ΣGS node: Derivations

Bart van Erp

June 7, 2021

1 Overview

The node factor is given by

$$p(x \mid s,\ n) = \mathcal{N}_{\mathcal{C}}(x \mid 0, e^s + e^n, 0), \tag{1}$$

where

$$\mathcal{N}_{\mathcal{C}}(x \mid \mu, \sigma^2, 0) = \frac{1}{\pi\sigma^2} e^{-\frac{1}{\sigma^2}|x-\mu|^2}, \tag{2}$$

so

$$p(x \mid s,\ n) = \frac{1}{\pi(e^s + e^n)} e^{-\frac{1}{e^s+e^n}|x|^2} \tag{3}$$

The approximate posterior is given by

$$q(x,\ s,\ n) = q(x)q(s)q(n), \tag{4}$$

where

$$q(x) = \mathcal{N}_{\mathcal{C}}(x \mid m_x, v_x, 0) \tag{5}$$
$$q(s) = \mathcal{N}(s \mid m_s, v_s) \tag{6}$$
$$q(n) = \mathcal{N}(n \mid m_n, v_n) \tag{7}$$

The log-pdf can be found as

$$\ln p(x \mid s, n) = -\ln(e^s + e^n) - \frac{1}{e^s + e^n}|x|^2 + const \tag{8}$$

1.1 Approximations

VMP will result in intractable computations because of the non-linear term in the node factor. Therefore we will approximate this non-linear function by a first-order (vector) Taylor expansion. The individual derivatives are given as

$$\frac{\partial}{\partial s} \ln(e^s + e^n) = \frac{e^s}{e^s + e^n} = \frac{1}{1 + e^{n-s}} = \sigma(s - n), \tag{9}$$

$$\frac{\partial}{\partial n} \ln(e^s + e^n) = \frac{e^n}{e^s + e^n} = \frac{1}{1 + e^{s-n}} = \sigma(n - s), \tag{10}$$

$$\frac{\partial}{\partial s} \frac{1}{e^s + e^n} = \frac{-e^s}{(e^s + e^n)^2} = -e^{-s}\sigma(s - n)^2 \tag{11}$$

and

$$\frac{\partial}{\partial n}\frac{1}{e^s+e^n} = \frac{-e^n}{(e^s+e^n)^2} = -e^{-n}\sigma(n-s)^2 \tag{12}$$

where we specify the sigmoid function $\sigma(x)$ as

$$\sigma(x) = \frac{1}{1+e^{-x}}. \tag{13}$$

2 Message $\nu(x)$

The message $\nu(x)$ can be calculated as

$$\ln \nu(x) = \mathrm{E}_{q(s)q(n)}\left[\ln p(x \mid s, n)\right] \tag{14}$$

The exponential terms cause issues. Let us focus on the first term $\log(e^s + e^n)$. The expectation with respect to this term cannot be calculated exactly. Instead we need to approximate it (see Zotero). Here we approximate it with the first-order Taylor expansion around its mean as

$$\ln(e^s+e^n) \approx \ln(e^{m_s}+e^{m_n})+\sigma(m_s-m_n)(s-m_s)+\sigma(m_n-m_s)(n-m_n). \tag{15}$$

Now the expectation simply becomes

$$\mathrm{E}_{q(s)q(n)}\left[\ln p(x \mid s,\ n)\right] \approx \ln(e^{m_s}+e^{m_n}). \tag{16}$$

Similarly the second term in the log-pdf can be approximated as

$$\mathrm{E}_{q(s)q(n)}\left[\frac{1}{e^s+e^n}\right] \approx \frac{1}{e^{m_s}+e^{m_n}}. \tag{17}$$

Using the above approximation we can approximate the log-message as

$$\begin{aligned}\ln \nu(x) &= \mathrm{E}_{q(s)q(n)}\left[\ln p(x \mid s,\ n)\right] \\ &\approx -\frac{1}{e^{m_s}+e^{m_n}}|x|^2 + \mathit{const}\end{aligned} \tag{18}$$

and identify it as

$$\boxed{\nu(x) \propto \mathcal{N}_{\mathcal{C}}(x \mid 0,\ e^{m_s}+e^{m_n},\ 0)} \tag{19}$$

3 Message $\nu(s)$

For the message $\nu(s)$ we again need to perform some approximations. Below you can find the approximations and their pros and cons. The expecation with respect to x is well-defined and simplifies the problem to

$$\begin{aligned}\ln \nu(s) &= \mathrm{E}_{q(n)q(x)}\left[\ln p(x \mid s,\ n)\right] + \mathit{const} \\ &= \mathrm{E}_{q(n)}\left[-\ln(e^s+e^n) - \frac{1}{e^s+e^n}(v_X + |m_x|^2)\right] + \mathit{const}\end{aligned} \tag{20}$$

3.1 Solution 1: Full Taylor expansion + LaPlace marginal

If we perform a full Taylor expansion (just like by $\nu(x)$) the message can be found as

$$\ln \nu(s) \approx -s \left(\sigma(m_s - m_n) - e^{-m_s}\sigma(m_s - m_n)^2\right) \tag{21}$$

This message has a linear log-pdf and is improper. An LaPlace approximation for this message is not useful, however, the resulting marginal $q(s)$ can be approximated using LaPlace.

3.2 Solution 2: Partial Taylor expansion + LaPlace message

Alternatively, we can linearize the log-message only with respect to n, yielding

$$\ln \nu(s) = -\ln(e^s + e^{m_n}) - \frac{1}{e^s + e^{m_n}}(v_y + |m_x|^2) + const \tag{22}$$

This message has a sigmoidal shape. The log-pdf is shaped horizontally and bends of linearly. This shape is not perse appropriate for approximating it with a Gaussian distribution, but if we want, we could try. The parial derivative can be found as

$$\frac{\partial}{\partial s} \ln \nu(s) = \frac{e^s}{e^s + e^{m_n}} \left(\frac{v_x + |m_x|^2}{e^s + e^{m_n}} - 1\right) \tag{23}$$

The mean can be found by calculating the corresponding zeros, which is found as

$$s_0 = \ln(v_x + |m_x|^2 - e^{m_n}), \tag{24}$$

from which it can be seen that this mode/mean is not always defined.

3.3 Solution 3: Partial Taylor expansion + LaPlace marginal

To make sure we can calculate the resulting marginal, the message

$$\ln \nu(s) = -\ln(e^s + e^{m_n}) - \frac{1}{e^s + e^{m_n}}(v_x + |m_x|^2) + const \tag{25}$$

can be propagated along the edge. Using the incoming message, the marginal distribution can be expressed as a function of s. Using numerical optimization (as no closed-form expression is present) the mode and variance around the mode can be determined, allowing for a Gaussian LaPlace approximation of the marginal.

3.4 Alternative solutions

Besides the LaPlace approximation, we can solve the problem by moment matching (using for example quadrature integration or importance sampling).

4 Message $\nu(n)$

By argument of symmetry, above strategies can be followed for the message $\nu(n)$.

5 Free energy

The local free energy can be calculated (by a Taylor expansion) as

$$\begin{aligned} U &= -\mathrm{E}_{q(x)q(s)q(n)} \left[\ln p(x \mid s, n)\right] \\ &\approx \ln(\pi) + \ln(e^{m_s} + e^{m_n}) + \frac{v_x + |m_x|^2}{e^{m_s} + e^{m_n}} \end{aligned} \tag{26}$$

6 Overview

The node factor is given by

$$p(y \mid x_1, \ldots, x_K) = \mathcal{N}_{\mathcal{C}}\left(y \;\middle|\; 0, \sum_{k=1}^{K} \exp(x_k), 0\right), \tag{27}$$

where

$$\mathcal{N}_{\mathcal{C}}(x \mid \mu, \sigma^2, 0) = \frac{1}{\pi\sigma^2} e^{-\frac{1}{\sigma^2}|x-\mu|^2}, \tag{28}$$

so

$$p(y \mid x_1, \ldots, x_K) = \frac{1}{\pi \sum_{k=1}^{K} \exp(x_k)} \exp\left\{ -\frac{1}{\sum_{k=1}^{K} \exp(x_k)} |y|^2 \right\} \tag{29}$$

The approximate posterior is given by

$$q(y, \boldsymbol{x}) = q(y) \prod_{k=1}^{K} q(x_k), \tag{30}$$

where

$$q(y) = \mathcal{N}_{\mathcal{C}}(y \mid m_y, v_y, 0) \tag{31}$$

$$q(x_k) = \mathcal{N}(x_k \mid m_{x_k}, v_{x_k}) \tag{32}$$

$$\tag{33}$$

The log-pdf can be found as

$$\ln p(y \mid x_1, \ldots, x_K) = -\ln(\pi) - \ln\left(\sum_{k=1}^{K} \exp(x_k)\right) - \frac{1}{\sum_{k=1}^{K} \exp(x_k)} |y|^2 \tag{34}$$

6.1 Approximations

VMP will result in intractable computations because of the non-linear term in the node factor. Therefore we will approximate this non-linear function by a first-order (vector) Taylor expansion. The individual derivatives are given as

$$\frac{\partial}{\partial x_k} \ln\left(\sum_{i=1}^{K} \exp(x_i)\right) = \frac{\exp(x_k)}{\sum_{i=1}^{K} \exp(x_i)} \tag{35}$$

and the corresponding gradient is given as

$$\nabla_x \ln\left(\sum_{i=1}^{K} \exp(x_i)\right) = \sigma(\boldsymbol{x}) \tag{36}$$

where we specify the softmax function $\sigma(\boldsymbol{x})$ as

$$\sigma(\boldsymbol{x})_k = \frac{\exp(x_k)}{\sum_{i=1}^{K} \exp(x_i)} \tag{37}$$

The other term can be approximated as

$$\frac{\partial}{\partial x_k}\frac{1}{\sum_{i=1}^{K}\exp(x_i)} = \frac{-\exp(x_k)}{\left(\sum_{i=1}^{K}\exp(x_i)\right)^2} \tag{38}$$

and the corresponding gradient is given as

$$\nabla_x \frac{1}{\sum_{i=1}^{K}\exp(x_i)} = -\exp.(\boldsymbol{x}) \circ \sigma(\boldsymbol{x}) \circ \sigma(\boldsymbol{x}) \tag{39}$$

Using the above approximations we can approximate the log-likelihood as

$$\begin{aligned} \ln p(y \mid x_1, \ldots, x_K) \approx & -\ln(\pi) - \ln\left(\sum_{k=1}^{K}\exp(m_{x_k})\right) - \frac{1}{\sum_{k=1}^{K}\exp(m_{x_k})}|y|^2 \\ & - \sigma(\boldsymbol{m}_x)^\top(\boldsymbol{x} - \boldsymbol{m}_x) + (\exp.(\boldsymbol{m}_x) \circ \sigma(\boldsymbol{m}_x) \circ \sigma(\boldsymbol{m}_x))^\top(\boldsymbol{x} - \boldsymbol{m}_x)|y|^2 \end{aligned} \tag{40}$$

7 Message $\nu(y)$

The message $\nu(y)$ can be calculated as

$$\ln \nu(y) = \mathrm{E}_{q(\boldsymbol{x})}\left[\ln p(y \mid x_1, \ldots, x_K)\right] \tag{41}$$

$$\begin{aligned} \ln \nu(y) &= \mathrm{E}_{q(\boldsymbol{x})}\left[\ln p(y \mid x_1, \ldots, x_K)\right] + \mathit{const} \\ &\approx \mathrm{E}_{q(\boldsymbol{x})}\Bigg[-\frac{1}{\sum_{k=1}^{K}\exp(m_{x_k})}|y|^2 \\ &\qquad +(\exp.(\boldsymbol{m}_x) \circ \sigma(\boldsymbol{m}_x) \circ \sigma(\boldsymbol{m}_x))^\top(\boldsymbol{x} - \boldsymbol{m}_x)|y|^2\Bigg] + \mathit{const} \\ &= -\frac{1}{\sum_{k=1}^{K}\exp(m_{x_k})}|y|^2 + \mathit{const} \end{aligned} \tag{42}$$

From this description the variational message can be identified as

$$\boxed{\nu(y) \propto \mathcal{N}_{\mathcal{C}}\left(y \,\middle|\, 0,\ \sum_{k=1}^{K}\exp\left(m_{x_k}\right),\ 0\right)} \tag{43}$$

8 Message $\nu(x_k)$

The message $\nu(x_k)$ can be calculated as

$$\begin{aligned}
\ln \nu(x_k) &= \mathrm{E}_{q(y)q(\boldsymbol{x}_{\setminus k})}\left[\ln p(y \mid x_1, \ldots, x_K)\right] + \mathit{const} \\
&= \mathrm{E}_{q(y)q(\boldsymbol{x}_{\setminus k})}\left[-\ln\left(\sum_{i=1}^{K} \exp(x_i)\right) - \frac{1}{\sum_{i=1}^{K} \exp(x_i)} |y|^2\right] + \mathit{const} \\
&= \mathrm{E}_{q(\boldsymbol{x}_{\setminus k})}\left[-\ln\left(\sum_{k=1}^{K} \exp(x_i)\right) - \frac{1}{\sum_{i=1}^{K} \exp(x_i)} (v_y + |m_y|^2)\right] + \mathit{const} \\
&= \mathrm{E}_{q(\boldsymbol{x}_{\setminus k})}\left[-\ln\left(\exp(x_k) + \sum_{i \neq k} \exp(x_i)\right) - \frac{1}{\exp(x_k) + \sum_{i \neq k} \exp(x_i)} (v_y + |m_y|^2)\right] + \mathit{const} \\
&\approx \mathrm{E}_{q(\boldsymbol{x}_{\setminus k})}\left[-\ln\left(\exp(x_k) + \sum_{i \neq k} \exp(m_{x_i})\right) - \frac{1}{\exp(x_k) + \sum_{i \neq k} \exp(m_{x_i})} (v_y + |m_y|^2)\right] \\
&\quad - \sigma(\boldsymbol{m}_{x_{\setminus k}})^\top (\boldsymbol{x}_{\setminus k} - \boldsymbol{m}_{x_{\setminus k}}) + (\exp.(\boldsymbol{m}_{x_{\setminus k}}) \circ \sigma(\boldsymbol{m}_{x_{\setminus k}}) \circ \sigma(\boldsymbol{m}_{x_{\setminus k}}))^\top (\boldsymbol{x}_{\setminus k} - \boldsymbol{m}_{x_{\setminus k}})(v_y + |m_y|^2) + \mathit{co} \\
&= -\ln\left(\exp(x_k) + \sum_{i \neq k} \exp(m_{x_i})\right) - \frac{1}{\exp(x_k) + \sum_{i \neq k} \exp(m_{x_i})} (v_y + |m_y|^2)
\end{aligned} \tag{44}$$

Here the non-linear function is approximated using a first-order Taylor approximation for all $\boldsymbol{x}$ except for x_k. From this description the variational message can be identified as

$$\boxed{\nu(x_k) \propto \exp\left(-\ln\left(\exp(x_k) + \sum_{i \neq k} \exp(m_{x_i})\right) - \frac{1}{\exp(x_k) + \sum_{i \neq k} \exp(m_{x_i})} (v_y + |m_y|^2)\right)} \tag{45}$$

or

$$\boxed{\nu(x_k) \propto \frac{1}{\exp(x_k) + \sum_{i \neq k} \exp(m_{x_i})} \exp\left(-\frac{1}{\exp(x_k) + \sum_{i \neq k} \exp(m_{x_i})} (v_y + |m_y|^2)\right)} \tag{46}$$

9 Free energy

The local free energy can be calculated (by a Taylor expansion) as

$$
\begin{aligned}
U &= -\mathrm{E}_{q(y)q(\boldsymbol{x})}\left[\ln p(y \mid x_1, \ldots, x_K)\right] \\
&= \mathrm{E}_{q(y)q(\boldsymbol{x})}\left[\ln(\pi) + \ln\left(\sum_{k=1}^{K} \exp(x_k)\right) + \frac{1}{\sum_{k=1}^{K} \exp(x_k)} |y|^2\right] \\
&= \mathrm{E}_{q(\boldsymbol{x})}\left[\ln(\pi) + \ln\left(\sum_{k=1}^{K} \exp(x_k)\right) + \frac{1}{\sum_{k=1}^{K} \exp(x_k)} (v_y + |m_y|^2)\right] \\
&\approx \ln(\pi) + \ln\left(\sum_{k=1}^{K} \exp(m_{x_k})\right) + \frac{1}{\sum_{k=1}^{K} \exp(m_{x_k})} (v_y + |m_y|^2)
\end{aligned}
\tag{47}
$$