamma, \text{ if } \hat{\beta} < 0 \text{ and } \gamma < |\hat{\beta}| \\ 0, \text{ if } \gamma > |\hat{\beta}| \end{cases}$$

, where $S(\hat{\beta}, \gamma)$ is called soft threshold. The basic idea of this function is to shrink all β coefficients between $-\gamma$ and γ.

Pathwise Coordinate-wise Algorithm

5-fold Cross-validation

Conclusions

Logistic Regression Model

Logistic-LASSO Model

Discussion

Contributions

We contributed to this project evenly.

Appendix

Reference