The Indian Buffet Process

Arthur Lui

Department of Statistics
Brigham Young University

10 December 2014

Introduction

- One challenge in latent variable modelling is predetermining the number of latent variables generating the observations
- One type of latent variable model is the latent *feature* model, which consists of quantitative observations and binary latent variables
- The Indian Buffet Process offers a distribution for sparse infinite binary matrices, which can serve as a prior distribution for feature matrices in latent feature models

Indian Buffet Process (IBP)

Figure 1: IBP(N=50,a=2)

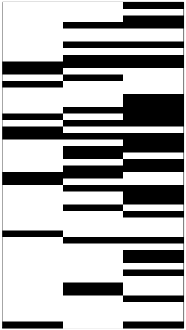
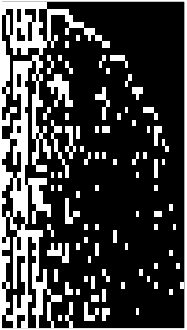


Figure 2: IBP(N=50,a=10)



Indian Buffet Process (IBP)

In the IBP, a customer (i) taking a dish (k) is analogous to an observation possessing a feautre. This is indicated by setting the value of z_{ik} to 1 if the customer takes the dish, and 0 otherwise.

An IBP(α) for N observations can be simulate as follows:

1. The 1^{st} customer takes Poisson(α) number of dishes
2. For customers $i = 2$ to N,
    - For each presiously sampled dish, customer i takes dish k with probability m_k/i
    - after sampling the last sampled dish, customer i samples Poisson(α/i) new dishes

Indian Buffet Process (IBP)

$$\underset{N\times\infty}{\mathbf{Z}} \sim \text{IBP}(\alpha), \text{ where } z_{ik} \in \{0,1\}$$

$$\implies \text{P}(\mathbf{Z}) = \frac{\alpha^{K_+}}{\prod_{i=1}^{N} {K_1}^{(i)}!} exp\{-\alpha H_N\} \prod_{k=1}^{K_+} \frac{(N-m_k)!(m_k-1)!}{N!},$$

where H_N is the harmonic number, $\sum_{i=1}^{N} i^{-1}$, K_+ is the number of non-zero columns in $\mathbf{Z}$, m_k is the k^{th} column sum of $\mathbf{Z}$, and $K_1^{(i)}$ is the "number of new dishes" sampled by customer i.

Indian Buffet Process (IBP)

To draw from $\mathbf{Z} \sim \text{IBP}(\alpha)$ using a **Gibbs** sampler,

1. Start with an arbitrary binary matrix of N rows
2. For each row, i,
   1. For each column, k,
   2. if $m_{-i,k} = 0$, delete column k. Otherwise,
      - set z_{ik} to 0
      - set z_{ik} to 1 with probability $P(z_{ik} = 1|\mathbf{z_{-i,k}}) = m_{-i,k}/i$
   3. at the end of row i, add Poisson(α/N) columns of 1's
3. iterate step 2 a large number of times

We can use this sampling algorithm to sample from the posterior distribution P$(\mathbf{Z}|\mathbf{X})$ where $\mathbf{Z} \sim IBP(\alpha)$ by sampling from the complete conditional

$$P(z_{ik} = 1|\mathbf{Z_{-(ik)}}, \mathbf{X}) \propto p(\mathbf{X}|\mathbf{Z})P(z_{ik} = 1|\mathbf{Z_{-(ik)}}). \qquad (1)$$

Indian Buffet Process (IBP)

Note that the conjugate prior for α is a Gamma distribution.

$$\begin{aligned}
\mathbf{Z}|\alpha &\sim IBP(\alpha) \\
\alpha &\sim Gamma(a, b), \text{where } b \text{ is the scale parameter}
\end{aligned}$$

$$\begin{aligned}
p(\alpha|\mathbf{Z}) &\propto p(\mathbf{Z}|\alpha)p(\alpha) \\
p(\alpha|\mathbf{Z}) &\propto \alpha^{K_+}e^{-\alpha H_N}\alpha^{a-1}e^{-\alpha/b} \\
p(\alpha|\mathbf{Z}) &\propto \alpha^{a+K_+-1}e^{-\alpha(1/b+H_N)}
\end{aligned}$$

$$\alpha|\mathbf{Z} \sim Gamma(a + K_+, (1/b + H_N)^{-1}) \tag{2}$$

Example: Linear-Gaussian Latent Feature Model with Binary Features

Suppose, we observe an $N \times D$ matrix $\mathbf{X}$, and we believe

$$\begin{aligned} \mathbf{X} &= \mathbf{ZA} + \mathbf{E}, \\ \text{where} \quad \mathbf{Z}|\alpha &\sim \text{IBP}(\alpha), \\ \mathbf{A}_i &\sim MVN(\mathbf{0}, {\sigma_A}^2\mathbf{I}), \\ \mathbf{E}_i &\sim MVN(\mathbf{0}, {\sigma_X}^2\mathbf{I}). \\ \alpha &\sim \text{Gamma}(a, b), \end{aligned}$$

Example: Linear-Gaussian Latent Feature Model with Binary Features

It can be shown that

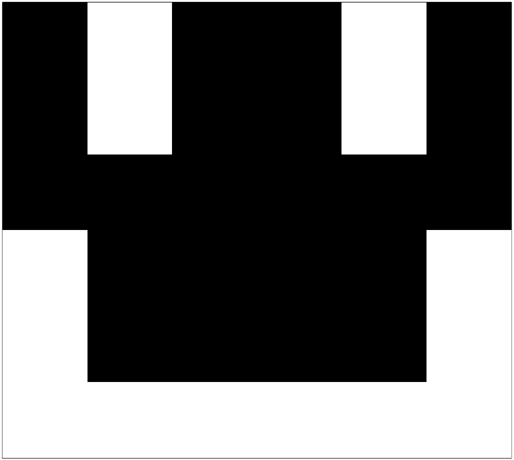
$$p(\mathbf{X}|\mathbf{Z}) = \frac{1}{(2\pi)^{ND/2}\sigma_X^{(N-K)D}\sigma_A^{KD}|\mathbf{Z}^T\mathbf{Z} + (\frac{\sigma_X}{\sigma_A})^2\mathbf{I}|^{D/2}} \exp\{-\frac{1}{2\sigma_X^2}tr(\mathbf{X}^T(\mathbf{I} - \mathbf{Z}(\mathbf{Z}^T\mathbf{Z} + (\frac{\sigma_X}{\sigma_A})^2\mathbf{I})^{-1}\mathbf{Z}^T)\mathbf{X}\} \quad (3)$$

Example: Linear-Gaussian Latent Feature Model with Binary Features

We can now use equation (1) to implement a Gibbs sampler to draw from the posterior posterior $\mathbf{Z}|\mathbf{X}, \alpha$.

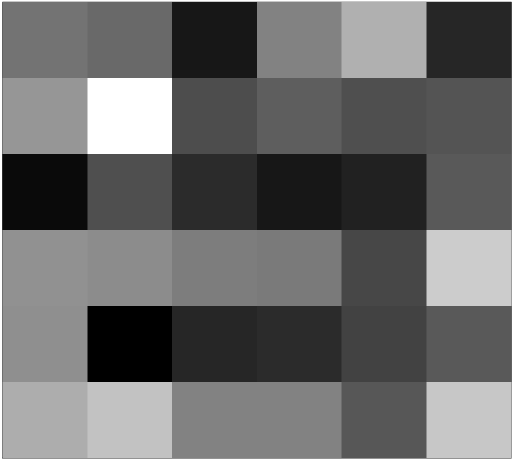
Data

Figure: One Observation Without Noise



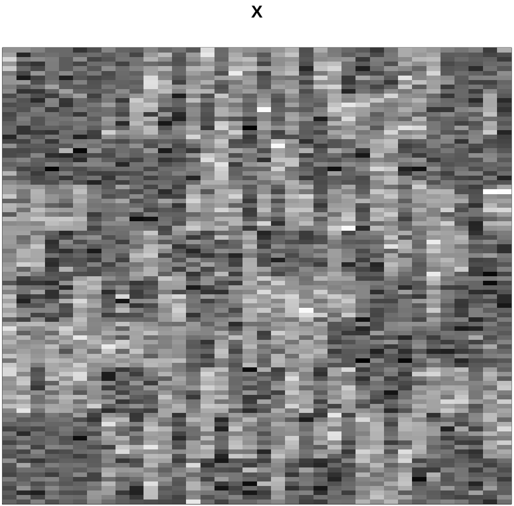
Data

Figure: One Observation + N(0,.5) Noise

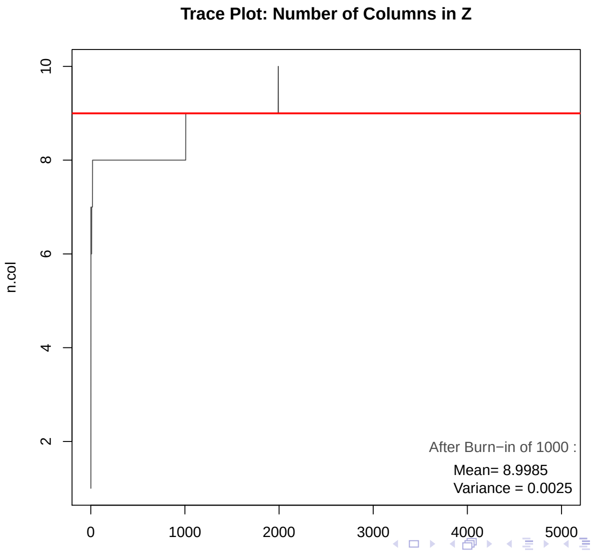


Data

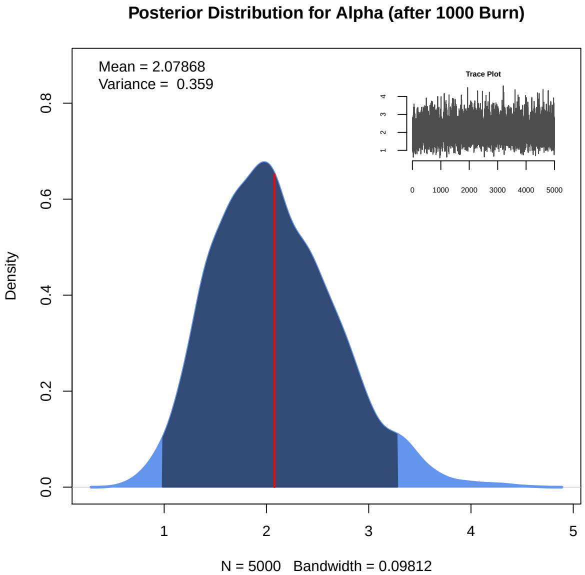
Figure: Data From Ten Stat666 Students + N(0,.5) Noise



Diagnostics



Diagnostics



Posterior

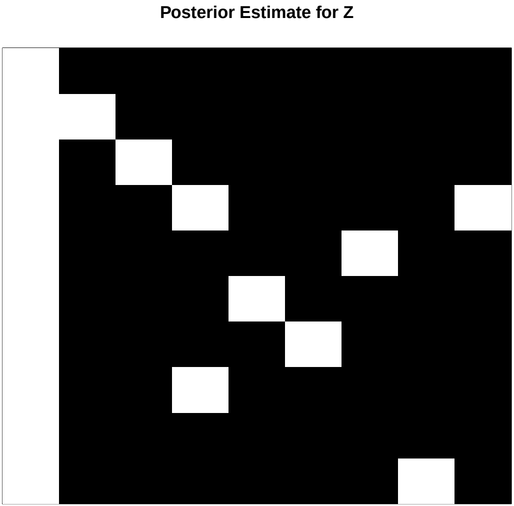


Figure: The posterior mean for **Z** computed by summing across all the **Z** matrices drawn, then dividing by 4000, the the number of draws.

Posterior

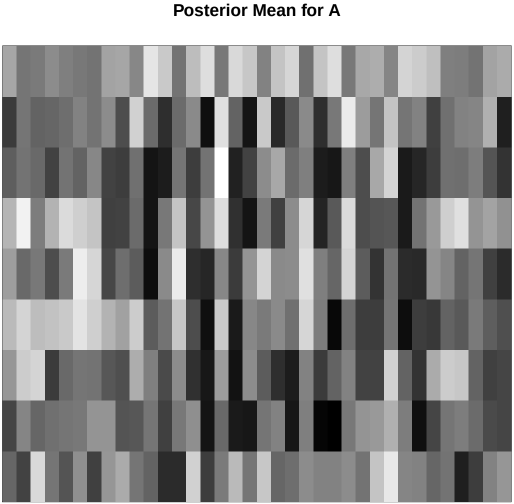


Figure: The posterior mean for **A** was computed by calculating $E[\mathbf{A}|\mathbf{Z}] = (\mathbf{Z}^T\mathbf{Z} + \frac{\sigma_X^2}{\sigma_A^2}\mathbf{I})^{-1}\mathbf{Z}^T\mathbf{X}$

Posterior

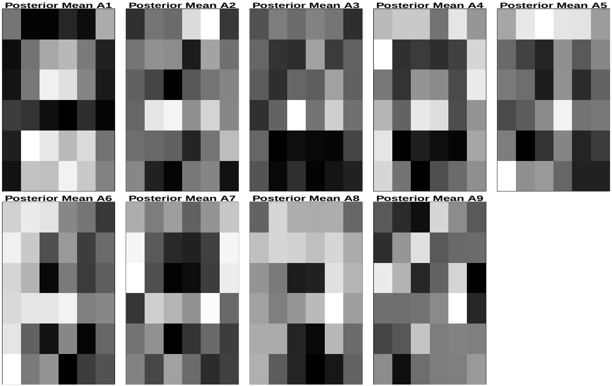


Figure: The latent features turned back into 6×6 images.

Posterior

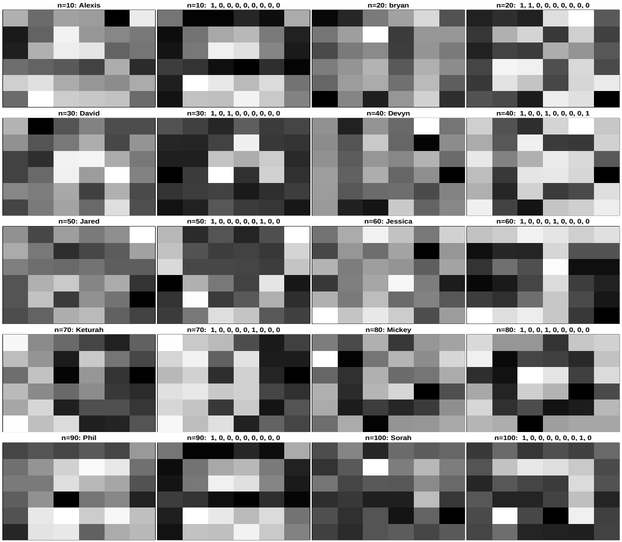
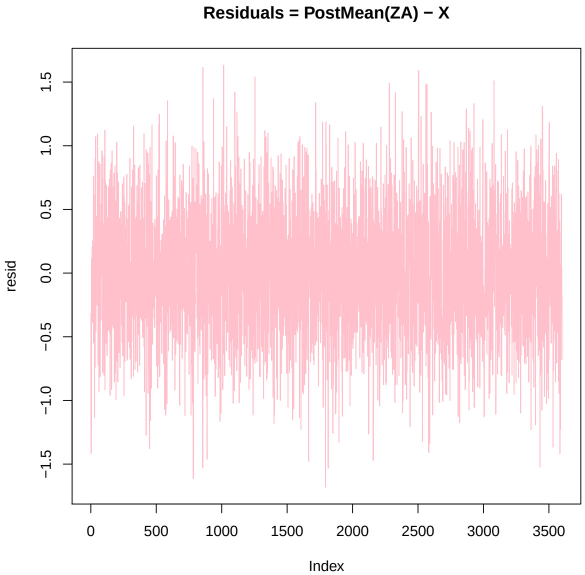


Figure: The latent features for each student = PostMean($\mathbf{Z}_i\mathbf{A}$).

Residuals


Residuals = PostMean(ZA) − X
resid
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
0
500
1000
1500
2000
2500
3000
3500
Index

Conclusions

- The IBP is flexible distribution on sparse binary matrices
- Number of latent features can be quickly learned, and latent features discovered in a Gaussian latent feature model
- Prior distributions can be set on σ_X and σ_A when the variance is unknown
- Distance information can be incorporated to enhance performance of the algorithm