Homework 2

Data

The data in 'auto.csv' (a cleaned data set from UCI) represent quantitative characteristics of a sample of $n = 392$ automobiles available in the 1970s and 1980s. For this homework assignment, we seek to construct a Bayesian regression model to predict miles per gallon (mpg) based on the number of engine cylinders (cylinders) and year of the car model (model.year).

Questions

Prepare a written response to the following, using Overleaf. The assignment shouldn't be longer than 10 (double-spaced, excluding title page, references, and appendices). Due Thurs., Feb. 9, at the beginning of the class period. Please submit the assignment as a PDF through CANVAS.

1. Develop a MCMC algorithm to fit a Bayesian regression model using a normal likelihood, multivariate normal prior for the coefficients $\boldsymbol{\beta}$, and normal prior for $\log(\sigma)$. Use a random walk proposal for $\log(\sigma)$ in the Metropolis-Hastings updates for $\log(\sigma)$.

2. Conduct a Bayesian regression analysis based on the data set using miles per gallon as the response variable and the three sets of covariates below. Compare the 3 models using DIC.

   (a) cylinders and model.year

   (b) cylinders

   (c) model.year

3. For the best performing model based on DIC, make inference about your findings using the associated MCMC sample.

4. For an automobile that is not in the data set but has `cylinders` $= 8$ and `model.year` $= 77$, predict the miles per gallon using the best performing model you identified above.

References

- `https://archive.ics.uci.edu/ml/datasets/auto+mpg`