Regression and Model Selection

Fall 2022, MATH8050: Homework 9
Your Name, Section XXX

Due November 9, 12:00 PM

General instructions for homeworks: Please follow the uploading file instructions according to the syllabus. Each answer must be supported by written statements as well as any code used. Your code must be completely reproducible and must compile. For writing mathematical expressions in R Markdown, refer to the homework template posted on Canvas, a 30-minute tutorial, or LaTeX/Mathematics.

Advice: Start early on the homeworks and it is advised that you not wait until the last day. While the professor and the TA's check emails, they will be answered in the order they are received and last minute help will not be given.

No late homeworks will be accepted.

R Working Environment

Please load all the packages used in the following R chunk before the function `sessionInfo()`

```r
# load packages


sessionInfo()
```

Total points on assignment: 10 (reproducibility) + 30 (Q1) + 60 (Q2)

Reproducibility component: 10 points.

1. (30pts total, equally weighted) PH Exercise 9.2 (the diabetes data)

2. (60pts total, equally weighted) Consider the diabetes data `azdiabetes.dat` in class. The goal here is to fit a Bayesian logistic regression model with the variable `diabetes` as the response and `npreg`, `bp`, `bmi`, `pred`, and `age` as the covariates. Suppose that the logistic regression model we consider is of the form $\Pr(Y_i = 1 \mid x_i, \beta, \gamma) = e^{\theta_i}/(1 + e^{\theta_i})$ where $\beta = (\beta_0, \ldots, \beta_5)$, $\gamma = (\gamma_1, \ldots, \gamma_5)$ and

$$\theta_i = \beta_0 + \sum_{j=1}^{5} \beta_j \gamma_j x_{i,j}$$

   Here $\gamma_j = 1$ if the jth variable is a predictor of diabetes and 0 otherwise. For example, $\gamma = (1, 1, 0, 0, 0)$ corresponds to the model $\theta_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2}$. Obtain posterior distribution of β and γ, assuming the following independent priors.

$$\gamma_j \sim \text{Ber}(0.5), \quad \beta_0 \sim \text{Normal}(0, 16), \quad \beta_j \sim \text{Normal}(0, 4)$$

for each $j > 0$.

a. Derive the full conditional distributions for β_j with $j = 0, 1, \ldots, 5$ and γ_j with $j = 1, \ldots, 5$.

b. Implement a Metropolis-Hastings algorithm to obtain MCMC samples from the joint posterior distribution and perform convergence diagnostics.

c. Report the 95% credible intervals for the parameters β_j with $j = 0, 1, \ldots, 5$ and report the posterior including probabilities for each covariate.