Bayesian Logistic Regression Model for Credit Score Data

Presented by Hongfei Li, Lai Wei

Data source

Our data set consists of 1312 applications for credit cards and their results (approved or rejected). The data comes from Professor William Greene's (New York University) on-line data for his book "Econometric Analysis, 5th Edition". (http://pages.stern.nyu.edu/ wgreene/Text/economet ricanalysis.htm)

Data source (Cont.)

Data set contains one categorical response variable and 10 explanatory variables.

- Approval = response/output. 1 if application for credit card accepted, 0 if not.
- MDR = Number of major derogatory reports.
- Age = Age n years plus twelfths of a year.
- Income = Yearly income (divided by 10,000)
- IncPer = Yearly income per dependent (divided by 10,000)
- Ownrent = Dummy variable, 1 if owns their home, 0 if rent
- Selfempl = Dummy variable, 1 if self employed, 0 if not.
- Dependent = 0 + number of dependents.
- Curadd = months living at current address.
- ActiveCard = number of active credit accounts
- MajorCard = number of major credit cards held.

Purpose

- Randomly choose 2/3 data as the training data, and the left as testing data.
- Fit linear logistic regression model. (Generalized Linear Model(GLM))
- Fit Bayesian logistic regression model(Bayesian Generalized Linear Model(BGLM)) based on different prior distribution.

MCMC by Bugs from R

- To perform MCMC sampling, we use the BUGS statistical package in R working directory.
- We discard the first 500 draws of the parameter values (the burn-in).
- After checking some criteria for evidence of convergence. we obtain 1000 more draws.

Posterior distribution

- We build a hierarchical Bayesian logistic regression model, with $\text{logit}(p_i) = X\beta$, where X is the covariate $n \times p$ matrix, and β is the coefficient $p \times 1$ vector.

- $$y_i | p_i \sim \text{Ber}(p_i)$$

  where $p_i = \text{logit}^{-1}(X\beta)$.

- The prior distribution for β_j is

  $$\beta_j \sim N(\beta_{0j}, \sigma_j^2).$$

- The posterior distribution for β is

$$\begin{aligned} f(\beta|y) &\propto f(y, \beta) \\ &\propto f(y|\beta) f(\beta) \\ &\propto \prod_{i=1}^{n} \left(\text{logit}^{-1}(X\beta)\right)^{y_i} \left(1 - \text{logit}^{-1}(X\beta)\right)^{1-y_i} \prod_{j=1}^{p} \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left(-\frac{(\beta_j - \beta_{0j})^2}{2\sigma_j^2}\right) \end{aligned}$$

Linear Logistic Regression

The best model in linear logistic regression is:

$$logit(p_i) = log(\frac{p_i}{1-p_i}) = \beta_0 + \beta_1(MDR) + \beta_2(Selfempl) + \beta_3(Ownrent) + \beta_4(Income) + \beta_5(Dependent) + \beta_6(ActiveCard)$$

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	0.81266	0.23525	3.454	0.000551	***
MDR	-1.63581	0.16102	-10.159	< 2e-16	***
Selfempl	-0.69401	0.35242	-1.969	0.048919	*
Ownrent	0.51542	0.22709	2.270	0.023228	*
Income	0.19760	0.07377	2.679	0.007392	**
Dependent	-0.30401	0.08109	-3.749	0.000177	***
ActiveCard	0.11425	0.02081	5.491	4.01e-08	***

BGLM based on the model selected by GLM

Add some noninformative prior to the parameters,

$$\beta \sim (0, 10^3 \times \mathbf{I})$$

where $\beta = (\beta_0, \beta_1, \cdots, \beta_6)^T$, and $\mathbf{I}$ are identity matrix.

parameter	2.5 %	25%	50%	75%	97.5%	mean	sd
intercept	0.350	0.644	0.810	0.971	1.267	0.807	0.239
MDR	-1.986	-1.777	-1.660	-1.546	-1.347	-1.663	0.166
Income	0.066	0.157	0.207	0.255	0.351	0.207	0.074
Ownrent	0.101	0.360	0.518	0.670	0.979	0.521	0.233
Selfempl	-1.394	-0.984	-0.721	-0.455	0.029	-0.710	0.364
Dependent	-0.459	-0.360	-0.307	-0.256	-0.133	-0.307	0.081
ActiveCard	0.075	0.099	0.115	0.129	0.157	0.115	0.022

Table 1: posterior distribution paramters summary with noninformative prior

Compare GLM with BGLM

- The MLE in GLm are very close to the Bayesian estimators(posterior mean and posterior median).
- They have the same 84.47% of correct classification for the testing data set.
- For GLM, all of the 95% CI don't include zero. However, 95% CI posterior CI for 'Selfempl' variable dose include zero for BGLM.

BGLM selection

We include all the explanatory variables in the model and set the noninformative priors. We select explanatory variables based on whether the 95% posterior CI include zero or not.

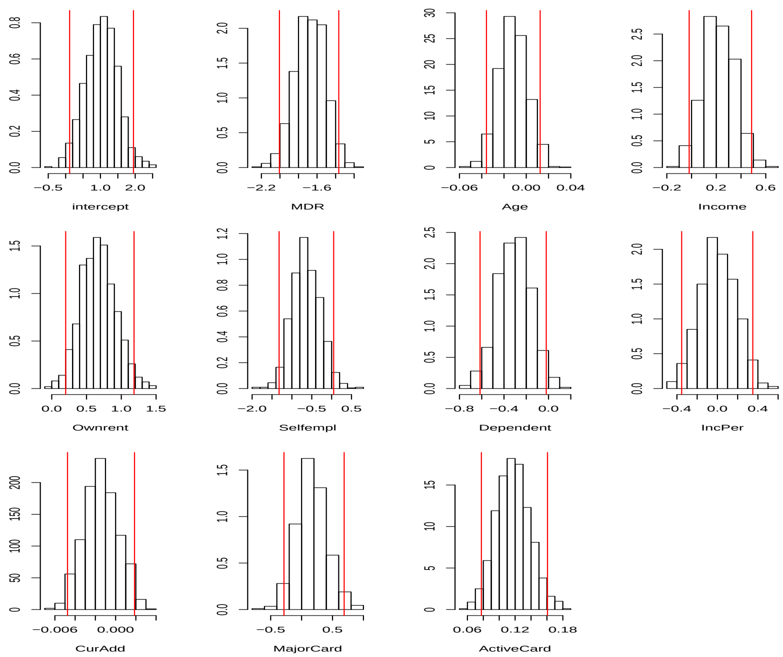

Figure 1: Posterior distribution of the parameters. (Red lines give the 95% posterior confidence interval

BGLM selection(Cont.)

Finally, we have the model:

$$logit(p_i) \quad = \quad \beta_0 + \beta_1(MDR) + \beta_2(Ownrent) + \beta_4(Dependent) + \beta_5(ActiveCard)$$

which is different from the model we got in GLM. However, it has few variables and 83.98% of correct classification, which is a little bit samller than before.

Prior selection

- Informative Prior:
  Although we do not have experts to provide any extra information, we can obtain that by the previous experiment, ie, the MLE given by GLM.

$$\begin{aligned}
\beta_{intercept} &\sim N(0.81266, 0.0553) \\
\beta_{MDR} &\sim N(-1.63581, 0.0259) \\
\beta_{Income} &\sim N(0.19760, 0.005442) \\
\beta_{Ownrent} &\sim N(0.51542, 0.05157) \\
\beta_{Selfempl} &\sim N(-0.69401, 0.1242) \\
\beta_{Dependent} &\sim N(-0.30401, 0.0065756) \\
\beta_{ActiveCard} &\sim N(0.11425, 0.000433)
\end{aligned}$$

Prior selection(Cont.)

- Mixed prior:
  We assign some informative priors on some β_j, since we have some confidence to believe the relationship between the covariates and the response. We assign information by using uniform distribution and put restrictions on the areas that the parameters can choose.

$$
\begin{aligned}
\beta_{intercept} &\sim N(0, 1000) \\
\beta_{MDR} &\sim Uniform(-5, 0) \\
\beta_{Income} &\sim N(0, 1000) \\
\beta_{Ownrent} &\sim N(0, 1000) \\
\beta_{Selfempl} &\sim N(0, 1000) \\
\beta_{Dependent} &\sim Uniform(-1, 0) \\
\beta_{ActiveCard} &\sim N(0, 1000)
\end{aligned}
$$

Prior selection

Prior	Percentages of correct classification
noninformative	0.8447
informative	0.8447
mixed	0.8592

Table 2: Predicitive accuracy results

- Noninformative prior doesn't improve the predictive accuracy compared to GLM.
- Comparing informative prior with noninformative prior, they have the same predictive accuracy.
- For mixed priors. The result indicates that the mixed priors improve the model.

Discussion

- Since Bayesian method takes more time and effect than the frequentist method, we explore if it is worth using Bayesian model for our dataset.
- Choosing priors is very important.
- Bayesian model sometimes is sensitive to the prior choosing.