Probabilistic Graphical Models

Homework 2

Author

Bastien Déchamps
bastien.dechamps@eleves.enpc.fr

Supervisors

Pierre Latouche

and

Nicolas Chopin

Gibbs Sampling and mean field VB for the probit model

Question 1. We chose an centered isotropic Gaussian distribution as the prior for β, so the input data needs to be normalized.

Question 2. If ϵ_i variance is σ^2, we could write:

$$y_i = \operatorname{sgn}\left(\beta^\top x_i + \epsilon_i\right) = \operatorname{sgn}\left(\frac{1}{\sigma}\beta^\top x_i + \frac{\epsilon_i}{\sigma}\right)$$

that reduces to the case where ϵ_i has a variance of one.

Question 3. We denote $z_i = \beta^\top x_i + \epsilon_i \sim \mathcal{N}(\beta^\top x_i, 1)$, and $y_i = \operatorname{sgn}(z_i)$. We want to estimate the posterior distribution of β, z given X, y by using Gibbs sampling. To this extent, we compute the posteriors $p(z|\beta, y)$ and $p(\beta|z)$.

By using Bayes formula, we have:

$$\begin{aligned}
p\left(\beta|z\right) &\propto p(\beta)p(z|\beta) \\
&\propto \mathcal{N}(\beta; 0, \tau I_p) \prod_{i=1}^{N} \mathcal{N}(z_i; \beta^\top x_i, 1) \\
&\propto \exp\left(-\frac{1}{2\tau}\|\beta\|^2 - \frac{1}{2}\sum_{i=1}^{N}(z_i - \beta^\top x_i)^2\right) \\
&\propto \exp\left(-\frac{1}{2\tau}\|\beta\|^2 - \frac{1}{2}\left\|z - X\beta\right\|^2\right)
\end{aligned}$$

Thus, $\beta|z \sim \mathcal{N}(\mu_n, \Sigma_n)$ where

$$\Sigma_n = \left(\frac{1}{\tau}I_p + X^\top X\right)^{-1}, \quad \mu_n = \Sigma_n X^\top z \tag{1}$$

Then, we compute $p(z_i|\beta, y_i)$ by using Bayes formula:

$$\begin{aligned}
p(z_i|\beta, y_i) &\propto p(z_i, \beta)p(y_i, z|\beta) \\
&\propto \exp\left(-\frac{1}{2}(z_i - \beta^\top x_i)^2\right) 1_{y_i z_i > 0} \\
&= f\left(z_i; \beta^\top x_i, 1, y_i\right)
\end{aligned}$$

where $f\left(z_i; \beta^\top x_i, 1, y_i\right)$ is the truncated Gaussian density for $z_i \in \{z \in \mathbb{R} \mid zy_i > 0\}$.

An implementation of the Gibbs sampling algorithm can be found in the file `gibbs.py`. By applying it on the German credit dataset, we compute the approximated marginal distribution of each parameters. Some examples are represented on figure 1. Plots for other parameters can be reproduced with the notebook `visualization.ipynb`.

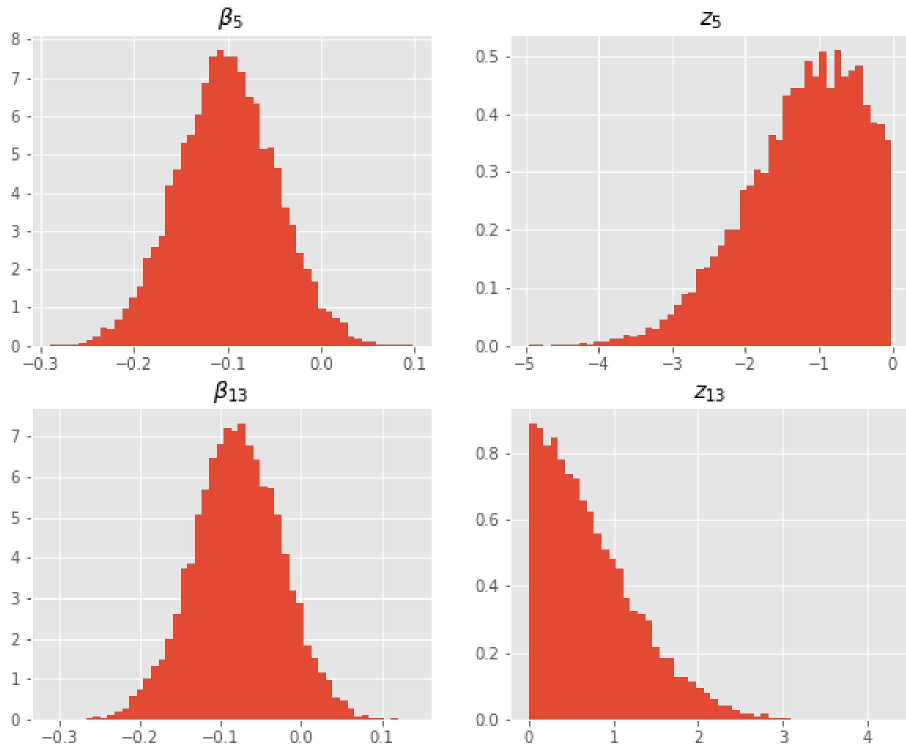

Figure 1: Some marginal distribution obtained with Gibbs sampling

Question 4. We now assume that $p(\beta, z|y)$ can be approximated by a simpler distribution $q(\beta, z) = q_\beta(\beta)q_z(z)$ (mean-field approximation). The optimal distribution q^* statisfies:

$$\begin{cases} \log q^*_\beta(\beta) = \mathbb{E}_z\left[\log p(\beta, z, y)|\beta, y\right] + \text{cst} \\ \log q^*_z(\beta) = \mathbb{E}_\beta\left[\log p(\beta, z, y)|\beta, y\right] + \text{cst} \end{cases} \tag{2}$$

Let's compute the optimal form for q_β:

$$\begin{aligned}
\log q^*_\beta(\beta) &= \mathbb{E}_z \left[\log p(y|z) + \log p(z|\beta) + \log p(\beta) \middle| \beta, y\right] + \text{cst} \\
&= \underbrace{\mathbb{E}_z \left[\log p(y|z) \middle| \beta, y\right]}_{\text{independent of } \beta} - \frac{1}{2}\mathbb{E}_z \left[\|z - X\beta\|^2 \middle| \beta, y\right] - \frac{1}{2\tau}\|\beta\|^2 + \text{cst} \\
&= -\frac{1}{2}\underbrace{\mathbb{E}_z \left[\|z\|^2 \middle| \beta, y\right]}_{\text{independent of } \beta} + \mathbb{E}_z \left[z^\top X\beta + \|X\beta\|^2 \middle| \beta, y\right] - \frac{1}{2\tau}\|\beta\|^2 + \text{cst} \\
&= \bar{z}^\top X\beta - \frac{1}{2}\beta^\top X^\top X\beta - \frac{1}{2\tau}\|\beta\|^2 + \text{cst} \\
&= -\frac{1}{2}(\beta - \bar{\beta})\Sigma_n^{-1}(\beta - \bar{\beta}) + \text{cst}
\end{aligned}$$

where

$$\Sigma_n = \left(\frac{1}{\tau}I_p + X^\top X\right)^{-1}, \quad \bar{z} = \mathbb{E}_z[z], \quad \bar{\beta} = \Sigma_n X^\top \bar{z}.$$

Thus, by normalizing, q^*_β has the following form:

$$q^*_\beta(\beta) = \mathcal{N}(\beta; \bar{\beta}, \Sigma_n) \tag{3}$$

Now, let's compute the optimal form for q_z:

$$\begin{aligned}
\log q^*_z(z) &= \mathbb{E}_\beta \left[\log p(y|z) - \frac{1}{2}\|z - X\beta\|^2 - \frac{1}{2\tau}\|\beta\|^2 \middle| z, y\right] + \text{cst} \\
&= \sum_{i=1}^{N} \left(\mathbb{1}_{y_i=1} \log \mathbb{1}_{z_i>0} + \mathbb{1}_{y_i=-1} \log \mathbb{1}_{z_i<0}\right) - \frac{1}{2}\left\|z - X\bar{\beta}\right\|^2 + \text{cst}
\end{aligned}$$

where $\bar{\beta} = \mathbb{E}_\beta[\beta] = \Sigma_n X^\top \bar{z}$. This is the form of a truncated Gaussian defined on $\left\{z \in \mathbb{R}^N \mid z_i y_i > 0, \forall i\right\}$ and we finally have:

$$q^*_z(z) = f\left(x; X\bar{\beta}, I_p, y\right) \tag{4}$$

We can now compute $\bar{z}$ which is given by:

$$\bar{z}_i = \bar{\beta}^\top x_i + y_i \frac{\phi\left(\bar{\beta}^\top x_i\right)}{\Phi\left(y_i \bar{\beta}^\top x_i\right)} \tag{5}$$

where ϕ (resp. Φ) is the density (resp. the cumulative function) of $\mathcal{N}(0, 1)$.

An implementation of mean-field algorithm can be found in the file `meanfield.py`. The convergence of the parameters are shown on figure 2 and some marginalization examples on figure 3. In terms of speed of convergence, Gibbs sampling took **8 seconds** to sample 10000 samples for a train accuracy of **0.787** while the mean-field algorithm took **0.5 second** to train for 250 iterations and sample 10000 samples, for a train accuracy of **0.785**.

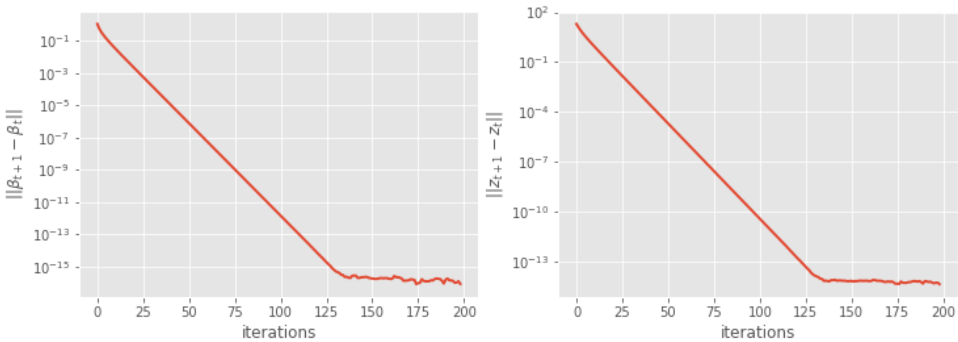

Figure 2: Convergence of $\bar{\beta}$ and $\bar{z}$.

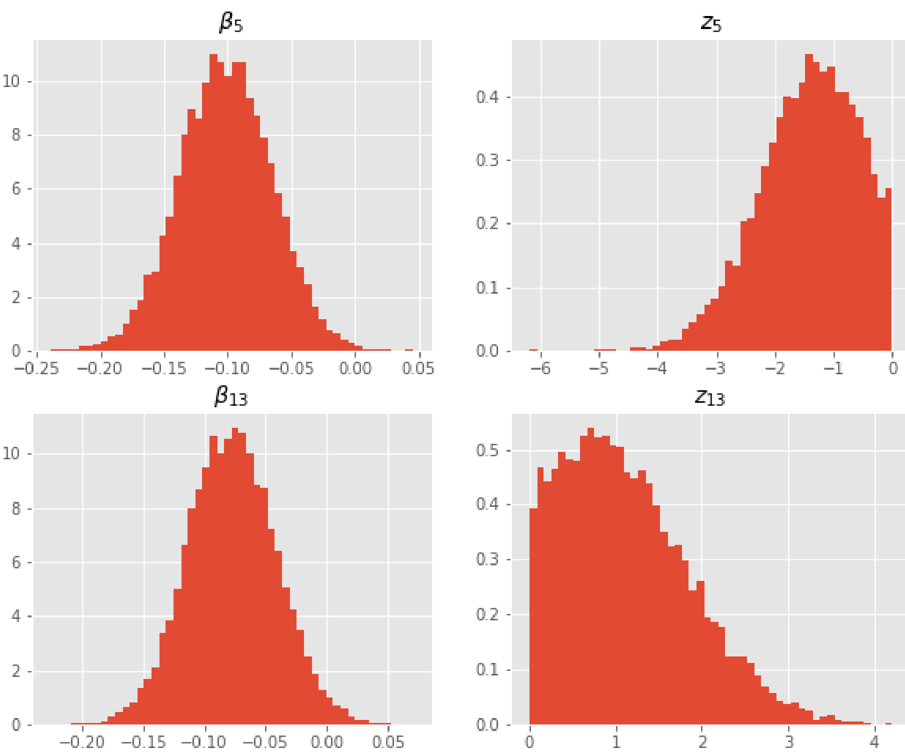

Figure 3: Some marginal distributions obtained with mean-field algorithm.