Lecture 14: Dirichlet Process Mixtures in a nutshell

Marina Meilă
mmp@stat.washington.edu

Department of Statistics
University of Washington

March 9, 2018

The Chinese Restaurant Process

Dirichlet Process Mixture

The Chinese Restaurant Process

- **Given** parameters $\alpha > 0$, G_0, with G_0 a continuous measure on measurable space $(\Theta, \mathcal{B})$.
- Assume we already have samples $\theta_{1:n} \in \Theta$.
- The probability of θ_{n+1} is then

$$\theta_{n+1} \mid \theta_{1:n} \sim \sum_{k=1}^{K} \frac{K}{n+\alpha} \delta_{\theta_k} + \frac{\alpha}{n+\alpha} G_0. \tag{1}$$

  In the above, K represents the number of distinct values among the n samples $\theta_{1:n}$.
- This defines a **Chinese Restaurant Process (CRP)**. It is easy to see that the process is exchangeable.
- One can also prove that for $n \to \infty$, $\theta_{1:n} \to G$ where $G \sim DP(\alpha, G_0)$.

Dirichlet Process

- A **Dirichlet Process (DP)** is distribution over measures.
- Let $(\Theta, \mathcal{B})$, α, G_0 be as above.
- We say that the random function G is drawn from $DP(\alpha, G_0)$ iff

$$\text{for any partition } B_{1:K} \subset \mathcal{B} \text{ of } \Theta,\ G(B_{1:K}) \sim Dirichlet(\alpha G_0(B_{1:K})). \tag{2}$$

Dirichlet Process Mixture

- Given: $DP(\alpha, G_0)$, family of distributions $\{f_\theta\}$ on $\mathcal{X}$.
- For $i = 1, 2, \ldots n$

$$\theta_i \sim CRP(\alpha, G_0, \theta_{1:i-1}) \qquad (3)$$

$$x_i \sim f_{\theta_i} \qquad (4)$$

Estimation of DP Mixture by Gibbs sampling

Input $\alpha, G_0, \{f\}, \mathcal{D} = \{x_1, \dots x_n\} \subseteq \mathcal{X}$

State cluster assignments c_i, $i = 1 : n$,
parameters θ_k for all distinct k

Iterate

1. for $i = 1 : n$(reassign data to clusters)
   1.1 if $n_{c_i} = 1$ delete this cluster and its θ_{c_i}
   1.2 resample c_i by

$$c_i = \begin{cases} \text{existing } k & \text{w.p} \propto \frac{n_k}{n-1+\alpha} f(x_i, \theta_k) \\ \text{new cluster} & \text{w.p} \frac{\alpha}{n-1+\alpha} \int f(x_i, \theta) G_0(\theta) d\theta \end{cases} \quad (5)$$

   1.3 if c_i is new label, sample a new θ_{c_i} from f_θ $G_0(\theta)$
2. (resample cluster parameters)
   for $k \in \{c_{1:n}\}$
   2.1 sample θ_k from posterior $f_{\theta_k} \propto G_0(\theta) \prod_{i \in C_k} f(x_i, \theta)$

   this can be computed in closed form if G_0 is conjugate prior

Output a state with high posterior