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

- Mixture of Gaussians
  - K-Means
- Dirichlet Process Mixture
  - DP-Means
- Indian Buffet Process
  - BP-Means
SVA for the Finite-State Hidden Markov Model

- Joint distribution: 
  \[ p(\mathcal{X}, Z) = p(z_1) \prod_{t=2}^{N} p(z_t | z_{t-1}) \prod_{t=1}^{N} \mathcal{N}(x_t | \mu_{z_t}, \sigma^2 I_d) \]

- Bijection between exponential families and Bregman (KL) divergences

- Exponential (scaled) tx. probabilities 
  \[ \Pr(z_t | z_{t-1,j} = 1) = \exp(-\tilde{\beta} d_\phi(z_t, m_j)) b_\phi(z_t), \]
  where \( \tilde{\phi} = \tilde{\beta} \phi \)

- SVA: 
  \[ \min_{Z, \mu, T} \left( \sum_{t=1}^{N} \| x_t - \mu_{z_t} \|^2_2 + \lambda \sum_{t=2}^{N} \text{KL}(z_t, m_{z_{t-1}}) \right) \]
SVA for the Infinite Hidden Markov Model

- Nonparametric Bayesian extension of the HMM

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

- Joint distribution: $p(X, Z) \propto p(Z|\alpha, \gamma, \lambda) \cdot p(z_1) \prod_{t=2}^{N} p(z_t|z_{t-1}) \cdot \prod_{t=1}^{N} N(x_t|\mu_{z_t}, \sigma^2 I_d) \cdot p(\mu_{1:K})$

- SVA result: K-Means like objective function with three penalties:
  $$\min_{K, Z, \mu, T} \sum_{t=1}^{N} \|x_t - \mu_{z_t}\|^2 + \lambda \sum_{t=2}^{N} KL(z_t, m_{z_{t-1}}) + \lambda_1 \sum_{k=1}^{K} (s_k - 1) + \lambda_2 (K - 1)$$

\[\text{Van Gael et al., Beam Sampling for the Infinite Hidden Markov Model, ICML 2008}\]
Some Results

- **NMI and Error** vs Lambda
- **Training Accuracy** vs # of states
- **Training Time** vs # of states
- **Indices** vs Trading day offset from 12/28/1999

Graphs comparing different states and models:
- AsymHMM
- Beam10
- Beam100
- Beam1000

Models include:
- AHMM prediction
- HMM prediction
- Market actual
- AiHMM prediction