Supplement: Belief Propagation in Conditional RBMs for Structured Prediction

Wei Ping
Computer Science, UC Irvine
weiping.thu@gmail.com

Alexander Ihler
Computer Science, UC Irvine
ihler@ics.uci.edu

A Gradients of Log-likelihood

Similar to Eq. (5) in the main text, the gradients of log-likelihood with other weights and biases are,

$$\frac{\partial \log p(\boldsymbol{v}^n|\boldsymbol{x}^n)}{\partial W^{vx}} = \boldsymbol{v}^n \boldsymbol{x}^{n\top} - \mathbb{E}_{p(\boldsymbol{v},\boldsymbol{h}|\boldsymbol{x}^n)}\big[\boldsymbol{v}\boldsymbol{x}^{n\top}\big],$$
$$\frac{\partial \log p(\boldsymbol{v}^n|\boldsymbol{x}^n)}{\partial W^{hx}} = \boldsymbol{\mu}^n \boldsymbol{x}^{n\top} - \mathbb{E}_{p(\boldsymbol{v},\boldsymbol{h}|\boldsymbol{x}^n)}\big[\boldsymbol{h}\boldsymbol{x}^{n\top}\big],$$
$$\frac{\partial \log p(\boldsymbol{v}^n|\boldsymbol{x}^n)}{\partial b^{v}} = \boldsymbol{v}^n - \mathbb{E}_{p(\boldsymbol{v},\boldsymbol{h}|\boldsymbol{x}^n)}\big[\boldsymbol{v}\big],$$
$$\frac{\partial \log p(\boldsymbol{v}^n|\boldsymbol{x}^n)}{\partial b^{h}} = \boldsymbol{\mu}^n - \mathbb{E}_{p(\boldsymbol{v},\boldsymbol{h}|\boldsymbol{x}^n)}\big[\boldsymbol{h}\big],$$

where $\boldsymbol{\mu}^n = \mathbb{E}_{p(\boldsymbol{h}|\boldsymbol{v}^n,\boldsymbol{x}^n)}\big[\boldsymbol{h}\big] = \sigma(W^{vh\top}\boldsymbol{v}^n + W^{hx}\boldsymbol{x}^n + \boldsymbol{b}^h)$. All the negative parts of these gradients are intractable to calculate, and must be approximated during learning.

B Derivation of Matrix-based BP

In this section we give additional proof details of our matrix-based BP update equations.

B.1 Proof of the update rule for M^{vh} in Eq. (13):

$$M^{vh}_{ij} = \frac{m_{j\to i}(v_i=1)}{m_{j\to i}(v_i=1)+m_{j\to i}(v_i=0)},$$
$$= \sigma\Big(\log \frac{\exp(W^{vh}_{ij})\cdot \frac{\tau^h_j}{M^{hv}_{ji}} + \frac{1-\tau^h_j}{1-M^{hv}_{ji}}}{\frac{\tau^h_j}{M^{hv}_{ji}} + \frac{1-\tau^h_j}{1-M^{hv}_{ji}}}\Big), \text{ by Eq. (8)}$$
$$= \sigma\Big(\log \frac{\exp(W^{vh}_{ij})\cdot(1-M^{hv}_{ji})\tau^h_j + M^{hv}_{ji}(1-\tau^h_j)}{(1-M^{hv}_{ji})\tau^h_j + M^{hv}_{ji}(1-\tau^h_j)}\Big).$$

Then, one can verify the update of M^{vh} (13) holds. The derivation is analogous for updating M^{hv} (14).

B.2 Proof of the update rule for $\boldsymbol{\tau}^v$ in Eq. (16):

$$\tau^v_i = \frac{\tau(v_i=1)}{\tau(v_i=1)+\tau(v_i=0)},$$
$$= \frac{1}{1+\frac{\exp\left(0+\sum_{j=1}^{|\boldsymbol{h}|}\log m_{j\to i}(v_i=0)\right)}{\exp\left(b_i+\sum_{j=1}^{|\boldsymbol{h}|}\log m_{j\to i}(v_i=1)\right)}}, \text{ by Eq. (11)}$$
$$= \frac{1}{1+\exp\big\{-b^1_i - \sum_{j=1}^{|\boldsymbol{h}|}\big(\log M^{vh}_{ij} - \log(1-M^{vh}_{ij})\big)\big\}},$$
$$= \sigma\Big(b^1_i + \big(\log M^{vh}_{i\bullet} - \log(\mathbf{1}^{\boldsymbol{h}\top} - M^{vh}_{i\bullet})\big)\cdot \mathbf{1}^h\big\}\Big).$$

Then, one can verify the update of $\boldsymbol{\tau}^v$ (16) holds. The update of $\boldsymbol{\tau}^h$ in Eq. (17) is derived similarly.

B.3 The (i,j) element of pairwise belief matrix:

$$\Gamma_{ij} = \frac{\tau(v_i=1,h_j=1)}{\sum_{v_i,h_j}\tau(v_i,h_j)} =$$
$$\frac{\frac{\exp(w^{vh}_{ij})\tau^v_i\tau^h_j}{M^{vh}_{ij}M^{hv}_{ji}}}{\frac{\exp(w^{vh}_{ij})\tau^v_i\tau^h_j}{M^{vh}_{ij}M^{hv}_{ji}} + \frac{(1-\tau^v_i)\tau^h_j}{(1-M^{vh}_{ij})M^{hv}_{ji}} + \frac{\tau^v_i(1-\tau^h_j)}{M^{vh}_{ij}(1-M^{hv}_{ji})} + \frac{(1-\tau^v_i)(1-\tau^h_j)}{(1-M^{vh}_{ij})(1-M^{hv}_{ji})}}$$

We can denote the intermediate terms

$$\Gamma^{11} = \exp(W^{vh})\circ(\boldsymbol{\tau}^v\cdot\boldsymbol{\tau}^{h\top})\circ(\mathbf{1}^{vh}-M^{vh})\circ(\mathbf{1}^{hv}-M^{hv})^\top,$$
$$\Gamma^{01} = \big((\mathbf{1}^v-\boldsymbol{\tau}^v)\cdot\boldsymbol{\tau}^{h\top}\big)\circ M^{vh}\circ(\mathbf{1}^{hv}-M^{hv})^\top,$$
$$\Gamma^{10} = \big(\boldsymbol{\tau}^v\cdot(\mathbf{1}^h-\boldsymbol{\tau}^h)^\top\big)\circ(\mathbf{1}^{vh}-M^{vh})\circ M^{hv\top},$$
$$\Gamma^{00} = \big((\mathbf{1}^v-\boldsymbol{\tau}^v)\cdot(\mathbf{1}^h-\boldsymbol{\tau}^h)\big)\circ M^{vh}\circ M^{hv\top}.$$

Then, the pairwise belief matrix $\Gamma = \frac{\Gamma^{11}}{\Gamma^{11}+\Gamma^{01}+\Gamma^{10}+\Gamma^{00}}$.