A nonparametric variable clustering model: supplementary material

Here we present the update equations for the DPVC Gibbs sampler.

Sampling the factor loading matrix G.

$$g_{dk}|\mathbf{Y}, \mathbf{G}_{-dk}, \mathbf{C}, \mathbf{X}, \sigma_g, \sigma_x, \sigma_d, \alpha \sim \mathcal{N}(\mu_g^*, \lambda_g^{-1}) \tag{1}$$

where

$$\lambda_g, \mu_g^* = \begin{cases} \sigma_g^2, \mu_g & \text{if } c_{dk} = 0 \\ \frac{\sum_n x_{kn}^2}{\sigma_d^2} + \frac{1}{\sigma_g^2}, \lambda_g^{-1}\left(\frac{1}{\sigma_d^2}\mathbf{Y_{d:}}\mathbf{X_{k:}^\intercal} + \frac{1}{\sigma_g^2}\mu_g\right) & \text{if } c_{dk} = 1 \end{cases} \tag{2}$$

Sampling the latent factor matrix X.

$$\mathbf{X_{:n}}|\mathbf{Y}, \mathbf{X}_{-:n}, \mathbf{G}, \mathbf{C}, \sigma_g, \sigma_x, \sigma_d, \alpha \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{X}_{:n}}, \boldsymbol{\Lambda}_{\boldsymbol{X}_{:n}}^{\mathbf{-1}}) \tag{3}$$

where

$$\boldsymbol{\Lambda}_{\boldsymbol{X}_{:n}} = (\mathbf{G}\cdot\mathbf{C})^\top \boldsymbol{\Sigma}_{\boldsymbol{y}}^{\mathbf{-1}}(\mathbf{G}\cdot\mathbf{C}) + \boldsymbol{\Sigma}_{\boldsymbol{x}}^{\mathbf{-1}}$$
$$\boldsymbol{\mu}_{\boldsymbol{X}_{:n}} = \boldsymbol{\Lambda}_{\boldsymbol{X}_{:n}}^{\mathbf{-1}}(\mathbf{G}\cdot\mathbf{C})^\top \boldsymbol{\Sigma}_{\boldsymbol{y}}^{\mathbf{-1}}\mathbf{Y_{:n}}$$

Cluster assignments c . When sampling the cluster assignments, **c** we found it beneficial to integrate out **g**, while instantiating **X**. We require

$$P(c_d = k|y_{d:}, x_{k:}, \sigma_g) = \int P(y_{d:}|x_{k:}, g_d)p(g_d|\sigma_g)dg_d$$
$$= \frac{1}{(2\pi\sigma_d^2)^{N/2}} e^{-\frac{y_{d:}y_{d:}^T}{2\sigma_d^2}} \frac{1}{\lambda_g^{\frac{1}{2}}\sigma_g} e^{\frac{\lambda_g {\mu^*}_g^2}{2}}$$

where λ_g, μ_g^* are defined in Equation 2.

Hyperparameters. We used slice sampling (**?**) to sample the CRP hyperparameter α, while the posterior updates of the σ_g and σ_d are as follows:

$$\sigma_g^2|\mathbf{G} \sim \mathcal{IG}\left(\frac{DK}{2} + 1, \frac{\sum_k \mathbf{g}_{:k}^\top \mathbf{g_{:k}}}{2} + 1\right)$$
$$\sigma_d^2|\mathbf{Y}, \mathbf{G}, \mathbf{C}, \mathbf{X} \sim \mathcal{IG}\left(\frac{DN+1}{2}, \frac{\sum_n (\mathbf{Y_{:n}} - \boldsymbol{\mu}_{:n})^2}{2} + 0.1\right)$$