Deep Probabilistic Programming - Week 5

Ahmad Salim Al-Sibahi

August 19, 2020

1 Weather Check Model

FiveThirtyEight did a survey on where and how often people check the weather report[1]. The goal of our model is to predict whether someone checks the weather daily $y \in \{0, 1\}$ based on the other available information $\mathbf{X}$ regarding their app usage, smartwatch usage, age, gender, earnings.

We would like to model this as a Bayesian hierarchical logistic regression problem, where specific region r predictions are pooled together to provide better overall accuracy.

We would like to construct the following model:

$$
\begin{aligned}
\beta_i &\sim \mathcal{N}(\mu = 0., \sigma^2 = 1000.) \qquad i \in [1..M] \\
\mu &\sim \mathcal{N}(\mu = 0., \sigma^2 = 1000.) \\
\tau &\sim \mathcal{N}^+(\sigma^2 = 1000.) \\
\gamma_j &\sim \mathcal{N}(\mu = \mu, \sigma^2 = \tau) \qquad j \in [1..R] \\
y_\ell &\sim \text{Ber}(\text{logit}(p) = \boldsymbol{\beta}\mathbf{X}_\ell + \gamma_{r_\ell})
\end{aligned}
$$

Here, M denotes the number of components for the covariate matrix $\mathbf{X}$, R denotes the number of regions, r denotes the specific region for a data point, y denotes the response variable, D denotes the size of the data, and $\mathcal{N}^+$ denotes the half-normal distribution.

References

[1] Hickey, W. Where people go to check the weather. FiveThirtyEight (Apr 2015).