AMS 598 Project 4 Report

Jiecheng Song

Nov 2021

1 Introduction

One big dataset has been partitioned into 10 chunks and are stored under the shared class space on SeaWulf: /gpfs/projects/AMS598/Projects/project4

Each file contains one column called 'y', which is a binary 0/1 response variable, and 25 other columns named x1 – x25, representing explanatory variables.

Use the ADMM algorithm to run a logistic regression on all data, and obtain one set of consensus estimates for the coefficients of the explanatory variables.

Write a report about how you did the analysis. Submit your scripts with the report.

2 Tools Selected

When we update beta, we can calculate the beta that maximum objective function (likelihood + multiplier) for each part of data simultaneously. This is a map reduce problem, but it takes long time to wait in queue for Seawulf, and the mission is relatively small (about 2-3 min for each part of dataset), so I chose to use local device with R code to conduct this project.

3 Methods (Ordinary)

We used ADMM algorithm to estimate the coefficients of a logistic regression, so we should follow the update equations below iteratively to get the results:

$$\beta_i^{k+1} = argmin_{\beta_i}(l_i(y_i, X_i^T \beta_i) + \rho/2||\beta_i - \bar{\beta}^k + u_i^k||_2^2)$$
$$u_i^{k+1} = u_i^k + (\beta_i^{k+1} - \bar{\beta}^{k+1})$$

where l_i is the negative log likelihood function. The second equation is not hard to update, but the first update equation, we need to use gradient descent to get the optimization values.

We need use derivatives to get the descents:

$$\frac{\partial f}{\partial \beta_i} = X^T(y - \mu) + \rho(\beta_i - \bar{\beta}^k + u_i^k)$$

where $\mu = g(X^T\beta)$ and $g(z = 1/(1+e^{-z})$. So β_i^{k+1} can be got by gradient descent:

$$\beta_{t+1} = \beta_t + \alpha * \frac{\partial f}{\partial \beta_i}$$

Until the improvement of objective function less than a threshold.

4 Parameters (Ordinary)

Due to we have 1,000,000 samples in each part of data, so we chose a large at first, then we check the improvement of mean likelihood of all the 10 data part with coefficients and when the likelihood improve slow, we replace the ρ to 100 and 10.

For gradient descent, I chose 0.01 as threshold, if the improvement of objective function less than 0.01 then stopped the iteration. And set the step size $\alpha = 1e-6$.

I used the coefficients of logistic regression on part 1 data as the initial β_i^0 for each data set.

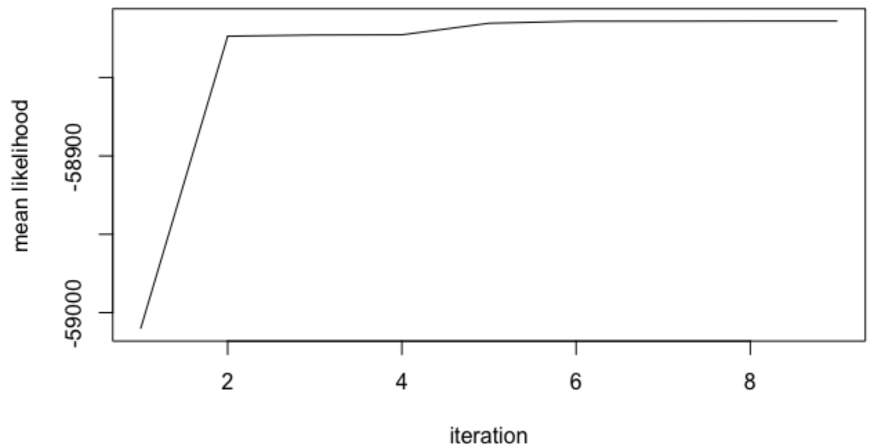
5 Results (Ordinary)

5.1 coefficients

coefficients	estimation	coefficients	estimation
Intercept	-1.962059	X13	-0.002833
X1	1.9807672	X14	-0.0101713
X2	0.7959822	X15	-0.0039285
X3	2.50758577	X16	0.00525062
X4	-0.0042246	X17	-0.0034817
X5	-0.0089611	X18	-0.0024831
X6	0.00071503	X19	-0.0055332
X7	-9.32E-05	X20	-0.0004567
X8	0.0106815	X21	0.00215759
X9	-0.000696	X22	-0.0004249
X10	0.69986891	X23	1.78446584
X11	0.00370527	X24	-0.0064923
X12	-0.0074298	X25	2.98971444

Significant variable (which absolute value of coefficients is greater than 0.02) is highlighted.

5.2 Likelihood Curve



At 4th iteration, the parameter is changed, so it increased in a faster speed again.

6 Methods (LASSO)

I also conducted a variable selection with LASSO, then coefficients are updated with following iterations:

$$\beta_i^{k+1} = argmin_{\beta_i}(l_i(y_i, X_i^T \beta_i) + \rho/2 * ||\beta_i - \beta^k + u_i^k||_2^2)$$
$$\beta^{k+1} = argmin_{\beta}(\lambda * \sum |\beta| + N\rho/2 * ||\beta - \bar{\beta}^{k+1}||)$$
$$u_i^{k+1} = u_i^k + (\beta_i^{k+1} - \bar{\beta}^{k+1})$$

The first and third equations are easy to update, following the steps in ordinary logistic regression, by calculation, the second equation solution will be that:

$$\beta^{k+1} = \bar{\beta}^{k+1} - sign(\bar{\beta}^{k+1}) * \frac{\lambda}{N\rho}, if|\beta^{k+1}| > \frac{\lambda}{N\rho}$$

$$\beta^{k+1} = 0, if|\beta^{k+1}| <= \frac{\lambda}{N\rho}$$

7 Parameters (LASSO)

In LASSO, I basically used the same parameters with ordinary logistic regression, and to be convenient, I directly set $\frac{\lambda}{N\rho} = 0.01$ as the threshold for penalty.

8 Results (LASSO)

8.1 (coefficients)

coefficients	estimation	coefficients	estimation
Intercept	-1.969469	X13	0
X1	1.9776288	X14	0
X2	0.7874780	X15	0
X3	2.4804172	X16	0
X4	0	X17	0
X5	0	X18	0
X6	0	X19	0
X7	0	X20	0
X8	0.000884	X21	0
X9	0	X22	0
X10	0.693365	X23	1.77494
X11	0	X24	0
X12	0	X25	2.980325

With LASSO, all the significant variable was selected out and other was penalized to 0, except X8, we can increase the penalty parameter to eliminate it.