Variational Inference Schemes

Mimno-SVI

The form of the local approximation in the Mimno-SVI method is

$$\log q_{\text{Mimno}}(\psi_i) = E_{q(\beta)}[\log p(\psi_i | y_{1:N}, \beta)]$$

$$= E_{q(\beta)} \left[ -\frac{\gamma_{\text{obs}}}{2} \| y_i - (z_i \circ w_i) \Phi \|^2 + \sum_k z_{ik} \log \left( \frac{\pi_k}{1 - \pi_k} \right) - \frac{\gamma_w}{2} w_i w_i^T \right] + \text{const}$$

$$= -\frac{c}{2d} \sum_k z_{ik} w_{ik} \left[ w_{ik} \left( \frac{\mu_k \mu_k^T}{\tau_k^2} + \frac{1}{\tau_k} \right) + \left( \sum_{j \neq k} z_{ij} w_{ij} \frac{\mu_k \mu_j^T}{\tau_k \tau_j} \right) - 2 \frac{\mu_k y_i}{\tau_k} \right]$$

$$+ \sum_k z_{ik} (\psi(a_k) - \psi(b_k)) - e f w_i w_i^T + \text{const}$$

It is clear that $\log q_{\text{Mimno}}$ is quadratic in each $w_{ik}$ and linear in each $z_{ik}$, therefore a Gibbs based sampler can easily be constructed to sample from $q_{\text{Mimno}}$, where $w_{ik}$ is Gaussian given all other local variables, and $z_{ik}$ is Bernoulli given all other local variables.

MF-SSVI

The local ELBO in the MF-SSVI framework is very similar to that of the MF-SVI, the difference being that samples of the global variables are used in MF-SSVI. The local ELBO has the following form

$$L_{\text{MF-SSVI}}^{\text{local}} = \frac{\gamma_{\text{obs}}}{2} \sum_{i,k} \theta_{ik} \phi_k \left[ \frac{2 \nu_{ik}}{\kappa_{ik}} - \left( \frac{\nu_{ik}^2}{\kappa_{ik}^2} + \frac{1}{\kappa_{ik}} \right) \phi_k^T \right] - \frac{\gamma_w}{2} \sum_{i,k} \left( \frac{\nu_{ik}^2}{\kappa_{ik}^2} + \frac{1}{\kappa_{ik}} \right) + \sum_{i,k} \theta_{ik} \left( \frac{\pi_k}{1 - \pi_k} \right)$$

$$- \frac{1}{2} \sum_{i,k} \log(\kappa_{ik}) - \sum_{i,k} \left[ \theta_{ik} \log \theta_{ik} + (1 - \theta_{ik}) \log(1 - \theta_{ik}) \right].$$
This is optimized as a function of \([\theta_{ik}, \nu_{ik}, \kappa_{ik}]\) using gradient descent. Once a local optimum is found, \(E_{q_{\psi_{1:N}^{(t)}}}(\eta_t)\) can be computed analytically as a function of the optimized parameters and global variable samples.

**Titsias-SSVI**

Recall that the Titsias-SSVI method maintains dependence between \(z_{ik}\) and \(w_{ik}\) for each \(k\). The local ELBO for Titsias-SSVI is

\[
L_{\text{Titsias-SSVI}}^{\text{local}} = \frac{\gamma_{\text{obs}}}{2} \sum_{i,k} \theta_{ik} \phi_k \left[ 2 \frac{\nu_{ik}}{\kappa_{ik}} y_i^\top - \left( \frac{\nu_{ik}^2}{\kappa_{ik}^2} + \frac{1}{\kappa_{ik}} \right) \phi_k^\top - \sum_{j \neq k} \theta_{ij} \frac{\nu_{ij}}{\kappa_{ij}} \frac{\nu_{ik}}{\kappa_{ik}} \phi_j^\top \right] - \frac{\gamma_w}{2} \sum_{i,k} \left( \frac{\nu_{ik}^2}{\kappa_{ik}^2} + \frac{1}{\kappa_{ik}} \right) + \sum_{i,k} \theta_{ik} \left( \frac{\pi_k}{1 - \pi_k} \right) - \frac{1}{2} \sum_{i,k} \left[ \theta_{ik} \left( \log(\kappa_{ik}) - 1 \right) + (1 - \theta_{ik}) \left( \log(\gamma_w) - 1 \right) \right] - \sum_{i,k} \left[ \theta_{ik} \log \theta_{ik} + (1 - \theta_{ik}) \log(1 - \theta_{ik}) \right].
\]

Again, this function is maximized as a function of \([\theta_{ik}, \nu_{ik}, \kappa_{ik}]\) using gradient descent, and the optimized parameters along with the global variable samples are used to compute \(E_{q_{\text{Titsias}}}(\psi_{1:N}|\beta^{(t)})[\eta_t]\) analytically.

**Gibbs-SSVI**

The Gibbs-SVI method uses the true posterior conditional distribution for local variables

\[
\log q_{\text{Gibbs}}(\psi_i) = \log p(\psi_i|y_{1:N}, \beta)
\]

\[
= -\frac{\gamma_{\text{obs}}}{2} \left\| y_i - (z_i \circ w_i) \Phi \right\|^2 + \sum_k z_{ik} \log \left( \frac{\pi_k}{1 - \pi_k} \right) - \frac{\gamma_w}{2} w_i w_i^\top + \text{const}
\]

\[
= -\frac{\gamma_{\text{obs}}}{2} \sum_k z_{ik} w_{ik} \phi_k \left[ w_{ik} \phi_k^\top + \left( \sum_{j \neq k} z_{ij} w_{ij} \phi_j^\top \right) - 2 y_i^\top \right]
\]

\[
+ \sum_k z_{ik} \log \left( \frac{\pi_k}{1 - \pi_k} \right) - \frac{\gamma_w}{2} w_i w_i^\top + \text{const}
\]

Just as was the case with Mimno-SVI, we notice that \(\log q_{\text{Gibbs}}\) is quadratic in each \(w_{ik}\) and linear in each \(z_{ik}\), therefore a Gibbs sampler can be designed to sample from \(q_{\text{Gibbs}}\).