Bayesian nonparametric latent feature models

François Caron

UBC

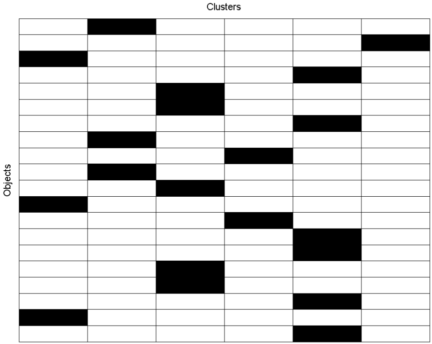
October 2, 2007 / MLRG

Overview

Overview

Finite mixture model

- n objects and K clusters
- Each object $k = 1, \ldots, n$ can belong to only one cluster $j \in \{1, \ldots, K\}$
- Represented by an allocation variable $c_k \in \{1, \ldots, K\}$



Finite mixture model

- Bayesian approach: Dirichlet-multinomial model

$$\pi_{1:K} \sim \mathcal{D}(\frac{\alpha}{K}, \dots, \frac{\alpha}{K})$$

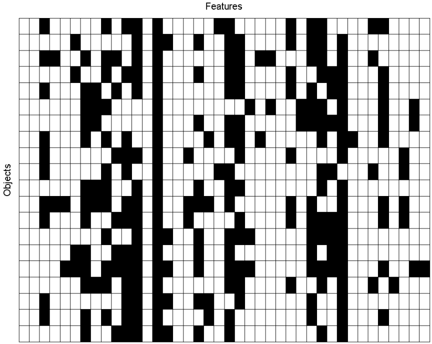
  and for $k = 1, \dots, n$,

$$c_k \sim \pi_{1:K}$$

- $K \to \infty$: Dirichlet Process

Finite latent feature model

- n objects and K features
- Each object k can have a finite number of latent features
- Represented by a binary vector $\mathbf{c}_k \in \{0, 1\}^K$



Finite latent feature model

- Bayesian approach: Prior distribution over a binary matrix of size $n \times K$
- $K \to \infty$: Nonparametric latent feature model

Overview

Finite mixture model

- Hierarchical model

$$\pi_{1:K} \sim \mathcal{D}(\frac{\alpha}{K}, \ldots, \frac{\alpha}{K})$$

- For $j = 1, \ldots, K$,

$$U_j \sim \mathbb{G}_0$$

- For $k = 1, \ldots, n$,

$$c_k \sim \pi_{1:K}$$

$$\mathbf{z}_k | c_k, U_{1:K} \sim f(\cdot | U_{c_k})$$

Finite mixture model

- Prior distribution

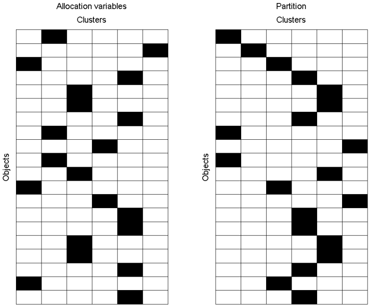
$$\Pr(\pi_{1:K}, U_{1:K}, c_{1:n}) = \Pr(\pi_{1:K}, c_{1:n}) \Pr(U_{1:K})$$

- Can integrate out analytically $\pi_{1:K}$ (Dirichlet-multinomial distribution)

$$\begin{aligned} \Pr(c_{1:n}) &= \int \Pr(\pi_{1:K}, c_{1:n}) d\pi_{1:K} \\ &= \frac{\Gamma(\alpha) \prod_{j=1}^{K} \Gamma(n_j + \frac{\alpha}{K})}{\Gamma(\alpha + n)\Gamma(\frac{\alpha}{K})^K} \end{aligned}$$

Distribution over partitions

- Let $\Pi_n = \{A_1, \ldots, A_{n(\Pi_n)}\}$ be a random partition of $\{1, \ldots, n\}$ and $\mathcal{P}_n$ the set of partitions of $\{1, \ldots, n\}$
- Ex: $\Pi_5 = \{\{1, 2, 5\}, \{3\}, \{4\}\}$
- Several different values of $c_{1:n}$ may induce the same partition
- Ex: $c_{1:5} = (1, 1, 2, 1, 3)$ and $c'_{1:5} = (2, 2, 3, 2, 1)$ both induce the same partition $\Pi_5 = \{\{1, 2, 4\}, \{3\}, \{5\}\}$
- Let $\Pi_n(c_{1:n})$ be the partition of $\{1, \ldots, n\}$ induced by the equivalence relationship $i \leftrightarrow j \iff c_i = c_j$



Distribution over partitions

- We have

$$\Pr(\Pi_n(c_{1:n})) = \frac{K!}{(K - n(\Pi_n))!} \Pr(c_{1:n})$$

for all $\Pi_n \in \mathcal{P}_K$ where $\mathcal{P}_K = \{\Pi_n \in \mathcal{P}_n | n(\Pi_n) \leq K\}$

- and thus for the Dirichlet-multinomial distribution

$$\Pr(\Pi_n) = \frac{K!}{(K - n(\Pi_n))!} \frac{\Gamma(\alpha) \prod_{j=1}^{K} \Gamma(n_j + \frac{\alpha}{K})}{\Gamma(\alpha + n)\Gamma(\frac{\alpha}{K})^K}$$

Dirichlet Process

- Limit when $K \to \infty$ of the finite model

$$\Pr(\Pi_n) = \frac{K!}{(K - n(\Pi_n))!} \frac{\Gamma(\alpha) \prod_{j=1}^{K} \Gamma(n_j + \frac{\alpha}{K})}{\Gamma(\alpha + n)\Gamma(\frac{\alpha}{K})^K}$$

- is given by

$$\Pr(\Pi_n) = \frac{\alpha^{n(\Pi_n)} \prod_{j=1}^{n(\Pi_n)} \Gamma(n_j)}{\prod_{i=1}^{n} (\alpha + i - 1)}$$

Dirichlet Process

- Connexion to the usual formulation

$$\mathbb{G} \sim DP(\alpha, \mathbb{G}_0)$$

for $k = 1, \ldots, n$

$$\theta_k | \mathbb{G} \sim \mathbb{G}$$

- Let $\Pi_n(\theta_1 \ldots, \theta_n)$ be the partition of $\{1, \ldots, n\}$ induced by the equivalence relationship $i \leftrightarrow j \iff \theta_i = \theta_j$ and $U_1, \ldots, U_{n(\Pi_n)}$ be the different values taken by $\theta_1, \ldots, \theta_n$
- When integrating out the unknown distribution $\mathbb{G}$

$$\Pr(\theta_1, \ldots, \theta_n) = \Pr(\Pi_n(\theta_1 \ldots, \theta_n)) \prod_{j=1}^{n(\Pi_n)} \mathbb{G}_0(U_j)$$

Chinese Restaurant

- Let Π_{-k} be the partition obtained by removing item k from Π_n
- Conditional association probability of item k is

$$\frac{\Pr(\Pi_n)}{\Pr(\Pi_{-k})}$$

- Item k is associated to an existing cluster $j = 1, \ldots, n(\Pi_{-k})$ with probability

$$\frac{n_{j,-k}}{n - 1 + \alpha}$$

and to a new cluster with probability

$$\frac{\alpha}{n - 1 + \alpha}$$

Stick-breaking representation

- Let $\Pi_1, \Pi_2, \ldots$ be the successive partitions obtained with sequential CRP updates
- Ex: $\Pi_1 = \{\{1\}\}$, $\Pi_2 = \{\{1\},\{2\}\}$, $\Pi_3 = \{\{1,3\},\{2\}\}$, $\Pi_4 = \{\{1,3\},\{2\},\{4\}\}, \ldots$
- Let $n_{j,t}$ the size of cluster j in Π_t, where the clusters are numbered in order of appearance
- Let $\lim_{t\to\infty} \frac{n_{j,t}}{t} = \pi_j$. We have

$$\pi_j = \beta_j \prod_{i=1}^{j-1} (1 - \beta_i)$$

where

$$\beta_j \sim \mathcal{B}(1, \alpha)$$

- Stick-breaking construction or residual allocation model

Overview

Finite feature model

- Assume the following model
- For each feature $j = 1, \ldots, K$,

$$\pi_j \sim \mathcal{B}(\frac{\alpha}{K}, 1)$$

- For each object k and each feature j

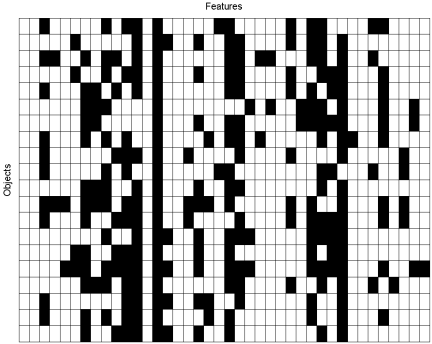
$$c_{k,j} \sim \text{Ber}(\pi_j)$$

Finite feature model

- Integrating out $\pi_{1:K}$

$$\Pr(c_{1:n}) = \prod_{j=1}^{K} \frac{\frac{\alpha}{K}\Gamma(m_j + \frac{\alpha}{K})\Gamma(n - m_j + 1)}{\Gamma(n + 1 + \frac{\alpha}{K})}$$

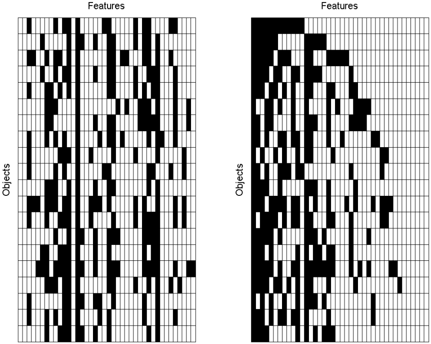
- Invariant to permutation of objects/features



Equivalent classes

- Distribution over binary matrix is invariant with respect to permutations of columns
- Left-ordered form

$$\Pr(\xi(\mathbf{c}_{1:n})) = \frac{K!}{\prod_{h=0}^{2^{n-1}} K_h!} \prod_{j=1}^{K} \frac{\frac{\alpha}{K}\Gamma(m_j + \frac{\alpha}{K})\Gamma(n - m_j + 1)}{\Gamma(n + 1 + \frac{\alpha}{K})}$$



Nonparametric latent feature model

- Limit when $K \to \infty$ of the finite model

$$\Pr(\xi(\mathbf{c}_{1:n})) = \frac{K!}{\prod_{h=0}^{2^{n-1}} K_h!} \prod_{j=1}^{K} \frac{\frac{\alpha}{K}\Gamma(m_j + \frac{\alpha}{K})\Gamma(n - m_j + 1)}{\Gamma(n + 1 + \frac{\alpha}{K})}$$

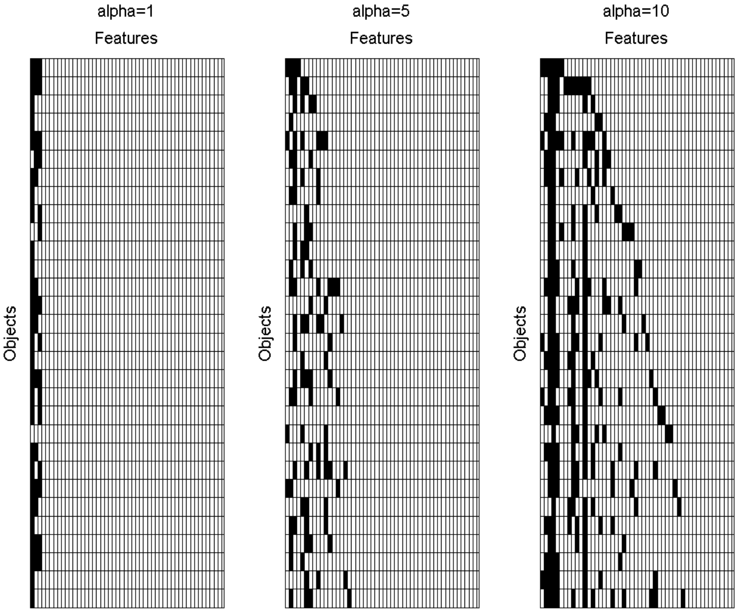
is given by

$$\Pr(\xi(\mathbf{c}_{1:n})) = \frac{K^+}{\prod_{h=1}^{2^n-1} K_h!} \exp(-\alpha H_n) \prod_{j=1}^{K^+} \frac{(n - m_j)!(m_j - 1)!}{n!}$$

Indian Buffet

- Buffet with infinite number of dishes
- First customer samples Poisson(α) dishes
- k^{eme} customer takes each dish with probability m_j/k and tries Poisson(α/k) new dishes
- Caution: Not in left-ordered form!

Indian Buffet



Properties

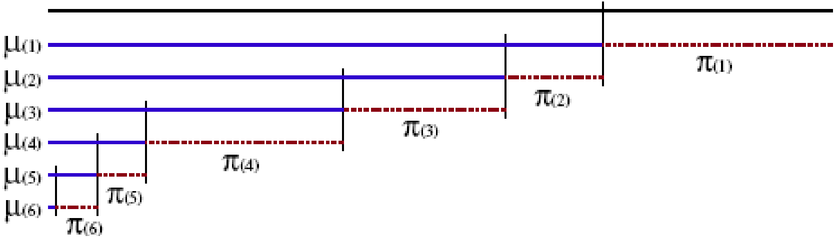
- Total number of different features K^+ follows Poisson$(\alpha(\sum_{k=1}^{n} \frac{1}{k}))$
- Number of features possessed by each object follows Poisson(α)
- Total number of entries in the matrix follows Poisson$(n\alpha)$

Stick-breaking construction

- Let $\pi_{(1)} > \pi_{(2)} > \ldots > \pi_{(K)}$ be a decreasing ordering of $\pi_{1:K}$ where $\pi_k \sim \mathcal{B}(\frac{\alpha}{K}, 1)$
- Limit when $K \to \infty$: stick-breaking contruction

$$\beta_{(k)} \sim \mathcal{B}(\alpha, 1) \text{ and } \pi_{(k)} = \pi_{(k-1)}\beta_{(k)}$$

- DP: stick lengths sum to one and are not decreasing (only in average)
- IBP: stick lengths does not sum to one and are decreasing



Overview

Applications

- Choice behaviour
- Protein interaction screens
- Structure of causal graphs

Bibliography

- Random partitions and Dirichlet Processes

J. Pitman.
Exchangeable and partially exchangeable random partitions
Probability Theory and Related Fields, 1995.

J. Pitman.
Some developments of the Blackwell-MacQueen urn scheme
In Statistics, probability and game theory; paper in honour of David Blackwell, 1996.

- Indian buffet

Z. Ghahramani, T.L. Griffiths and P. Sollich.
Bayesian nonparametric latent feature models
Proc. Valencia / ISBA 8th World Meeting on Bayesian Statistics, 2006.

T.L. Griffiths and Z. Ghahramani.
Infinite latent feature models and the Indian Buffet Process
Technical Report, University College, London, 2005.

Y.W. Teh, D. Gorur and Z. Ghahramani.
Stick-breaking construction for the Indian Buffet Process

Bibliography

- Applications of Indian buffet

D. Gorur, F. Jakel and C. Rasmussen.
A choice model with infinitely many latent features.
ICML 2006.

W. Chu, Z. Ghahramani, R. Krause and D.L. Wild.
Identifying protein complexes in high-throughput protein interactions screens using an infinite latent feature model.
Biocomputing 2006.

F. Wood, T.L. Griffiths and Z. Ghahramani.
A nonparametric Bayesian method for inferring hidden causes.
UAI 2006.

- Still hungry?

M. McGlohon.
Fried Chicken Bucket Processes
SIGBOVIK, Pittsburgh, 2007.