Mean-field variational inference

Simon Leglaive

November 13, 2020

L'inférence variationnelle consiste à chercher une distribution variationnelle sur les variables latentes, de densité de probabilité $q \in \mathcal{F}$, qui minimise la divergence de Kullback-Leibler

$$D_{\mathrm{KL}}(q(\mathbf{z})||p(\mathbf{z}|\mathbf{x};\boldsymbol{\theta})), \tag{1}$$

ou de façon équivalent qui maximise l'énergie variationnelle libre

$$\mathcal{L}(q;\boldsymbol{\theta}) = \mathbb{E}_q[\ln p(\mathbf{x},\mathbf{z};\boldsymbol{\theta})]. \tag{2}$$

Pour résoudre ce problème nous devons en général formuler certaines hypothèses sur la famille variationnelle $\mathcal{F}$. L'approximation de champ moyen est pour cela très répandue, elle consiste à supposer que la famille variationnelle correspond à l'ensemble des densités de probabilités qui se factorisent sous la forme suivante :

$$q(\mathbf{z}) = \prod_{i=1}^{M} q_i(z_i), \tag{3}$$

où l'ensemble des variables latentes est supposé pouvoir se factoriser en M partitions z_i, $i = 1, ..., M$.

En injectant (3) dans (2) on a:

$$\begin{aligned}
\mathcal{L}(q;\boldsymbol{\theta}) &= \int \prod_{i=1}^{M} q_i(z_i) \ln\left(\frac{p(\mathbf{x},\mathbf{z};\boldsymbol{\theta})}{\prod_{i=1}^{M} q_i(z_i)}\right) d\mathbf{z} \\
&= \int \prod_{i=1}^{M} q_i(z_i) \Big[\ln p(\mathbf{x},\mathbf{z};\boldsymbol{\theta}) - \sum_{i=1}^{M} \ln q_i(z_i) \Big] d\mathbf{z} \\
&= \int \prod_{i=1}^{M} q_i(z_i) \ln p(\mathbf{x},\mathbf{z};\boldsymbol{\theta}) d\mathbf{z} - \sum_{i=1}^{M} \int \prod_{k=1}^{M} q_k(z_k) \ln q_i(z_i) d\mathbf{z} \\
&= \int \prod_{i=1}^{M} q_i(z_i) \ln p(\mathbf{x},\mathbf{z};\boldsymbol{\theta}) d\mathbf{z} - \sum_{i=1}^{M} \int q_i(z_i) \ln q_i(z_i) dz_i \qquad \text{car} \int q_k(z_k) dz_k = 1 \\
&= \int q_j(z_j) \left[\int \ln p(\mathbf{x},\mathbf{z};\boldsymbol{\theta}) \prod_{i \neq j} q_i(z_i) dz_i \right] dz_j - \int q_j(z_j) \ln q_j(z_j) dz_j \\
&\qquad - \sum_{i \neq j} \int q_i(z_i) \ln q_i(z_i) dz_i.
\end{aligned} \tag{4}$$

On pose :

$$\ln \tilde{p}(\mathbf{x}, z_j; \boldsymbol{\theta}) = \int \ln p(\mathbf{x}, \mathbf{z}; \boldsymbol{\theta}) \prod_{i \neq j} q_i(z_i) dz_i = \langle \ln p(\mathbf{x}, \mathbf{z}; \boldsymbol{\theta}) \rangle_{\prod_{i \neq j} q_i}. \tag{5}$$

Alors,

$$\begin{aligned}
\mathcal{L}(q; \boldsymbol{\theta}) &= \int q_j(z_j) \ln \tilde{p}(\mathbf{x}, z_j; \boldsymbol{\theta}) dz_j - \int q_j(z_j) \ln q_j(z_j) dz_j - \sum_{i \neq j} \int q_i(z_i) \ln q_i(z_i) dz_i \\
&= \int q_j(z_j) \ln \frac{\tilde{p}(\mathbf{x}, z_j; \boldsymbol{\theta})}{q_j(z_j)} dz_j - \sum_{i \neq j} \int q_i(z_i) \ln q_i(z_i) dz_i \\
&= -D_{\mathrm{KL}}(q_j(z_j) || \tilde{p}(\mathbf{x}, z_j; \boldsymbol{\theta})) - \sum_{i \neq j} \int q_i(z_i) \ln q_i(z_i) dz_i. \qquad (6)
\end{aligned}$$

On suppose maintenant que les $\{q_i\}_{i \neq j}$ sont fixes et on cherche à maximiser $\mathcal{L}(q; \boldsymbol{\theta})$ par rapport à $q_j(z_j)$. D'après l'équation (6) on voit que cela est équivalent à minimiser la divergence de Kullback-Leibler $D_{\mathrm{KL}}(q_j(z_j) || \tilde{p}(\mathbf{x}, z_j; \boldsymbol{\theta}))$, qui est minimale quand $q_j(z_j) = \tilde{p}(\mathbf{x}, z_j; \boldsymbol{\theta})$. La distribution optimale $q_j^\star(z_j)$ doit donc satisfaire :

$$\ln q_j^\star(z_j) = \langle \ln p(\mathbf{x}, \mathbf{z}; \boldsymbol{\theta}) \rangle_{\prod_{i \neq j} q_i} + \text{constante}, \tag{7}$$

où la constante peut être déterminée par normalisation ou en reconnaissant une certaine forme de loi de probabilité.

Les équations (7) pour $j = 1, ..., M$ correspondent à un ensemble de conditions de consistance que doivent satisfaire les distributions $q_j^\star(z_j)$ qui maximisent l'énergie libre. Elles définissent un ensemble de solutions couplées car $q_j^\star(z_j)$ dépend des autres facteurs $q_i(z_i)$ pour $i \neq j$. Une solution consistante est alors trouvée en cyclant sur ces facteurs $q_j^\star(z_j)$ pour $j = 1, ..., M$ et en utilisant à chaque cycle les estimées précédentes.