Project Report, CPSC-540

Bita Nejat 45113115
Sohrab Salehi 86711132
Xi Laura Cang 40460024

November 21, 2014

1 Intro

1.1 Dirichlet Process Mixture Models

- brief description - establish notation.
- the inference problem

We would like to sample from the posterior:

$$p(\eta_1, \eta_2, ..., \eta_N | y_1, y_2, ..., y_N)$$

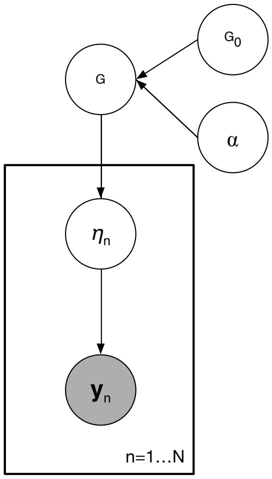

Figure 1: PGM for DPMMs

1.2 Chinese Restaurant Process Representation

- brief description
- generative model
- exact inference
  - posterior distribution
  - predictive distribution

- Sampling scheme
  - Conjugate case Gaussian-Gaussian mean /Gamma variance Categorical data
    - Non-collapsed
    - Collapsed
  - Non-conjugate case
- Sampling Algorithm Pseudocode [**?**]

1.3 Stick Breaking Representation

- brief description
- hierarchical model
- Sampling Algorithm Pseudocode

2 Experiments

- One-dimensional data
- Higher dimensions

2.1 Synthetic Data

2.2 Real Data

2.3 Results

- describe clustering accuracy measures used
- scatter clustering plot for CRP
- scatter clustering plot for SBR

3 Discussion

3.1 Which one was better?

3.2 Future Work

4 Appendix

Sample new cluster assignments as follows:

Algorithm 1 Rao-Blackwellaized Gibbs Sampler for DPMMs CRP Representation [**?**]

```
Require: $z^{t-1} \geq 0 \vee Kcurrentclusterstatistics$
  (1) $\phi\{1...N\} \sim perm(\{1...N\})$
  (2) $z^t \leftarrow z^{t-1}$
  for $i \in \{\phi(1), \phi(2), ..., \phi(N)\}$ do
    (a)
    for each of the $K$ existing clusters, determine predictive likelihood do
      $f_k(x_i) = p(x_i|\{x_j|z_j = k, j \neq i\}, \lambda)$
    end for
    $f_{\bar{k}}(x_i) = p(x_i|z_i = \bar{k}, z_{-i}, x_{-i}, \lambda) = p(x_i|\lambda)$ // reference in the text as how to calculate this
    (b) $z_i \sim \frac{1}{Z_i}(\alpha f_{\bar{k}}(x_i)\delta(z_i, \bar{k}) + \sum_{k=1}^{K} N_k^{-i} f_k(x_i)\delta(z_i, k))$ where
    $Z_i = \alpha f_{\bar{k}}(x_i) + \sum_{k=1}^{K} N_k^{-i} f_k(x_i)$ and $N_k^{-i} = \#\{x_j : z_j = k\}$
    (c) Update cached sufficient statistics to reflect the assignment of $x_i$ to cluster $z_i$
    if $z_i = \bar{k}$ then
      Create a new cluster
      $K \leftarrow K + 1$
    end if
  end for
  (3) set $z_t \leftarrow z$ // sample mixture parameters current clusters via step 3 of alg 2.1. [? ]
  (4) $K \leftarrow K - \#\{k : N_k == 0\}$
```