Stochastic Variational Inference

Practice

Gaussian Processes for Big Data

- Variational inducing variables:

$$p(\mathbf{y}\,|\,\mathbf{f}) = \mathcal{N}\left(\mathbf{y}|\mathbf{f}, \beta^{-1}\mathbf{I}\right),$$
$$p(\mathbf{f}\,|\,\mathbf{u}) = \mathcal{N}\left(\mathbf{f}|\mathbf{K}_{nm}\mathbf{K}_{mm}^{-1}\mathbf{u}, \widetilde{\mathbf{K}}\right),$$
$$p(\mathbf{u}) = \mathcal{N}\left(\mathbf{u}|\mathbf{0}, \mathbf{K}_{mm}\right),$$

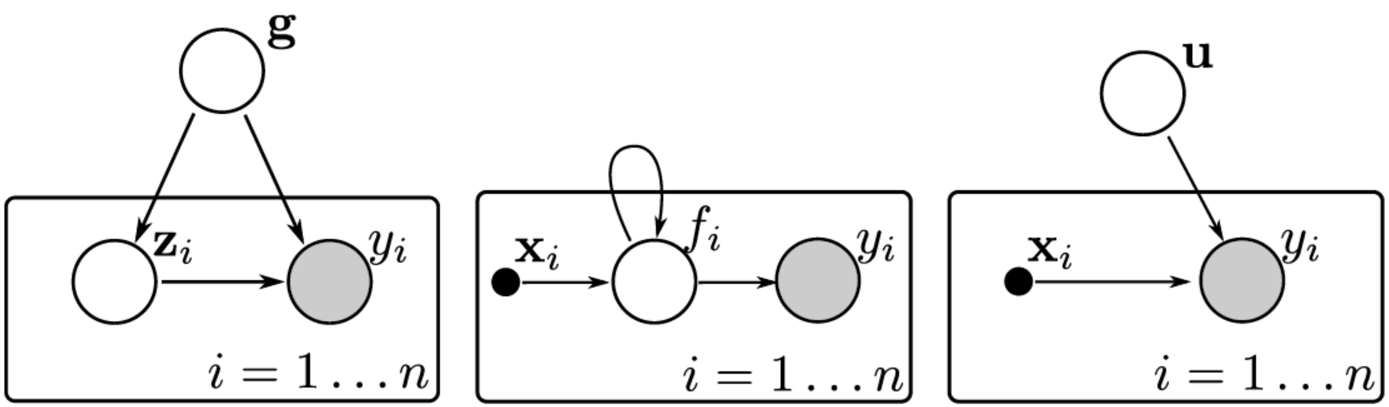
-
- Marginal likelihood:

$$\log p(\mathbf{y}\,|\,\mathbf{X}) = \log \int p(\mathbf{y}\,|\,\mathbf{u})p(\mathbf{u})\,\mathrm{d}\mathbf{u}$$
$$\geq \log \int \exp\{\mathcal{L}_1\}\, p(\mathbf{u})\,\mathrm{d}\mathbf{u} \triangleq \mathcal{L}_2,$$

- Lower bound:

$$\mathcal{L}_2 = \log \mathcal{N}\left(\mathbf{y}|\mathbf{0}, \mathbf{K}_{nm}\mathbf{K}_{mm}^{-1}\mathbf{K}_{mn} + \beta^{-1}\mathbf{I}\right) - \frac{1}{2}\beta \mathrm{tr}\left(\widetilde{\mathbf{K}}\right)$$

Gaussian Processes for Big Data



(a) Requirements for SVI (b) Gaussian Process regression (c) Variational GP regression

- Let u be the global variable with q(u) =N(m; S), leave m and S!

$$\log p(\mathbf{y} \mid \mathbf{X}) \geq \langle \mathcal{L}_1 + \log p(\mathbf{u}) - \log q(\mathbf{u}) \rangle_{q(\mathbf{u})} \triangleq \mathcal{L}_3.$$

$$\mathcal{L}_3 = \sum_{i=1}^{n} \left\{ \log \mathcal{N}\left(y_i | \mathbf{k}_i^\top \mathbf{K}_{mm}^{-1} \mathbf{m}, \beta^{-1}\right) - \frac{1}{2}\beta \widetilde{k}_{i,i} - \frac{1}{2} \mathrm{tr}\left(\mathbf{S}\mathbf{\Lambda}_i\right) \right\} - \mathrm{KL}\left(q(\mathbf{u}) \,\|\, p(\mathbf{u})\right)$$

Stochastic Variational Inference for Hidden Markov Models

- Joint distribution

$$p(\mathbf{x}, \mathbf{y}) = \pi_0(x_1)p(y_1|x_1)\prod_{t=2}^{T} p(x_t|x_{t-1}, A)p(y_t|x_t, \phi)$$

- Structured mean field

$$q(\theta, \mathbf{x}) = q(A)q(\phi)\boxed{q(\mathbf{x})}$$

- Lower bound

$$\ln p(\mathbf{y}) \geq E_q\left[\ln p(\theta)\right] - E_q\left[\ln q(\theta)\right] + E_q\left[\ln p(\mathbf{y}, \mathbf{x}|\theta)\right] - E_q\left[\ln q(\mathbf{x})\right] := \mathcal{L}(q(\theta), q(\mathbf{x}))$$

- If non-sequential

$$\mathcal{L} = E_{q(\theta)}\left[\ln p(\theta)\right] - E_{q(\theta)}\left[\ln q(\theta)\right] + \sum_{i=1}^{T} E_{q(x_i)}\left[\ln p(y_i, x_i|\theta)\right] - E_{q(\mathbf{x})}\left[\ln q(\mathbf{x})\right]$$

Stochastic Variational Inference for Hidden Markov Models

- Objective:

$$\ln p(\mathbf{y}) \geq E_q\left[\ln p(\theta)\right] - E_q\left[\ln q(\theta)\right] + E_q\left[\ln p(\mathbf{y}, \mathbf{x}|\theta)\right] - E_q\left[\ln q(\mathbf{x})\right] := \mathcal{L}(q(\theta), q(\mathbf{x}))$$

- For HMM, sample subchains $\mathbf{y}^S = (y_1^S, \ldots, y_L^S)$, and decompose the lower bound:

$$\ln p(\mathbf{y}, \mathbf{x}|\theta) = \ln \pi(x_1) + \sum_{t=2}^{T} \ln A_{x_{t-1}, x_t} + \sum_{i=1}^{T} \ln p(y_t|x_t).$$

- global update

$$E_S\left[E_q \ln p(\mathbf{y}^S, \mathbf{x}^S|\theta)\right] \approx p(S) E_q\left[\boxed{\sum_{t=1}^{T-L+1}} \ln \pi(x_t) + \boxed{(L-1)} \sum_{t=2}^{T} \ln A_{x_{t-1}, x_t} + L \sum_{t=1}^{T} p(y_t|x_t)\right]$$

- local update

$$q^*(\boxed{\mathbf{x}^S}) \propto \exp\left(E_{q(A)}\left[\ln \pi(x_1^S)\right] + \sum_{\ell=2}^{L} E_{q(A)}\left[\ln A_{x_{\ell-1}^S, x_\ell^S}\right] + \sum_{\ell=1}^{L} E_{q(\phi)}\left[\ln p(y_\ell^S | x_\ell^S)\right]\right)$$

Stochastic Variational Inference for Hidden Markov Models

- Improvements: Buffering subchains

Algorithm 2 `GrowBuf` procedure.

Input: subchain S, min buffer length $u \in \mathbb{Z}_+$, error tolerance $\epsilon > 0$.
Initialize $q^{\text{old}}(\mathbf{x}^S) = \texttt{ForwardBackward}(\mathbf{y}^S, \hat{\pi}, \widetilde{A}, \widetilde{p}_S)$ and set $S^{\text{old}} = S$.
while true **do**
  Grow $\boxed{\text{buffer } S^{\text{new}} \text{ by extending } S^{\text{old}}}$ by u observations in each direction.
  $q^{\text{new}}(\mathbf{x}^{S^{\text{new}}}) = \texttt{ForwardBackward}(\mathbf{y}^{S^{\text{new}}}, \hat{\pi}, \widetilde{A}, \widetilde{p}_{S^{\text{new}}})$, reusing messages from S^{old}.
  if $\boxed{\left|\left|q^{\text{new}}(\mathbf{x}^S) - q^{\text{old}}(\mathbf{x}^S)\right|\right|} < \epsilon$ **then**
    return $q^*(\mathbf{x}^S) = q^{\text{new}}(\mathbf{x}^S)$
  end if
  Set $S^{\text{old}} = S^{\text{new}}$ and $q^{\text{old}} = q^{\text{new}}$.
end while