0.1 Spike and Slab with sample-wise noise precision

0.1.1 The model

We have a data set $\mathbf{Y}$ of M input views with dimensionality $\mathbf{Y}^m \in \mathbb{R}^{N\times D_m}$. The sparse group factor analysis model is defined as follows:

$$y_{nd} \approx \mathcal{N}\left(y_{nd}^m \,|\, \mathbf{w}_{d,:}^m \mathbf{z}_{n,:}, 1/\tau_{dn}^m\right) \tag{1}$$
$$w_{dk}^m \approx \theta\mathcal{N}\left(w_{dk} \,|\, 0, 1/\alpha_k^m\right) + (1-\theta)\delta_0(w_{dk}^m) \tag{2}$$
$$z_{nk} \approx \mathcal{N}\left(z_{nk} \,|\, 0, 1\right) \tag{3}$$

The precision of observed data is allowed to vary for samples and features to incorporate a pseudo-data view on the Jaakola bound. This parameter is updated along with the variational parameter of the pseudodata and is not treated as a variational node itself.

0.1.2 Changes to standard spike and slab

The updates affected by this change are the one of $w = s\hat{w}$ and z. In the ELB the Bernoulli likelihood of the truly observed data is used. For all other updates and terms refer to spike_slab.tex.

0.1.3 Joint probability density function:

$$p(\mathbf{Y}, \hat{\mathbf{W}}, \mathbf{S}, \mathbf{Z}) = \prod_{m=1}^{M} \prod_{n=1}^{N} \prod_{d=1}^{D_m} \mathcal{N}\left(y_{nd}^m \,|\, \sum_{k=1}^{K} s_{dk}^m \hat{w}_{dk}^m z_{nk}, 1/\tau_{nd}\right) \times$$
$$\prod_{d=1}^{D_m} \prod_{k=1}^{K} \mathcal{N}\left(\hat{w}_{dk}^m \,|\, 0, 1/\alpha_k^m\right) \theta^{s_{dk}^m} (1-\theta)^{1-s_{dk}} \times$$
$$\prod_{n=1}^{N} \prod_{k=1}^{K} \mathcal{N}\left(z_{nk} \,|\, 0, 1\right)$$

Full variational distribution:

$$q(\hat{\mathbf{W}}, \mathbf{S}, \mathbf{Z}) = \prod_{m=1}^{M} \prod_{d=1}^{D_m} \prod_{k=1}^{K} q(\hat{w}_{dk}^m, s_{dk}^m) \prod_{k=1}^{K} \prod_{n=1}^{N} q(z_{n,k})$$

0.1.4 Derivation of update equations

The optimal distribution $\hat{q}_i$ for each variable $\mathbf{x}_i$, is the following:

$$\log \hat{q}_i(\mathbf{x}_i) = \mathbb{E}_{i\neq j}[\log p(\mathbf{Y}, \mathbf{X})] + \text{const.} \tag{4}$$

where $\mathbb{E}_{i\neq j}$ denotes an expectation with respect to the q distributions over all variables $\mathbf{x}_j$ except for $\mathbf{x}_i$. The additive constant is set by normalising the distribution $\hat{q}_i(\mathbf{z}_i)$:

$$\hat{q}_i(\mathbf{x}_i) = \frac{\exp(\mathbb{E}_{i\neq j}[\log p(\mathbf{Y}, \mathbf{X})])}{\int \exp(\mathbb{E}_{i\neq j}[\log p(\mathbf{Y}, \mathbf{X})]) d\mathbf{X}}$$

Latent variables

Term from the likelihood $p(\mathbf{Y}|\hat{\mathbf{W}}, \mathbf{Z}, \boldsymbol{\tau}, \mathbf{S})$:

$$\sum_{m=1}^{M} \sum_{d=1}^{D=m} \langle \tau_{nd}^m \rangle \langle s_{dk}^m \hat{w}_{dk}^m \rangle y_{nd}^m z_{nk} - \frac{1}{2} \sum_{m=1}^{M} \sum_{d=1}^{D_m} \langle \tau_{nd}^m \rangle \langle (s_{dk}^m \hat{w}_{dk}^m)^2 \rangle z_{nk}^2$$
$$- \frac{1}{2} \sum_{m=1}^{M} \frac{1}{2} \sum_{d=1}^{D} \langle \tau_{nd}^m \rangle \sum_{j \neq k} (\langle s_{dj}^m \hat{w}_{dj}^m \rangle \langle z_{nj} \rangle) \langle \hat{w}_{dk}^m s_{dk}^m \rangle \langle z_{nk} \rangle + \text{const.}$$

Term from the prior $p(z_{nk})$:

$$-\frac{1}{2} z_{nk}^2 + \text{const.}$$

Variational distribution:

$$q(\mathbf{Z}) = \prod_{k=1}^{K} \prod_{n=1}^{N} q(z_{nk}) = \prod_{k=1}^{K} \prod_{n=1}^{N} \mathcal{N}\left(z_{nk} \,|\, \mu_{z_{nk}}, \sigma_{z_{nk}}\right)$$

where

$$\sigma_{z_{nk}}^2 = \Big(\sum_{m=1}^{M} \sum_{d=1}^{D_m} \tau_{nd}^m \langle (s_{dk}^m \hat{w}_{dk}^m)^2 \rangle + 1 \Big)^{-1}$$

$$\mu_{z_{nk}} = \sigma_{z_{nk}}^2 \sum_{m=1}^{M} \sum_{d=1}^{D_m} \langle \tau_{nd}^m \rangle \langle s_{dk}^m \hat{w}_{dk}^m \rangle \Big(y_{nd}^m - \sum_{j \neq k} \langle s_{dj}^m \hat{w}_{dj}^m \rangle \langle z_{nj} \rangle \Big)$$

Spike and Slab Weights

Variational distribution:

$$q(\hat{\mathbf{W}}, \mathbf{S}) = \prod_{m=1}^{M} \prod_{d=1}^{D_m} \prod_{k=1}^{K} q(\hat{w}_{dk}^m, s_{dk}^m) = \prod_{m=1}^{M} \prod_{d=1}^{D_m} \prod_{k=1}^{K} q(\hat{w}_{dk}^m | s_{dk}^m) q(s_{dk}^m)$$

Update for $q(s_{dk}^m)$:

$$\gamma_{dk} = q(s_{dk} = 1) = \frac{1}{1 + \exp(-\lambda_{dk})} = logit^{-1}(\lambda_{dk})$$

where

$$\lambda_{dk}^m = \langle \log \frac{\theta}{1-\theta} \rangle + \frac{\Big(\sum_{n=1}^{N} y_{nd}^m \langle z_{nk} \rangle \tau_{nd}^m - \sum_{j \neq k} \langle s_{dj}^m \hat{w}_{dj}^m \rangle \sum_{n=1}^{N} \langle z_{nk} \rangle \langle z_{nj} \rangle \tau_{nd}^m \Big)^2}{\sum_{n=1}^{N} \tau_{nd}^m \langle z_{nk}^2 \rangle + \langle \alpha_k^m \rangle}$$
$$+ 0.5 \log(\langle \alpha_k^m \rangle) - 0.5 \log(\sum_{n=1}^{N} \tau_{nd}^m \langle z_{nk}^2 \rangle + \langle \alpha_k^m \rangle)$$

Update for $q(\hat{w}_{dk}^m)$:

$$q(\hat{w}_{dk}^m | s_{dk}^m = 0) = \mathcal{N}\left(\hat{w}_{dk}^m \,|\, 0, 1/\alpha_k^m\right)$$
$$q(\hat{w}_{dk}^m | s_{dk}^m = 1) = \mathcal{N}\left(\hat{w}_{dk}^m \,|\, \mu_{w_{dk}^m}, \sigma_{w_{dk}^m}^2\right)$$

where

$$\mu_{w^m_{dk}} = \sigma^2_{w^m_{dk}} \left(\sum_{n=1}^{N} \tau^m_{nd} y^m_{nd} \langle z_{nk} \rangle - \sum_{j \neq k} \langle s^m_{dj} \hat{w}^m_{dj} \rangle \sum_{n=1}^{N} \tau^m_{nd} \langle z_{nk} \rangle \langle z_{nj} \rangle \right)$$

$$\sigma^2_{w^m_{dk}} = \frac{1}{\sum_{n=1}^{N} \tau^m_{nd} \langle z^2_{nk} \rangle + \langle \alpha^m_k \rangle}$$

Taken together this means that we can update $q(\hat{w}^m_{dk}, s^m_{dk})$ using:

$$q(\hat{w}^m_{dk} | s^m_{dk}) \times q(s^m_{dk}) = \mathcal{N}\left(\hat{w}^m_{dk} \mid s^m_{dk} \mu_{w^m_{dk}}, s^m_{dk} \sigma^2_{w^m_{dk}} + (1 - s^m_{dk})/\alpha^m_k\right) \quad \times \quad (\lambda^m_{dk})^{s^m_{dk}} (1 - \lambda^m_{dk})^{1 - s_{dk}}$$

0.1.5 Lower bound

Likelihood term

$$\log p(y^m_{nd} | Z, W) = y^m_{nd} \sum_{k=1}^{K} \langle s^m_{dk} \hat{w}^m_{dk} \rangle \langle z_{nk} \rangle - \log \left(1 + \exp \left(\sum_{k=1}^{K} \langle s^m_{dk} \hat{w}^m_{dk} \rangle \langle z_{nk} \rangle \right) \right)$$