The Dirichlet process and clustering

Consider first a finite mixture model with K clusters, e.g., a GMM:

all 1 vector with K comp.

$$p(\pi|\alpha) = \mathrm{Dir}(\pi; \tfrac{\alpha}{K} \mathbb{1}_K)$$

$$p(\theta^k|\lambda) = H(\theta^k; \lambda)$$ conjugate prior to

$$p(z_i = k|\pi) = \pi_k$$

$$p(x_i|z_i = k, \theta) = p(x_i|\theta^k)$$

We can write

$$\theta^k \sim H(\lambda)$$

$$x_i \sim F(\theta^{z_i})$$

Instead we can view this as

$\sim \mathrm{Dir}$ $\sim H$ as above

$$\theta_i \sim G(\theta) = \sum_{k=1}^{K} \pi_k \delta_{\theta^k}(\theta)$$ Notice θ_i data, θ^k cluster

$$x_i \sim F(\theta_i).$$

firs part

This gives sh clusters. DP generaliz to "any" k.

We define the DP using 2 concrete processes.

<u>The stickbreaking process</u>

Repeatedly:

(1) break a stick, using a Beta$(1,\alpha)$ dist.
(2) continue with non-picked part.

That is,

- start with length 1 stick

for k=1 to ∞
- pick new fraction $\beta_k \sim \mathrm{Beta}(1,\alpha) \propto (1-x)^{\alpha-1}$
- let π_k be the absolut length

$$\pi_k = \beta_k \underbrace{\prod_{\ell=1}^{k-1}(1-\beta_\ell)}_{\text{left by earlier breaks}} = \beta_k\left(1-\sum_{\ell=1}^{k-1}\pi_\ell\right)$$

This is denoted

$$\pi \sim GEM(\alpha)$$

By adding

$$\theta^k \sim H(\lambda)$$

$$G(\theta)=\sum_{k=1}^{\infty}\pi_k\,\delta_{\theta^k}(\theta)$$

we get $G \sim DP(\alpha, H)$.

Note, by construction discrete.

We will tend to se the same θ^k many times, this induces the clusters.

Alternative: The Chinese restaurant process.

Let

$$p(\theta_{N+1} = \theta \mid \theta_{1:N}, \alpha, H) = \frac{1}{\alpha + N}\left(\alpha H(\theta) + \sum_{k=1}^{K} N_k \, \delta_{\theta^k}(\theta)\right)$$

where

K = # different observations

N_k = # obs. of θ^k in $\theta_{1:N}$.

It is also possible to introduce a class variable z_i (so, $\theta_i = \theta^{z_i}$). Then

$$p(z_{N+1} = k \mid z_{1:N}, \alpha) = \begin{cases} \frac{\alpha}{\alpha + N} & \text{if } k \notin [K] \\ \frac{N_k}{\alpha + N} & \text{otherwise} \end{cases}$$

Interpretation

- $[K]$ tables / clusters
- new customer sits with others or at new table

• "large" tables have large prob. of growing.

Theo

$$P(\#\text{tables} \propto \alpha \log N) \xrightarrow{N \to \infty} 1.$$

DP mixture model

Up to now, we have generated clusters, or a distribution for each cluster.

Now also data from each sample cluster:

$$\pi \sim GEM(\alpha)$$
$$z_i \sim \pi \quad (P(z_i = k \mid \pi) = \pi_k$$
$$\theta^k \sim H(\lambda)$$
$$x_i \sim F(\theta^{z_i}).$$

A collapsed Gibbs sampler for DPMM.

As before the states are of the form $z = (z_1, \dots, z_N)$.

Notice

$$p(z_i = k \mid z_{-i}, X, \alpha, \lambda) \propto \underbrace{p(z_i = k \mid z_{-i}, \alpha)}_{1} \underbrace{p(x_i \mid X_{-i}, z_i = k, z_{-i}, \lambda)}_{2}$$

1

By interchangability (order of sampling irrelevant) we can assume that i is the last customer.

Hence,

$$\begin{cases} \frac{N_{k,-i}}{\dots} & \text{if } k \in [K] \end{cases}$$

$$p(z_i = k \mid z_{-i}, \alpha) = \begin{cases} \dfrac{\;}{\alpha + N - 1} & \\ \dfrac{\alpha}{\alpha+N-1} & \text{otherwise} \end{cases}$$

where

$K = \#$ cluster for z_{-i}

$N_{k,-i} = \#$ of z_{-i} in cluster k.

2

Let $x_{k,-i} := \{x_j : z_j = k, j \neq i\}$.

Notice,

$$p(x_i \mid x_{-i}, z_{-i}, z_i = k, \lambda) = p(x_i \mid \bar{x}_{k,-i}, \lambda) = \frac{p(x_i, x_{k,-i} \mid \lambda)}{p(x_{k,-i} \mid \lambda)}$$

where den. and num. are marginal likelihood of data.

E.g.

$$p(x_i, x_{k,-i} \mid \lambda) = \int p(x_i \mid \theta) \prod_{j \neq i : z_j = k} p(x_j \mid \theta) H(\theta \mid \lambda)\, d\theta$$

If $z_i = K+1$ (a new cluster), then

$$p(x_i \mid x_{-i}, z_{-i}, z_i = k, \lambda) = \int p(x_i \mid \theta) H(\theta \mid \lambda)\, d\theta.$$

So for conj. prior H of F, we are typically ok.

Gaussian F with fixed variance and uniform H on its exp.

$$\int_0^1 \prod_i \mathcal{N}(x_i; \mu, \sigma^2)\, d\mu$$

fixed

can be solved numerically.