MVA – Probabilistic Graphical Models
Homework 3: Gibbs Sampling

Wilson Jallet*

December 19, 2019

Q1. This operation puts all the data on the same scale – this is especially useful because the prior on β assigns the same variance in each direction.

Q2. If we supposed that ε_i had a variance of σ^2, we could write $\varepsilon_i = \sigma\varepsilon_i'$ where $\varepsilon_i' \sim \mathcal{N}(0,1)$, and we'd have

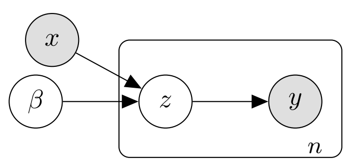
$$y_i = \operatorname{sgn}(\beta^T x_i + \varepsilon_i) = \operatorname{sgn}({\beta'}^T x_i + \varepsilon_i')$$

where $\beta' = \beta/\sigma$.

Q3. We define the following graphical model:

- observed features $x_i \in \mathbb{R}^p$, $i \in \{1,\dots,n\}$
- random variable $\beta \sim \mathcal{N}(0, \tau I_p)$
- latent variables $z_i = \beta^T x_i + \varepsilon_i$, $\varepsilon_i \sim \mathcal{N}(0,1)$
- observed labels $y_i = \operatorname{sgn}(z_i) \in \{-1, 1\}$

This model has the following representation:



We want the posterior distribution of β given y. We will need the conditional posteriors to do Gibbs sampling. Evidently,

$$p(y_i|\beta) = \Phi(y_i \beta^T x_i)$$
$$p(z_i|\beta) \sim \mathcal{N}(\beta^T x_i, 1)$$
$$p(y_i, z_i|\beta) = \mathbb{1}_{\{y_i z_i > 0\}}$$

and by Bayes' theorem we have the posteriors

$$\begin{aligned} p(\beta|z) \propto p(\beta)p(z|\beta) &\propto \exp\left(-\frac{1}{2\tau}\|\beta\|^2 - \frac{1}{2}\sum_{i=1}^n (z_i - \beta^T x_i)^2\right) \\ &= \exp\left(-\frac{1}{2\tau}\|\beta\|^2 - \frac{1}{2}\|z - X\beta\|^2\right) \end{aligned}$$

*wilson.jallet@polytechnique.org

and

$$p(z|\beta, y) \propto p(z|\beta)p(z, y|\beta) \propto \exp\left(-\frac{1}{2}\|z - X\beta\|^2\right) \prod_{i=1}^{n} \mathbb{1}_{\{y_i z_i > 0\}}$$

where $X = (x_1| \dots |x_n)^T \in \mathbb{R}^{n \times p}$ is the design matrix. By identification $\beta|z \sim \mathcal{N}(\mu, \Sigma)$ where

$$\Sigma^{-1} = \frac{1}{\tau} I_p + X^T X, \quad \mu = \Sigma X^T z$$

and $z|\beta, y \sim \mathrm{T}\mathcal{N}(X\beta, I_n; \mathcal{P}_y)$ where $\mathrm{T}\mathcal{N}(\cdot; \mathcal{P}_y)$ is the truncated Gaussian with support in the polytope $\mathcal{P}_y = \{z \in \mathbb{R}^n : z_i y_i > 0,\ i = 1, \dots, n\}$.