Lab 11

Multiple Linear Regression

Ex. 1. (3.5p) What factors determined the price of a PC at the time of their introduction? The file Prices.csv describes a sample of 500 PC sales collected between 1993 and 1995 in the United States. In addition to the sale price, information was collected about processor frequency in MHz, hard disk size in MB, RAM size in MB, and whether the manufacturer is premium (e.g., IBM, COMPAQ).

Assume the regression model $y \sim \mathcal{N}(\mu, \sigma)$, where

$$\mu = \alpha + \beta_1 x_1 + \beta_2 x_2,$$

where y is the sale price, x_1 is the processor frequency, and x_2 is the natural logarithm of the hard disk size.

a) (1p) Using weakly informative prior distributions for the parameters α, β_1, β_2 and σ, use PyMC to simulate a sufficiently large sample from the posterior distribution.
b) (0.5p) Obtain 95% HDI estimates for the parameters β_1 and β_2.
c) (0.5p) Based on the results, are processor frequency and hard disk size useful predictors of the sale price?
d) (0.5p) Now suppose a consumer is interested in a computer with a frequency of 33 MHz and a hard disk of 540 MB. Simulate draws from the expected sale price (μ) and construct a 90% HDI interval for this price.
e) (1p) Instead, suppose this consumer wants to predict the sale price of a computer with this frequency and hard disk size. Simulate draws from the posterior predictive distribution and use these simulated draws to find a 90% HDI prediction interval.

Bonus. (0.5p) Does the fact that the manufacturer is premium affect the price in any way? Justify.