In-class exercise

Instructions

- **Don't look at the solution yet!** This is for your benefit.
- This exercise must be submitted within 48 hours of the lecture in which it was given.
- As long as you do the exercise on time, you get full credit—your performance does not matter.
- Without looking at the solution, take 5 minutes to try to solve the exercise.
- Pre-assessment: Write down how correct you think your answer is, from 0 to 100%.
- Post-assessment: Now, study the solution and give yourself a "grade" from 0 to 100%.
- Submit your work on the course website, including the pre- and post- assessments.

Exercise

You need to model the probability of recovery for patients admitted to the hospital in severe cardiac distress. (Let's say that "recovery" means the patient survived long enough to be discharged from the hospital.) You have past data from a number of patients from 6 different hospitals. For each patient i, you have various information $x_i = (x_{i1}, \ldots, x_{ip})$ (e.g., gender, weight, age, blood pressure, etc.), and a binary outcome y_i (did the patient recover or not). Design a Bayesian model for this problem.

Solution

There are multiple ways of approaching this, and each will be better or worse depending on the particulars of the situation. One natural way is to use a hierarchical Probit regression model, as follows. Let's divide up the patients according to hospital, and let y_{hi} be the outcome for the ith patient in hospital h, and let $x_{hi} = (x_{hi1}, \ldots, x_{hip})$ be the corresponding information.

$$\begin{aligned}
&Y_{hi}|\theta_h \sim \text{Bernoulli}(\Phi(\theta_h^{\mathsf{T}} x_{hi})) \text{ independently}\\
&\theta_h|\mu, \Sigma \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \Sigma)\\
&\mu \sim \mathcal{N}(\mu_0, \Sigma_0)\\
&\Sigma \sim \text{InverseWishart}(S, \nu).
\end{aligned}$$

This would be appropriate if the effect of the predictors/covariates x was expected to be different for the different hospitals.

Another way would be to augment the covariate vector with indicator variables for the hospital, i.e., for $j = 1, \ldots, 6$, define $x_{h,i,p+j} = \mathbb{1}(j = h)$, and use a Probit regression model with a Normal prior on $\theta = (\theta_1, \ldots, \theta_{p+6})$. This would be appropriate if the effect of the predictors/covariates was expected to be the same for the different hospitals, but the overall recovery rate varied among hospitals.