10-801: Advanced Topics in Graphcal Models 10-801, Spring 2007

Bayesian Admixtures

Lecturer: Eric P. Xing *Scribe: Wenjie Fu*

Three Models

Admixture Model is also known as *Mixed Membership Model* and *Latent Dirichlet Allocation(LDA)*. Here are three applications in different fields.

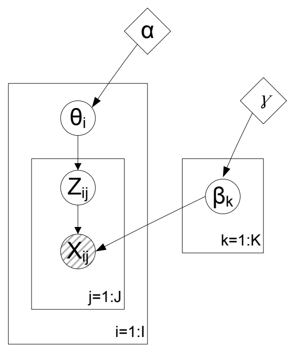

Figure 1: Example in Information Retrieval

In Figure 1, θ is Admixture Vector, aka Topic Vector. Each θ_i correspond to one paper, and they follow a Dirichlet Process parameterized by α. i, j and k denotes respectively the index of papers, words and topics. β is a $K \times M$ matrix.

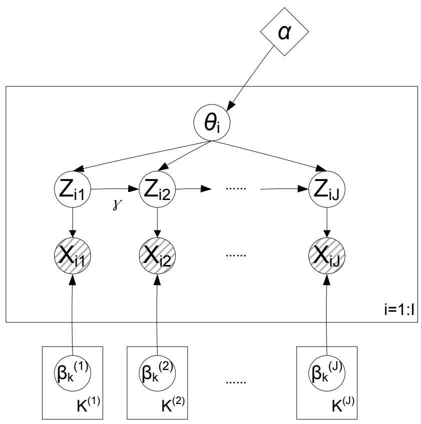

Figure 2: Example in Genomics

In Figure 2, θ is Admixture Vector, aka Population Structure. z_{ij} denotes marker. γ is the switching probability – with prob γ the topic will be regenerated according to θ_i and with prob $1-\gamma$ it will be the same as previous one. i, j and k denotes respectively the index of individuals, loci and genotypes.

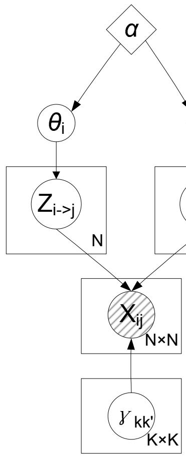

Figure 3: Example in Networks

In social networks, nodes are grouped in clusters, so that are just like topics. Also, one node can have several topics, which make this kind of problems similar to the previous ones. It is also called Admixture of Stochastic Blocks.
In Figure 3, θ is Admixture Vector, aka Aspect Vector. $z_{i\to j}$ denotes the edge from i to j. $\gamma_{kk'}$ indicates pair of functionality.

Inference

Take the first model above as an example.

$$P(x, z, \vec{\theta}, \vec{\beta}) = \prod_{i=1}^{N}\prod_{j=1}^{J_i} P(x_{ij}|z_{ij}, \vec{\beta})P(z_{ij}|\vec{\theta_i})P(\vec{\theta_i}|\vec{\alpha}) \prod_{k=1}^{K} P(\vec{\beta_k}|\vec{\gamma})$$
$$= \int \prod_{i}\prod_{k} \theta_{ik}^{z_{ik}+\alpha-1} \prod_{m=1}^{M} \beta_{km}^{\delta(z_i,k)n(m|x)+\gamma_k-1} d\vec{\theta}d\vec{\beta}$$

Gibbs Sampling

To solve the inference problem, one approach is the Gibbs sampling method. The unknown variables here are $z, \vec{\theta}, \vec{\beta}$. $z|\vec{\theta}, x, \vec{\beta}$ and $\vec{\theta}|z$ are easy to sample. $\vec{\beta}|x, z$ can be written as follows.

$$P(\vec{\beta}|x, z) = \frac{P(x|\vec{\beta}, z)P(\vec{\beta}|\vec{\gamma})}{\int P(x|\vec{\beta}, z)P(\vec{\beta}|\vec{\gamma})d\vec{\beta}}$$

Mean Field

Another apporach is the Mean Field method. There are two steps:

1. Break the probability function into factors
2. Replace hidden variables with approximation

In the first step, take $P(\beta, \theta, z|x)$ as an example.

$$P(\beta, \theta, z|x) = q(\beta)q(\theta)q(z)$$

To minimize the Kullback-Leibler divengence of q and p, q must has the form of $p(v|$ markov blanket of $v)$ For example,

$$\begin{aligned} & q(\theta \mid \alpha, z) \\ =& p(z \mid \theta) p(\theta \mid \alpha) \\ =& \theta_{ik}^{<z_{ik}>+\alpha-1} \end{aligned}$$

$$< z >= \mathbb{E}_{q(z|\theta)}[z]$$

The actual effect is that the prior of θ shifts to $\alpha+ < z >$.

$$\begin{aligned} & q(z \mid \theta, \beta, x) \\ \propto & \ p(z \mid \theta) \ p(x \mid z, \beta) \\ = & \ \theta_{ik}^{z_{ijk}} \beta_{km}^{z_{ijk}\delta(x_i,m)} \\ = & \ \exp\{z_{ijk} \ln \theta_{ik} + z_{ijk}\delta(x_i, m) \ln \beta_{km}\} \\ = & \ \exp\{z_{ijk}\gamma\} \\ \gamma = & < \ln \theta_{ik} > +\delta(x_i, m) < \ln \beta_{km} > \end{aligned}$$

Parameter Estimation

$$\begin{aligned} & p(n \mid \theta)p(\theta \mid \alpha) \\ =& \frac{\Gamma(|\alpha|)}{\prod_{i=1}^{K} \Gamma(\alpha_i)} \theta_i^{n_i+\alpha_i-1} \end{aligned}$$

α is norm 1 in the above equation.

$$L(\alpha \mid x) = \frac{\Gamma(|\alpha|)}{\Gamma(|\alpha| + |n|)} \prod_i \frac{\Gamma(\alpha_i + n_i)}{\Gamma(\alpha_i)}$$

$$\alpha = \arg\max L(\alpha)$$

Since $\alpha > 0$, let $\alpha_i = e^{\omega_i}$.

$$\omega_i = \arg\max L(\alpha)$$

We have to use gradient method.

$$
\begin{aligned}
\omega =& \omega + k\Delta\omega \\
\Delta\omega =& \frac{\partial l}{\partial \alpha} \cdot \frac{\partial \alpha}{\partial \omega} \qquad \Big(\frac{\partial \alpha}{\partial \omega} = \alpha\Big) \\
=& \alpha \frac{\partial}{\partial \alpha}\Big(\log\Gamma(|\alpha|) - \log\Gamma(|\alpha| + |n|) + \sum_i \log\Gamma(\alpha_i + n_i) - \sum_i \log\Gamma(\alpha_i)\Big)
\end{aligned}
$$

Replace all the $\log\Gamma$ in the equation with estimated values.
$|\alpha|$ describes how much we believe in the prior.

Structure

$$
\begin{aligned}
& \theta \sim \mathrm{Dir} \\
& P(n \mid \theta(\mu))\ P(\mu \mid \mu_i, \Sigma) \\
=& \exp\Big\{ n \ln\theta_i(\mu) - \frac{1}{2}(\mu - \mu_i)^{\mathrm{T}}\Sigma^{-1}(\mu - \mu_i)\Big\} \\
=& \exp\Big\{ n\mu - N\ln(1 + \sum \exp\mu_i) - \frac{1}{2}(\mu - \mu_i)^{\mathrm{T}}\Sigma^{-1}(\mu - \mu_i)\Big\} \\
& \mu_i = \ln\frac{\theta_i}{\theta_k} \qquad (\mu_k = 0) \\
& \theta_i = \frac{e^{\mu_i}}{1 + \sum_i e^{\mu_i}}
\end{aligned}
$$

This cannot be solved using exact inference.
Three trick may apply.

- Sampling (but we haven't seen any one use it, maybe infeasible)
- Method by David Blei
- Laplace