

Methods

Jiahao XU

May 10, 2018

Methods

We have N observations X_1, \dots, X_N , and for each X_i , we have following assumption with fixed σ^2 .

$$X_i \sim \mathcal{N}(\mu_i, \sigma^2)$$

We consider X_i and X_j should be in one cluster, if $\mu_i = \mu_j$. We want to propose a prior \mathcal{H} for μ . If \mathcal{H} is a continuous distribution, we have $Pr(\mu_i = \mu_j) = 0$, which is infeasible for clustering. Therefore, we introduce a discrete approximation for our prior by using Dirichlet Process(DP).

Dirichlet Process (DP)

We have a measure space (Θ, Σ) . Define a measurable finite partitioning of Θ to be a finite collection of sets A_1, A_2, \dots, A_K such that:

- (1) Finite: $K < \infty$. (2) Measureable: $A_k \in \Sigma$. (3) Disjoint: $A_j \cap A_k = \emptyset, \forall j \neq k$. (4) Complete: $\cup_k A_k = \Theta$.

A Dirichlet process is a random probability measure G over a (Θ, Σ) with property that given any measurable finite partitioning of Θ , we have

$$[G(A_1), \dots, G(A_K)] \sim \text{Dirichlet}(\alpha H(A_1), \dots, \alpha H(A_K))$$

where α is scale, and H is base measure, and $G \sim DP(\alpha, H)$ will be discrete. [1]

Dirichlet Process Gaussian Mixture Model (DPGMM)

With introduction of DP, we can reformulate our model as following with given $\sigma^2, \alpha, \mu_0, \sigma_0^2$.

$$\begin{aligned} X_i &\sim \mathcal{N}(\mu_i, \sigma^2) \\ \mu_i | G &\sim G \\ G &\sim DP(\alpha, \mathcal{N}(\mu_0, \sigma_0^2)) \end{aligned}$$

Notation

We slice space of μ into K partitions, and use Z_i to indicate which partition μ_i falls in, which is our cluster assignment for X_i .

$$\begin{aligned} Z_i = k &\Leftrightarrow \mu_i \text{ in } k\text{th partition} \quad \text{where } k \in [1, 2, \dots, K] \\ p_k &\triangleq P(Z_j = k) \quad \forall j \in [1, 2, \dots, N] \\ p &\triangleq \{p_1, \dots, p_K\}, \quad -i \triangleq \{1, 2, \dots, i-1, i+1, \dots, N\} \\ n_{k,-i} &= \text{count of } j \quad \text{s.t. } Z_j = k \text{ and } j \neq i \end{aligned}$$

If we have infinite partitions $K \rightarrow \infty$, we can rewrite our DPGMM model as following:

$$\begin{aligned} X_i | \mu &\sim \mathcal{N}(\mu_i, \sigma^2) \\ Z_i | p &\sim \text{Discrete}(p_1, \dots, p_K) \\ \mu_i &\sim \mathcal{N}(\mu_0, \sigma_0^2) \\ p &\sim \text{Dir}(\alpha/K, \dots, \alpha/K) \end{aligned}$$

Predictive Distribution

We have following predictive distribution for Z_i , and detail of result derivation is discussed in appendix A.

$$\begin{aligned} P(Z_i = m | Z_{-i}) &= \frac{P(Z_i = m, Z_{-i})}{P(Z_{-i})} \\ &= \frac{\int_p P(Z_i = m, Z_{-i} | p_1, \dots, p_K) P(p_1, \dots, p_K) dp}{\int_p P(Z_{-i} | p_1, \dots, p_K) P(p_1, \dots, p_K) dp} \quad [A.1] \\ &= \frac{\frac{\alpha}{K} + n_{m,-i}}{\alpha + N - 1} \end{aligned}$$

Chinese Restaurant Process (CRP)

Based on pervious predictive distribution, when $K \rightarrow \infty$, we have Chinese restaurant process.

$$\begin{aligned} P(Z_i = m | Z_{-i}) &= \frac{n_{m,-i}}{\alpha + N - 1} \quad \text{existing cluster } m \\ P(Z_i = \text{new} | Z_{-i}) &= \frac{\alpha}{\alpha + N - 1} \quad \text{new cluster} \end{aligned}$$

Gibbs Sampler for DPGMM

Based on exchangeability, we have following predictive probability for Gibbs sampling [2] [3] [4]. Detail of derivation for new cluster case is discussed in appendix A.

$$\begin{aligned} P(Z_i = m | Z_{-i}, X) &\propto P(Z_i = m | Z_{-i}, \alpha) \cdot P(X_i | Z_i = m, \mu_i) \\ &\propto \begin{cases} n_{m,-i} \cdot \mathcal{N}(x_i; \mu_{[m]}, \sigma^2) & \text{existing cluster } m \\ \alpha \int_{\mu} Pr(X_i | \mu) \cdot Pr(\mu | \mu_0) = \alpha \mathcal{N}(x_i; \mu_0, \sigma^2 + \sigma_0^2) & \text{new cluster} \end{cases} \quad [A.2] \\ &\text{where } \mu_{[m]} = \mu \text{ of cluster } m \end{aligned}$$

Sampling Algorithm

Initialization

Assign all data in one cluster s.t. $Z_1^{(0)} = Z_2^{(0)} = \dots = Z_n^{(0)} = 1$, and $K^{(0)} = 1$. Sample $\mu_{[1]}^{(0)}$ based on posterior of $\mu | X, Z$ ([A.3]), where n_l is count for all $Z_l = j$, detail discussed in appendix A.

Run detail

For i in $[1, \dots, N]$ sample $Z_i^{(t+1)}$ based on

$$P(Z_i^{(t+1)} = m) \propto \begin{cases} n_{m,-i}^{(t)} \cdot \mathcal{N}(x_i; \mu_{[m]}^{(t)}, \sigma^2) & \text{existing cluster } m \\ \alpha \mathcal{N}(x_i; \mu_0, \sigma^2 + \sigma_0^2) & \text{new cluster s.t. } m = K + 1 \end{cases}$$

$K = K + 1$ each time if our sampled assignment is new. After sampling assignment, we sample $\mu_{[k]}^{(t+1)}$ for all $k \in [1, \dots, K]$ based on

$$P(\mu_{[k]}^{(t+1)} | X_l, \text{ where } Z_l^{(t+1)} = 1) \sim \mathcal{N}\left(\frac{\sum x_l + \frac{\mu_0}{\sigma_0^2}}{\frac{n_k^{(t+1)}}{\sigma^2} + \frac{1}{\sigma_0^2}}, \left[\frac{n_k^{(t+1)}}{\sigma^2} + \frac{1}{\sigma_0^2}\right]^{-1}\right) \quad [A.3]$$

Appendix A

Reference

- [1] Ferguson. (1973). "A Bayesian Analysis of Some Nonparametric Problems" Annals of Statistics
- [2] David M. Blei, Michael I. Jordan. (2006). "Variational Inference for Dirichlet Process Mixtures" Bayesian Analysis
- [3] Radford M. Neal. (2000). "Markov Chain Sampling Methods for Dirichlet Process Mixture Models" Journal of Computational and Graphical Statistics
- [4] Samuel Harris. (2015) " Dirichlet Process Gaussian Mixture Model Gibbs Sampler for a 1-dimensional Behavioural Time Series Segmentation"