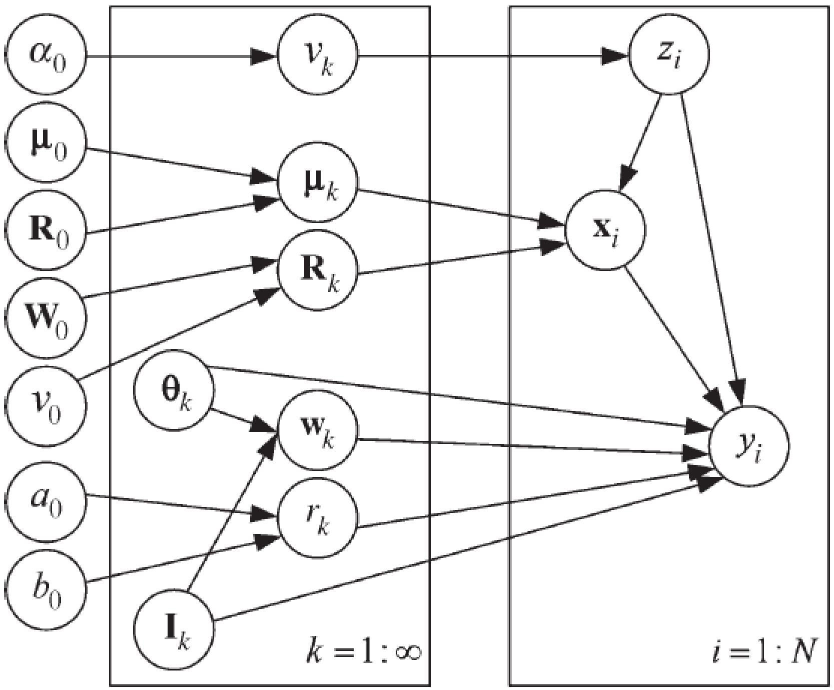
Variational Inference and MCMC Practice

2017/03/24

Mixtures of Gaussian Processes



$$p(\mathbf{x}|z=k,\boldsymbol{\mu}_k,\mathbf{R}_k) = \mathcal{N}\left(\mathbf{x}|\boldsymbol{\mu}_k,\mathbf{R}_k^{-1}\right)$$

$$\boldsymbol{\mu}_k \sim \mathcal{N}\left(\boldsymbol{\mu}_0,\mathbf{R}_0^{-1}\right), \quad \mathbf{R}_k \sim \mathcal{W}(\mathbf{W}_0,\nu_0)$$

$$p(y|\mathbf{x},z=k,\mathbf{w}_k,r_k) = \mathcal{N}(y|\mathbf{w}_k^\top\phi_k(\mathbf{x}),r_k^{-1})$$

$$\mathbf{w} \sim \mathcal{N}(\mathbf{w}_k|\mathbf{0},\mathbf{U}_k^{-1})\ \Gamma(r_k|a_0,b_0) \propto r_k^{a_0-1}e^{-b_0 r_k}$$

$$p(z|\mathbf{x}) = \frac{p(z)p(\mathbf{x}|z)}{\sum_z p(z)p(\mathbf{x}|z)}$$

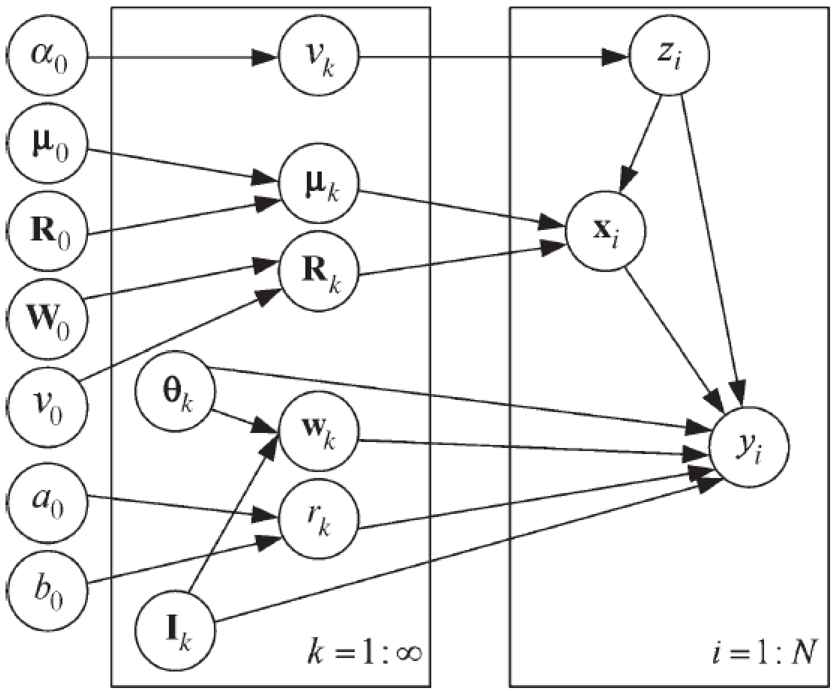
$$p(z) = \text{multinomial}\{\pi_1,\ldots,\pi_\infty\}$$

$$G = \sum_{i=1}^{\infty}\pi_i\delta_{\Phi_i} \qquad \pi_i = \nu_i\prod_{j=1}^{i-1}(1-\nu_j)$$

$$\nu_i \sim \text{Beta}(1,\alpha_0)$$

$$G \sim DP(\alpha_0, H) \qquad \Phi_i \sim H$$

MGP: Joint Distribution



$$
\begin{aligned}
p(\mathbf{D}, \Omega) &= p(\bar{\boldsymbol{\nu}})p(\bar{\boldsymbol{\mu}})p(\bar{\mathbf{R}})p(\bar{\mathbf{w}})p(\bar{r}) \\
&\qquad \times \prod_{i=1}^{N} p(z_i|\bar{\boldsymbol{\nu}})p(\mathbf{x}_i|z_i, \bar{\boldsymbol{\mu}}, \bar{\mathbf{R}})p(y_i|\mathbf{x}_i, z_i, \bar{\mathbf{w}}, \bar{r}) \\
&= \prod_{k=1}^{\infty} p(\nu_k)p(\boldsymbol{\mu}_k)p(\mathbf{R}_k)p(\mathbf{w}_k)p(r_k) \\
&\qquad \times \prod_{i=1}^{N} p(z_i|\bar{\boldsymbol{\nu}})p(\mathbf{x}_i|z_i, \bar{\boldsymbol{\mu}}, \bar{\mathbf{R}})p(y_i|\mathbf{x}_i, z_i, \bar{\mathbf{w}}, \bar{r})
\end{aligned}
$$

$$\ln p(\mathbf{D}) = \mathcal{L}(q) + \mathrm{KL}(q\|p)$$

$$q(\Omega) = \prod_{t=1}^{T-1} q(\nu_t) \prod_{k=1}^{T} q(\boldsymbol{\mu}_k)q(\mathbf{R}_k)q(\mathbf{w}_k)q(r_k) \prod_{n=1}^{N} q(z_n)$$

$$\ln q(\boldsymbol{\omega}) = \mathbb{E}_{\Omega\setminus\boldsymbol{\omega}}\left[\ln p(\mathbf{D}, \Omega)\right] + \mathrm{const}$$

MGP: Variational Inference

$$\ln q(\nu_t) = \ln p(\nu_t) + \boxed{\sum_{n=1}^{N} \mathbb{E}_{\Omega \setminus \nu_t} \left[\ln p(z_n | \bar{\boldsymbol{\nu}}) \right]} + \text{const.}$$

$$\begin{aligned} p(\mathbf{D}, \Omega) &= p(\bar{\boldsymbol{\nu}}) p(\bar{\boldsymbol{\mu}}) p(\bar{\mathbf{R}}) p(\bar{\mathbf{w}}) p(\bar{r}) \\ &\quad \times \prod_{i=1}^{N} p(z_i | \bar{\boldsymbol{\nu}}) p(\mathbf{x}_i | z_i, \bar{\boldsymbol{\mu}}, \bar{\mathbf{R}}) p(y_i | \mathbf{x}_i, z_i, \bar{\mathbf{w}}, \bar{r}) \\ &= \prod_{k=1}^{\infty} p(\nu_k) p(\boldsymbol{\mu}_k) p(\mathbf{R}_k) p(\mathbf{w}_k) p(r_k) \\ &\quad \times \prod_{i=1}^{N} p(z_i | \bar{\boldsymbol{\nu}}) p(\mathbf{x}_i | z_i, \bar{\boldsymbol{\mu}}, \bar{\mathbf{R}}) p(y_i | \mathbf{x}_i, z_i, \bar{\mathbf{w}}, \bar{r}) \end{aligned}$$

MGP: Variational Inference

$$\ln q(\nu_t) = \ln p(\nu_t) + \sum_{n=1}^{N} \mathbb{E}_{\Omega \backslash \nu_t} \left[\ln p(z_n | \bar{\boldsymbol{\nu}})\right] + \text{const}$$

$$\ln q(\boldsymbol{\mu}_k) = \ln p(\boldsymbol{\mu}_k) + \sum_{n=1}^{N} \mathbb{E}_{\Omega \backslash \boldsymbol{\mu}_k} \left[\ln p(\mathbf{x}_n | z_n, \bar{\boldsymbol{\mu}}, \bar{\mathbf{R}})\right] + \text{const}$$

$$\ln q(\mathbf{R}_k) = \ln p(\mathbf{R}_k) + \sum_{n=1}^{N} \mathbb{E}_{\Omega \backslash \mathbf{R}_k} \left[\ln p(\mathbf{x}_n | z_n, \bar{\boldsymbol{\mu}}, \bar{\mathbf{R}})\right] + \text{const}$$

$$\begin{aligned}\ln q(\mathbf{w}_k) &= \ln p(\mathbf{w}_k) \\ &\quad + \sum_{n=1}^{N} \mathbb{E}_{\Omega \backslash \mathbf{w}_k} \left[\ln p(y_n | \mathbf{x}_n, z_n, \bar{\mathbf{w}}, \bar{r})\right] + \text{const}\end{aligned}$$

$$\ln q(r_k) = \ln p(r_k) + \sum_{n=1}^{N} \mathbb{E}_{\Omega \backslash r_k} \left[\ln p(y_n | \mathbf{x}_n, z_n, \bar{\mathbf{w}}, \bar{\mathbf{r}})\right] + \text{const}$$

$$\begin{aligned}\ln q(z_n) + \text{const} = \mathbb{E}_{\Omega \backslash z_n} &\left[\ln p(z_n | \bar{\boldsymbol{\nu}}) + \ln p(\mathbf{x}_n | z_n, \bar{\boldsymbol{\mu}}, \bar{\mathbf{R}})\right. \\ &\left. + \ln p(y_n | \mathbf{x}_n, z_n, \bar{\mathbf{w}}, \bar{r})\right].\end{aligned}$$

$$\begin{aligned}p(\mathbf{D}, \Omega) &= p(\bar{\boldsymbol{\nu}}) p(\bar{\boldsymbol{\mu}}) p(\bar{\mathbf{R}}) p(\bar{\mathbf{w}}) p(\bar{r}) \\ &\quad \times \prod_{i=1}^{N} p(z_i | \bar{\boldsymbol{\nu}}) p(\mathbf{x}_i | z_i, \bar{\boldsymbol{\mu}}, \bar{\mathbf{R}}) p(y_i | \mathbf{x}_i, z_i, \bar{\mathbf{w}}, \bar{r}) \\ &= \prod_{k=1}^{\infty} p(\nu_k) p(\boldsymbol{\mu}_k) p(\mathbf{R}_k) p(\mathbf{w}_k) p(r_k) \\ &\quad \times \prod_{i=1}^{N} p(z_i | \bar{\boldsymbol{\nu}}) p(\mathbf{x}_i | z_i, \bar{\boldsymbol{\mu}}, \bar{\mathbf{R}}) p(y_i | \mathbf{x}_i, z_i, \bar{\mathbf{w}}, \bar{r})\end{aligned}$$

MGP: MCMC

- To infer the posterior $p(Z, \Theta | \mathcal{D})$:Gibbs sampling
  - Update indicator Z， sample from $p(z_i | Z_{-i}, \Theta, \mathcal{D})$
  - Update other hidden variables
    - Canonical distributions；
    - HMC；
    - Metropolis；
  - Optimize hyper-parameters