

SBM Laplace prior for concentration matrix

March 14, 2019

1 Model

Data: X , $i \in \{0, \dots, n\}$ observations. $X \sim \mathcal{N}(0, \Sigma^{-1} = \Omega)$, Ω concentration matrix. Assumption: Q hidden clusters in the data. For each observation i , $Z_{iq} = \mathbb{1}_{i \in q}$, $q \in \{1, \dots, Q\}$. Z_{iq} = independant latent variables.

- $\alpha_q = \mathbb{P}(i \in q)$ (Probability that observation i belongs to cluster q);
- $\sum_q \alpha_q = 1$;
- $\underline{Z}_i \sim \mathcal{M}(1, \underline{\alpha})$ with $\underline{\alpha} = (\alpha_1, \dots, \alpha_Q)$ and $\underline{Z}_i = (Z_{i1}, \dots, Z_{iQ})$

Assumption:

$$\Omega_{ij} | \{Z_{iq} Z_{jl}\} \sim \text{Laplace}(0, \lambda_{ql}), \quad (i, j) \in \{1, \dots, n\}^2, (q, l) \in \{1, \dots, Q\}^2$$

Laplace distribution:

$$\forall x \in \mathbb{R}, \quad f_{ql}(x) = \frac{1}{\lambda_{ql}} \exp - \frac{|x|}{2\lambda_{ql}} \quad \text{if } q \neq l \quad \text{and} \quad f_0(x) = \frac{1}{\lambda_0} \exp - \frac{|x|}{2\lambda_0} \quad \text{otherwise}$$

where $\lambda_{ql}, \lambda_0 > 0$ are scaling parameters and $\lambda_{ql} = \Lambda_{ql}$. Λ matrix of parameters λ_{ql} , $(q, l) \in \{1, \dots, Q\}^2$.

2 Complete likelihood

Assumptions reminder:

$$X | \Omega \sim \mathcal{N}(0, \Omega^{-1}), \quad \Omega | Z \sim \text{Laplace}(0, \Lambda), \quad Z \sim \mathcal{M}(1, \underline{\alpha})$$

Complete likelihood decomposition:

$$\begin{aligned} L_c(X, \Omega, Z) &= \mathbb{P}(X, \Omega, Z) \\ &= \mathbb{P}(X | \{\Omega, Z\}) \mathbb{P}(\Omega | Z) \mathbb{P}(Z) \quad \text{conditional probabilities definition/formula} \\ &= \mathbb{P}(X | \Omega) \mathbb{P}(\Omega | Z) \mathbb{P}(Z) \end{aligned}$$

$\mathbb{P}(X | \{\Omega, Z\}) = \mathbb{P}(X | \Omega)$, known distribution (knowing Z is equivalent to know Ω ?).

Complete log-likelihood formula: (when $i \neq j$, otherwise replace λ_{ql} with λ_0)

$$\begin{aligned} \log L_c(X, \Omega, Z) &= \log \mathbb{P}(X, \Omega, Z) \\ &= \underbrace{\log \mathbb{P}(X | \Omega)}_{(1)} + \underbrace{\log \mathbb{P}(\Omega | Z)}_{(2)} + \underbrace{\log \mathbb{P}(Z)}_{(3)} \\ &= \frac{n}{2} (\log(|\Omega|) - \text{tr}(S\Omega) - p \log(2\pi)) + \sum_{q,l,i,j,i \neq j} Z_{iq} Z_{jl} \left(-\log(2\lambda_{ql}) - \frac{|\Omega_{ij}|}{\lambda_{ql}} \right) \\ &\quad + \sum_{i=1}^n \sum_{q=1}^Q Z_{iq} \log \alpha_q \end{aligned}$$

Penalty term (lasso approach) in the article: $\sum_{q,l,i,j,i \neq j} Z_{iq} Z_{jl} \frac{|\Omega_{ij}|}{\lambda_{ql}}$. In the paper: notation $\|\rho_Z(\Omega)\|_{l_1}$.

(1): $X|\Omega \sim \mathcal{N}(0, \Omega^{-1})$

Notation: $|\Omega| = \det(\Omega) = 1/|\Omega^{-1}|$

$$\begin{aligned}
\log \mathbb{P}(X|\Omega) &= \sum_{i=1}^n \log \left(\frac{1}{(2\pi)^{\frac{p}{2}} |\Omega^{-1}|} \exp\left(\frac{1}{2} x_i^T \Omega x_i\right) \right) \\
&= -\frac{n}{2} (p \log(2\pi) + \log(|\Omega^{-1}|)) - \frac{1}{2} \text{tr}(X^T X \Omega) \\
&= -\frac{n}{2} (p \log(2\pi) - \log(|\Omega|)) - \frac{n}{2} \text{tr}(S \Omega) \text{ En notant } S = \frac{1}{n} X^T X \\
&= \frac{n}{2} (-p \log(2\pi) + \log(|\Omega|)) - \frac{n}{2} \text{tr}(S \Omega) \\
&= \frac{n}{2} (\log(|\Omega|) - \text{tr}(S \Omega) - p \log(2\pi))
\end{aligned}$$

(2): $\Omega|Z \sim \text{Laplace}(0, \Lambda)$

If $i \neq j$:

$$\begin{aligned}
\log \mathbb{P}(\Omega|Z) &= \log \prod_{i,j,q,l} \mathbb{P}(\Omega_{ij} | \{Z_{iq} Z_{jl}\})^{Z_{iq} Z_{jl}} \text{ Astuce termes puissance 0 ne participent pas au produit} \\
&= \sum_{\substack{q,l,i,j \\ i \neq j}} Z_{iq} Z_{jl} \left(-\log(2\lambda_{ql}) - \frac{|\Omega_{ij}|}{\lambda_{ql}} \right) \\
&= - \underbrace{\sum_{\substack{q,l,i,j \\ i \neq j}} Z_{iq} Z_{jl} \log(2\lambda_{ql})}_{\sum_{q,l} n_{ql} \log(2\lambda_{ql})} - \underbrace{\sum_{\substack{q,l,i,j \\ i \neq j}} Z_{iq} Z_{jl} \frac{|\Omega_{ij}|}{\lambda_{ql}}}_{\|\rho_Z(\Omega)\|_1}
\end{aligned}$$

Otherwise:

$$\log \mathbb{P}(\Omega|Z) = \sum_i \left(-\log(2\lambda_0) - \frac{|\Omega_{ii}|}{\lambda_0} \right)$$

(3): $Z \sim \mathcal{M}(1, \underline{\alpha})$

$$\begin{aligned}
\log \mathbb{P}(Z) &= \log \prod_{i=1}^n \mathbb{P}(Z_i) = \log \prod_{i=1}^n \prod_{q=1}^Q \alpha_q^{Z_{iq}} \\
&= \sum_{i=1}^n \sum_{q=1}^Q \log \alpha_q^{Z_{iq}} = \sum_{i=1}^n \sum_{q=1}^Q Z_{iq} \log \alpha_q
\end{aligned}$$

3 EM algorithm and variational estimation

Steps

1. Expectation (E): Calculate the expected value of the likelihood under the current parameters. Estimate Z_i , knowing Ω from previous step.
2. Maximization (M): Find parameters that maximizes the likelihood. Compute parameters knowing Z_i .

3.1 Definitions

Let \mathcal{Z} the space of all possibilities.

Definitions give us:

$$\log \mathbb{P}(Z|X, \Omega) = \log \mathbb{P}(X, \Omega, Z) - \log \mathbb{P}(X, \Omega)$$

3.2 E step: impossibility of direct computation and variational approach

In this part, Ω is known (as $\Omega^{(t)}$) from a previous M step.

Considering $\mathbb{P}(X, \Omega)$ being constant with respect to Z :

$$\begin{aligned}\log \mathbb{P}(X, \Omega) &= \log \mathbb{P}(X, \Omega, Z) - \log \mathbb{P}(Z|X, \Omega) \\ &= \mathbb{E}_{Z|\Omega^{(t)}}[\log \mathbb{P}(X, \Omega, Z)] - \mathbb{E}_{Z|\Omega^{(t)}}[\log \mathbb{P}(Z|X, \Omega)] \\ &= \sum_{z \in \mathcal{Z}} \mathbb{P}(Z = z|\Omega^{(t)}) \log \mathbb{P}(X, \Omega, Z) - \sum_{z \in \mathcal{Z}} \mathbb{P}(Z = z|\Omega^{(t)}) \log \mathbb{P}(Z|X, \Omega) \\ &= \mathbb{E}_{Z|\Omega^{(t)}}[\log \mathbb{P}(X, \Omega, Z)] - \mathcal{H}(Z|X)\end{aligned}$$

The expected complete likelihood under the current parameters is:

$$\mathcal{Q}(\Omega|\Omega^{(t)}) = \mathbb{E}_{Z|\Omega^{(t)}}[\log \mathbb{P}(X, \Omega, Z)] = \sum_{z \in \mathcal{Z}} \mathbb{P}(Z = z|\Omega^{(t)}) \log \mathbb{P}(X, \Omega, Z)$$

$\mathbb{P}(Z_{iq}Z_{jl} = 1|\Omega_{ij}^{(t)})$ is unknown as Z_{iq} and Z_{jl} are not independant. Variational approach: take an approximation for $\mathbb{P}(Z|\Omega^{(t)}) := R_t(Z)$ for E step.

Now we consider the lower bound \mathcal{J} of $\mathbb{P}(X, \Omega)$:

$$\mathcal{J}_\tau(X, \Omega, R(Z)) := \log \mathbb{P}(X, \Omega) - D_{KL}\{R(Z)||\mathbb{P}(Z|\Omega)\},$$

where $D_{KL}\{R(Z)||\mathbb{P}(Z|\Omega)\}$ is the Küllback-Leibler divergence (i.e. distance between these two distributions). We take the following distribution for $R(\cdot)$ ¹

$$R_\tau(Z) = \prod_{i=1}^n h_{\tau_i}(Z_i),$$

where h_{τ_i} is the density of the multinomial probability distribution $\mathcal{M}(1, \tau_i)$ and $\tau_i = (\tau_{i1}, \dots, \tau_{iQ})$ is a random vector containing the parmeters to optimize in the variational approach. Approximation of the probability that vertex i belongs to cluster q , τ_{iq} estimates $\mathbb{P}(Z_{iq} = 1|\Omega)$, under the constraint $\sum_q \tau_{iq} = 1$.

Küllback-Leibler divergence:

$$\begin{aligned}D_{KL}\{R_\tau(Z)||\mathbb{P}(Z|\Omega)\} &= \sum_{Z \in \mathcal{Z}} R_\tau(Z) \log \frac{R_\tau(Z)}{\mathbb{P}(Z|\Omega)} \\ &= \sum_{Z \in \mathcal{Z}} R_\tau(Z) (\log R_\tau(Z) - \log \mathbb{P}(Z|\Omega)) \\ &= -\mathcal{H}(R_\tau(Z)) - \sum_{Z \in \mathcal{Z}} R_\tau(Z) \log \mathbb{P}(Z|\Omega)\end{aligned}$$

New formula for the bound²:

$$\begin{aligned}\mathcal{J}_\tau(X, \Omega, R(Z)) &= \log \mathbb{P}(X, \Omega) - D_{KL}\{R(Z)||\mathbb{P}(Z|\Omega)\} \\ &= \log \mathbb{P}(X, \Omega) + \mathcal{H}(R_\tau(Z)) + \sum_{Z \in \mathcal{Z}} R_\tau(Z) \log \mathbb{P}(Z|\Omega) \\ &= \mathcal{H}(R_\tau(Z)) + \sum_{Z \in \mathcal{Z}} R_\tau(Z) (\log \mathbb{P}(Z|\Omega) + \log \mathbb{P}(X, \Omega)) \\ &= \mathcal{H}(R_\tau(Z)) + \underbrace{\sum_{Z \in \mathcal{Z}} R_\tau(Z) \log L_c(X, \Omega, Z)}_{\hat{Q}_\tau(\Omega) = \mathbb{E}_{\mathcal{R}_\tau}[\log \mathbb{P}(X, \Omega, Z)]}\end{aligned}$$

Objective: maximization of $\mathcal{J}(X, \Omega, R(Z))$.

¹cf. Mariadassous and Robin, Uncovering latent structure in valued graphs: a variational approach, Technical Report 10, Statistics for Systems Biology, 2007

²En considérant que $\mathbb{P}(Z|\Omega) = \mathbb{P}(Z|\Omega, X)$ et en remarquant que $\log \mathbb{P}(X, \Omega)$ est indep de Z

Expression for $\hat{Q}_\tau(\Omega)$ Assume $R_\tau(Z) = \prod_{i=1}^n h_{\tau_i}(Z_i)$

$$\begin{aligned}\hat{Q}_\tau(\Omega) &= \sum_{Z \in \mathcal{Z}} R_\tau(Z) \log L_c(X, \Omega, Z) = \mathbb{E}_{R_\tau}[\log L_c(X, \Omega, Z)] \\ &= \frac{n}{2} (\log(|\Omega|) - \text{tr}(S\Omega) - p \log(2\pi)) + \sum_{i,j=1, j \neq i}^n \sum_{q,l=1}^Q \tau_{iq} \tau_{jl} \left(-\log(2\lambda_{ql}) - \frac{|\Omega_{ij}|}{\lambda_{ql}} \right) + \sum_{i=1}^n \sum_{q=1}^Q \tau_{iq} \log \alpha_q\end{aligned}$$

Expression for $\mathcal{H}(R_\tau(Z))$ Assume $R_\tau(Z) = \prod_{i=1}^n h_{\tau_i}(Z_i)$

$$\begin{aligned}\mathcal{H}(R_\tau(Z)) &= - \sum_{i=1}^n \sum_{q=1}^Q \mathbb{P}(Z_{iq} = 1) \log \mathbb{P}(Z_{iq} = 1) \\ &= - \sum_{i=1}^n \sum_{q=1}^Q \tau_{iq} \log \tau_{iq}\end{aligned}$$

Complete expression of $\mathcal{J}_\tau(X, \Omega, R(Z))$

$$\begin{aligned}\mathcal{J}_\tau(X, \Omega, R(Z)) &= \mathcal{H}(R_\tau(Z)) + \hat{Q}_\tau(\Omega) \\ &= \underbrace{\frac{n}{2} (\log(|\Omega|) - \text{tr}(S\Omega) - p \log(2\pi))}_{\text{Constant par rapport à } Z, \text{ et autres, désigné par } c \text{ dans l'article}} \\ &\quad + \sum_{\substack{i,j=1 \\ j \neq i}}^n \sum_{q,l=1}^Q \tau_{iq} \tau_{jl} \left(-\log(2\lambda_{ql}) - \frac{|\Omega_{ij}|}{\lambda_{ql}} \right) + \sum_{i=1}^n \sum_{q=1}^Q \tau_{iq} \log \alpha_q - \sum_{i=1}^n \sum_{q=1}^Q \tau_{iq} \log \tau_{iq}\end{aligned}$$

Similar to the one in the complete likelihood expression, the penalty term is: $\sum_{i,j=1, j \neq i}^n \sum_{q,l=1}^Q \tau_{iq} \tau_{jl} \frac{|\Omega_{ij}|}{\lambda_{ql}}$.

3.3 E step: parameters estimation

Strategy: estimate τ_{iq} with fixed α_q and λ_{lq} , then estimate α_q and λ_{lq} considering $\hat{\tau}_{iq}$.

Estimating $\hat{\tau}_{iq}$: Introducing the constraint using Lagrange multiplier method. Constraint: $\sum_q \tau_{iq} = 1$.

$$\begin{aligned}\frac{\partial \mathcal{J}_\tau(X, \Omega, Z) - \lambda(\sum_{q=1}^Q \tau_{iq} - 1)}{\partial \tau_{iq}} &= \sum_{\substack{j=1 \\ j \neq i}}^n \sum_{l=1}^Q \tau_{jl} \left(-\log(2\lambda_{ql}) - \frac{|\Omega_{ij}|}{\lambda_{ql}} \right) + \log \alpha_q - 1 - \log \tau_{iq} - \lambda \\ \tau_{iq} &= \exp \left[\sum_{\substack{j=1 \\ j \neq i}}^n \sum_{l=1}^Q \tau_{jl} \left(-\log(2\lambda_{ql}) - \frac{|\Omega_{ij}|}{\lambda_{ql}} \right) + \log \alpha_q - 1 - \lambda \right] = \exp(-1) \exp(-\lambda) \alpha_q \exp \left[\sum_{\substack{j=1 \\ j \neq i}}^n \sum_{l=1}^Q \tau_{jl} \left(-\log(2\lambda_{ql}) - \frac{|\Omega_{ij}|}{\lambda_{ql}} \right) \right] \\ &= \exp(-1) \exp(-\lambda) \alpha_q \prod_{\substack{j=1 \\ j \neq i}}^n \prod_{l=1}^Q \exp \left[\tau_{jl} \left(-\log(2\lambda_{ql}) - \frac{|\Omega_{ij}|}{\lambda_{ql}} \right) \right] = \exp(-1) \exp(-\lambda) \alpha_q \prod_{\substack{j=1 \\ j \neq i}}^n \prod_{l=1}^Q \left(\frac{1}{2\lambda_{ql}} \exp \left(\frac{-|\Omega_{ij}|}{\lambda_{ql}} \right) \right)^{\tau_{jl}}\end{aligned}$$

Using the constraint to find λ :

$$\sum_{q=1}^Q \tau_{iq} = 1 = \exp(-1) \exp(-\lambda) \alpha_q \prod_{\substack{j=1 \\ j \neq i}}^n \prod_{l=1}^Q \left(\frac{1}{2\lambda_{ql}} \exp\left(\frac{-|\Omega_{ij}|}{\lambda_{ql}}\right) \right)^{\tau_{jl}}$$

$$\Rightarrow \exp(\lambda) = \left[\sum_{q=1}^Q \exp(-1) \alpha_q \prod_{\substack{j=1 \\ j \neq i}}^n \prod_{l=1}^Q \left(\frac{1}{2\lambda_{ql}} \exp\left(\frac{-|\Omega_{ij}|}{\lambda_{ql}}\right) \right)^{\tau_{jl}} \right]^{-1}$$

$$\hat{\tau}_{iq} = \frac{\exp(-1) \alpha_q \prod_{\substack{j=1 \\ j \neq i}}^n \prod_{l=1}^Q \left(\frac{1}{2\lambda_{ql}} \exp\left(\frac{-|\Omega_{ij}|}{\lambda_{ql}}\right) \right)^{\tau_{jl}}}{\sum_{q=1}^Q \exp(-1) \alpha_q \prod_{\substack{j=1 \\ j \neq i}}^n \prod_{l=1}^Q \left(\frac{1}{2\lambda_{ql}} \exp\left(\frac{-|\Omega_{ij}|}{\lambda_{ql}}\right) \right)^{\tau_{jl}}}$$

In the paper:

$$\hat{\tau}_{iq} \propto \alpha_q \prod_{\substack{j=1 \\ j \neq i}}^n \prod_{l=1}^Q \left(\frac{1}{2\lambda_{ql}} \exp\left(\frac{-|\Omega_{ij}|}{\lambda_{ql}}\right) \right)^{\tau_{jl}}$$

Estimating $\hat{\alpha}_q$: With constraint $\sum_q \alpha_q = 1$.

$$\left. \frac{\partial \mathcal{J}_\tau(X, \Omega, Z)}{\partial \alpha_q} \right|_{\tau_{iq}} = \sum_{i=1}^n \sum_{q=1}^Q \tau_{iq} \frac{1}{\alpha_q}$$

The value if α_q will be derived from the following expression:

$$\left. \frac{\partial \mathcal{J}_\tau(X, \Omega, Z) - \eta(\sum_q \alpha_q - 1)}{\partial \alpha_q} \right|_{\tau_{iq}} = \sum_{i=1}^n \tau_{iq} \frac{1}{\alpha_q} - \eta$$

$$0 = \sum_{i=1}^n \tau_{iq} \frac{1}{\alpha_q} - \eta \Leftrightarrow \alpha_q = \frac{\sum_i \tau_{iq}}{\eta}$$

$$\Leftrightarrow \sum_q \alpha_q = \frac{\sum_q \sum_i \tau_{iq}}{\eta} \Leftrightarrow 1 = \frac{n}{\eta} \Leftrightarrow \boxed{\eta = n}$$

$$0 = \sum_{i=1}^n \tau_{iq} \frac{1}{\alpha_q} - \eta \Leftrightarrow \alpha_q = \frac{\sum_i \tau_{iq}}{\eta}$$

$$\hat{\alpha}_q = \frac{\sum_i \tau_{iq}}{n}$$

Estimating $\hat{\lambda}_{lq}$:

$$\left. \frac{\partial \mathcal{J}_\tau(X, \Omega, Z)}{\partial \lambda_{ql}} \right|_{\tau_{iq}} = \sum_{\substack{j=1 \\ j \neq i}}^n \tau_{iq} \tau_{jl} \left(-\frac{1}{2\lambda_{ql}} + \frac{|\Omega_{ij}|}{\lambda_{ql}^2} \right)$$

$$\begin{aligned}
0 &= \sum_{\substack{j=1 \\ j \neq i}}^n \tau_{iq} \tau_{jl} \left(-\frac{1}{2\lambda_{ql}} + \frac{|\Omega_{ij}|}{\lambda_{ql}^2} \right) \\
\Rightarrow 0 &= \sum_{\substack{j=1 \\ j \neq i}}^n \tau_{iq} \tau_{jl} \left(-\frac{1}{2} + \frac{|\Omega_{ij}|}{\lambda_{ql}} \right)
\end{aligned}$$

$$\hat{\lambda}_{ql} = \frac{\sum_{\substack{j=1 \\ j \neq i}}^n \tau_{iq} \tau_{jl} |\Omega_{ij}|}{\sum_{\substack{j=1 \\ j \neq i}}^n \tau_{iq} \tau_{jl}}$$