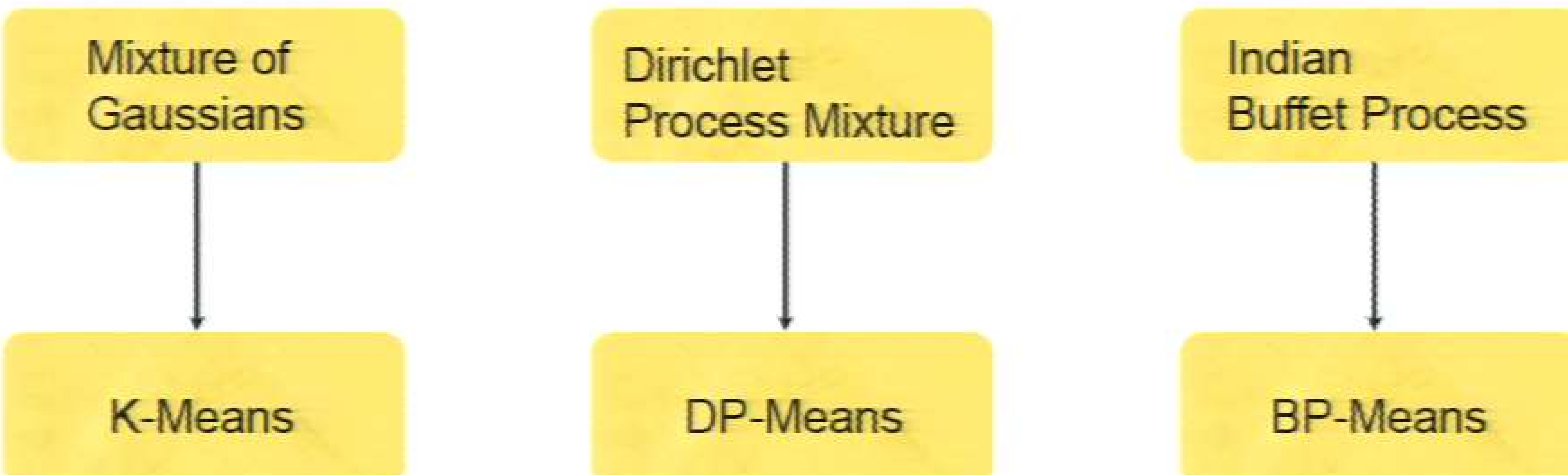
Small-Variance Asymptotics for Hidden Markov models

Anirban Roychowdhury, Ke Jiang, and Brian Kulis

Small-Variance Asymptotics has been used to learn fast and scalable non-probabilistic algorithms from parametric and nonparametric mixture models



THE OHIO STATE UNIVERSITY

SVA for the Finite-State Hidden Markov Model

- Joint distribution : $p(\mathcal{X}, \mathcal{Z}) = p(\boldsymbol{z}_1) \prod_{t=2}^{N} p(\boldsymbol{z}_t|\boldsymbol{z}_{t-1}) \prod_{t=1}^{N} \mathcal{N}(\boldsymbol{x}_t|\boldsymbol{\mu}_{z_t}, \sigma^2 I_d)$

- Bijection between exponential families and Bregman (KL) divergences

- Exponential (scaled) tx. probabilities $\Pr(\boldsymbol{z}_t|z_{t-1,j} = 1) = \exp(-\hat{\beta} d_{\phi}(\boldsymbol{z}_t, \boldsymbol{m}_j)) b_{\tilde{\phi}}(\boldsymbol{z}_t)$, where $\tilde{\phi} = \hat{\beta}\phi$

- SVA : $\min_{Z,\boldsymbol{\mu},T} \left(\sum_{t=1}^{N} \|\boldsymbol{x}_t - \boldsymbol{\mu}_{z_t}\|_2^2 + \lambda \sum_{t=2}^{N} \mathrm{KL}(\boldsymbol{z}_t, \boldsymbol{m}_{\boldsymbol{z}_{t-1}}) \right)$

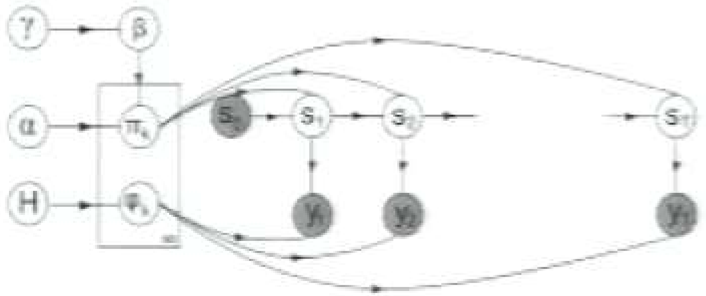
- Nonparametric Bayesian extension of the HMM

- HDP prior : $G_k \sim DP(\alpha, G_0), G_0 \sim DP(\gamma, H)$

- Joint distribution : $p(\mathcal{X}, \mathcal{Z}) \propto p(\mathcal{Z}|\alpha, \gamma, \lambda) \cdot p(\boldsymbol{z}_1) \prod_{t=2}^{N} p(\boldsymbol{z}_t|\boldsymbol{z}_{t-1}) \cdot \prod_{t=1}^{N} \mathcal{N}(\boldsymbol{x}_t|\boldsymbol{\mu}_{z_t}, \sigma^2 I_d) \cdot p(\boldsymbol{\mu}_{1:K})$

- SVA result : K-Means like objective function with three penalties:-

$$\min_{K,Z,\boldsymbol{\mu},T} \quad \sum_{t=1}^{N} \|\boldsymbol{x}_t - \boldsymbol{\mu}_{z_t}\|^2 + \lambda \sum_{t=2}^{N} \mathrm{KL}(\boldsymbol{z}_t, \boldsymbol{m}_{\boldsymbol{z}_{t-1}}) + \lambda_1 \sum_{k=1}^{K} (s_k - 1) + \lambda_2 (K - 1)$$



Van Gael et al., Beam Sampling for the Infinite Hidden Markov Model, ICML 2008

Some Results