Weekly Report II

Donglai Wei

2010.4.21

1. Generalized Formula

1.0 ME Algorithm settings

0) Notation:
Observations:
$\vec{x} = (x^{(1)}, ..., x^{(N)})$
Hidden Variables:
$\vec{z} = (z^{(1)}, ..., z^{(N)})$: assignments from stochastic process;
θ: parameter from exponential distribution;
Hyperparameters:
λ:(λ_0, λ_a for θ and α, γ for $\vec{z}$)

The goal is to find the log probability: $log(p(\vec{x}|\lambda)) = log(\sum_{\vec{z}}[\int_\theta p(\vec{x}, \vec{z}, \theta|\lambda)\, d\theta])$

1) θ:
For Exponential Family,we can integrate out θ in close form:

likelihood of the data: $p(x|\theta) = v(x)exp\{\sum_{a\in\mathcal{A}} \theta_a\phi_a(x) - \Phi(\theta)\}$
Prior for the parameter: $p(\theta|\lambda) = exp\{\sum_{a\in\mathcal{A}} \theta_a\lambda_0\lambda_a - \lambda_0\Phi_(\theta) - \Omega(\lambda)\}$
Posterior for the parameter : $p(\theta|x^{(1)}, ..., x^{(N)}, \lambda) = p(\theta|\bar{\lambda})$
where $\bar{\lambda}_0 = \lambda_0 + N, \bar{\lambda}_a = \frac{\lambda_0\lambda_a + \sum_{l=1}^N \phi_a(x^{(l)})}{\lambda_0+N}\lambda_0 + N$
Thus, θ can be easily integrate out: $log(p(x^{(1)}, ..., x^{(N)}|\lambda)) = \Omega(\bar{\lambda}) - \Omega(\lambda) + \sum_{l=1}^N v(x^{(l)})$

2) $\vec{z}$:
But the discrete variable $\vec{z}$ is hard to sum out.
So instead, we pick the MAP estimator $\vec{z}^*$:

$$
\begin{aligned}
log(p(\vec{x}|\lambda)) &= log(\sum_{\vec{z}}[\int_{\theta} p(\vec{x},\vec{z},\theta|\lambda)\,d\theta]) \\
&\approx log(\int_{\theta} p(\vec{x},\vec{z}^{*},\theta|\lambda)\,d\theta) \\
&= log(\int_{\theta} p(\vec{x},\theta|\vec{z}^{*},\lambda)\,d\theta) + log(p(\vec{z}|\lambda)) \\
&= \Omega(\bar{\lambda}) - \Omega(\lambda) + \sum_{l=1}^{N} v(x^{(l)}) + log(p(\vec{z}|\lambda))
\end{aligned}
$$

1.1 θ: Conjugated Exponential Family

K: number of clusters in $\vec{z}$
N: number of observations
N_c: the number of points in cluster c

a)Nomal-Inverse-Wishart

$$
\begin{aligned}
p(\theta|\lambda) &= p(\mu,\Sigma|m_0,\eta_0,\xi_0,B_0) \\
&= NIW(m_0,\eta_0,\xi_0,B_0) \\
&= \frac{1}{Z}|\Sigma|^{-((\eta_0+d)/2+1)} exp(-\frac{1}{2}tr(B_0\Sigma^{-1}) - \frac{\xi_0}{2}(\mu-m_0)^T\Sigma^{-1}(\mu-m_0))
\end{aligned}
$$

$Z = \dfrac{2^{\eta_0 d/2}\Gamma_D(\eta_0/2)(2\pi/\xi)^{d/2}}{|\Sigma|^{\eta_0/2}}$,
$\Gamma_D(x) = \pi^{\frac{D(D-1)}{4}}\Pi_{i=1}^{D}\Gamma(x+\frac{1-i}{2})$

$$
\begin{aligned}
log(p(\vec{x},\theta|\vec{z},\lambda)) &= log(p(\theta|\lambda)) + \sum_{n=1}^{N}[log(p(x_n|z_n,\theta))] \\
&= log(\mathcal{N}(\mu|m_0,\xi_0\Omega)\mathcal{W}(\Omega|\eta_0,B_0)) + \sum_{n=1}^{N}[log(\mathcal{N}(x_n|z_n,\mu,\Omega))] \\
&= -\frac{DN}{2}log(2\pi) + log(\mathcal{N}(\mu|m_c,\xi_c\Omega_c)\mathcal{W}(\Omega_c|\eta_c,B_c)) + log(p(\vec{x}|\vec{z},\lambda))
\end{aligned}
$$

S_c: the covariance of datas in cluster c
$\phi_c = \phi_0 + N_c$
$m_c = \frac{N_c\bar{x}_c + \xi_0 m_0}{\xi_c}$
$B_c = B_0 + N_c S_c + \frac{N_c\xi_0}{\xi_c}(\bar{x}_c - m_0)(\bar{x}_c - m_0)^T$

$\eta_c = \eta_0 + N_c$
$\xi_c = \xi_0 + N_c$

$log(\int_\theta p(\vec{x}, \theta|\vec{z}, \lambda)\, d\theta)$

$$\begin{aligned}
&= -\frac{DN}{2}log(2\pi) + log(\int_\theta \mathcal{N}(\mu|m_c, \xi_c\Omega_c)\mathcal{W}(\Omega_c|\eta_c, B_c)\, d\theta) + log(p(\vec{x}|\vec{z}, \lambda)) \\
&= -\frac{DN}{2}log(2\pi) + 0 + \sum_{c=1}^{K} log(Z(\bar{\lambda}_c)) - log(Z(\lambda_c)) \\
&= -\frac{DN}{2}log(2\pi) + \sum_{c=1}^{K} log(\frac{2^{\eta_c D/2}\Gamma_D(\eta_c/2)(2\pi/\xi_c)^{D/2}|\Sigma|^{\eta_0/2}}{2^{\eta_0 D/2}\Gamma_D(\eta_0/2)(2\pi/\xi_0)^{D/2}|\Sigma|^{\eta_c/2}}) \\
&= -\frac{DN}{2}log(2\pi) + \sum_{c=1}^{K}[\frac{DN}{2}log(2) - (\frac{D}{2}log\frac{\xi_c}{\xi_0}) + (\frac{\eta_0}{2}logdet(B_0) - \frac{\eta_c}{2}logdet(B_c)) + (log\frac{\Gamma_D(\frac{\eta_c}{2})}{\Gamma_D(\frac{\eta_0}{2})})]
\end{aligned}$$

b)Dirichlet-Multinomial

$$\begin{aligned}
p(\theta|\lambda) &= \mathcal{D}(\beta|\phi_0) \\
&= \frac{1}{Z}\Pi_{i=1}^{i=K}\beta_i^{(\phi_0 - 1)}
\end{aligned}$$

$$Z = \frac{\Gamma(\phi_0)^K}{\Gamma(K\phi_0)}$$

$$\begin{aligned}
p(\vec{x}, \theta|\vec{z}, \lambda) &= \Pi_{n=1}^{N}[p(x_n|z_n, \theta)]p(\theta|\lambda) \\
&= \Pi_{n=1}^{N}[\mathcal{M}(x_n|z_n, \alpha)] \times \mathcal{D}(\alpha|\phi_0) \\
&= \mathcal{D}(\alpha|\phi_c) \times p(\vec{x}|\vec{z}, \lambda)
\end{aligned}$$

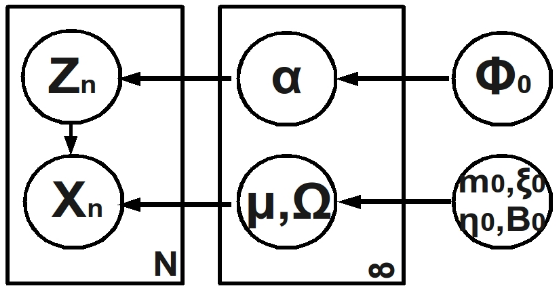
$$\phi_c = \phi_0 + N_c$$

$$\begin{aligned}
log(\int_\theta p(\vec{x}, \theta|\vec{z}, \lambda)\, d\theta) &= log(\int_\theta \mathcal{D}(\alpha|\phi_c)\, d\theta) + log(p(\vec{x}|\vec{z}, \lambda)) \\
&= 0 + log(Z(\phi_c)) - log(Z(\phi_0)) \\
&= log(\frac{\Pi_{c=1}^{K}\Gamma(\phi_c)}{\Gamma(N + K\phi_0)} / \frac{\Gamma(\phi_0)^K}{\Gamma(K\phi_0)}) \\
&= log(\frac{\Gamma(K\phi_0)}{\Gamma(N + K\phi_0)}) + \sum_{c=1}^{K} log(\frac{\Gamma(\phi_c)}{\Gamma(\phi_0)})
\end{aligned}$$

1.2 $\vec{z}$: Stochastic Process

a)DP mixture(with Chinese Restaurant Process)

Graphical Model for DP mixture



K: number of clusters in $\vec{z}$
N: number of observations
N_c: the number of points in cluster c
$p(\vec{z}|\lambda) = p(z_1, ..., z_N|\phi_0) = \Pi_{j=1}^{j=N} p(z_j|z_1, ..., z_{j-1}, \phi_0)$

$p(z_j|z_1, ..., z_{j-1}, \phi_0) = \sum_{k=1}^{K(j)} \frac{m_{k(j)}}{j-1+\phi_0}\delta_{z_j=k} + \frac{\phi_0}{j-1+\phi_0}\delta_{z_j=K(j)+1}$

1) Partition:$\Pi_{j=1}^{j=N} \frac{1}{\phi_0+j-1} = \frac{\Gamma(\phi_0)}{\Gamma(N+\phi_0)}$
2) Forming new clusters:ϕ_0^K
3) Accumulating for all clusters:$\Pi_{j=1}^{j=N}(N_j-1)! = \Pi_{j=1}^{j=N}\Gamma(N_j)$

So,$p(\vec{z}|\lambda) = \frac{\Gamma(\phi_0)}{\Gamma(N+\phi_0)}\Pi_{c=1}^{K}[\Gamma(N_c)]\phi_0^K$

b)HDP(with Chinese Franchise Process)

Notation:

1) Global: J Restaurants, K global dishes, T tables, N datas,
t_{ji} : the table that customer i in Restaurant j sits;
k_{jt} : the dish that table t in Restaurant j serves;
$\vec{k}$: new dish different from what has been served;
$\vec{t}$: the table different from what has been occupied;
2) Counting so far (during the process)
m_{jk} : number of tables in Restaurant j serving dish k;
$m_{.k}$: number of tables serving dish k;
$m_{j.}$: number of tables in Restaurant j;
$m_{..}$: number of tables;
n_{jtk} : number of customers in Restaurant j at table t eating dish k;
$n_{jt.}$: number of customers in Restaurant j at table t;

$n_{j..}$: number of customers in Restaurant j;

$p(\vec{z}, \vec{k}|\lambda) = p(k_{11}, ..., k_{JT_J}|\gamma)p(t_{11}, ..., t_{JN_J}|\alpha)$

$p(k_{11}, ..., k_{JT_J}|\gamma) = \Pi_{j=1}^{J}(\Pi_{i=1}^{T_j} p(k_{ji}|\vec{k}_1, ...\vec{k}_{j-1}, k_{j1}, ..., k_{j(i-1)}, \gamma))$

$p(t_{11}, ..., t_{JN_J}|\alpha) = \Pi_{j=1}^{J}(\Pi_{i=1}^{N_j} p(t_{ji}|t_{j1}, ..., t_{j(i-1)}, \alpha))$

$p(k_{jt}|\vec{k}_1, ...\vec{k}_{j-1}, k_{j1}, ..., k_{j(t-1)}, \gamma) = \sum_{k=1}^{K} \frac{m_{.k}}{m_{..}-1+\gamma}\delta_{k_{jt}=k} + \frac{\gamma}{m_{..}-1+\gamma}\delta_{k_{ji}=\vec{k}}$

$p(t_{ji}|t_{j1}, ..., t_{j(i-1)}, \alpha) = \sum_{t=1}^{m_{j.}} \frac{n_{jt.}}{i-1+\alpha}\delta_{t_{ji}=t} + \frac{\alpha}{i-1+\alpha}\delta_{t_{ji}=\vec{t}}$

1) Partition:

k: $\Pi_{w=1}^{\sum_{j=1}^{J} m_{j.}} \frac{1}{\gamma+w-1} = \Pi_{w=1}^{m_{..}} \frac{1}{\gamma+w-1} = \frac{\Gamma(\gamma)}{\Gamma(T+\gamma)}$

t: $\Pi_{j=1}^{J}\Pi_{i=1}^{i=n_{j..}} \frac{1}{\alpha+i-1} = \Pi_{j=1}^{j=J} \frac{\Gamma(\alpha)}{\Gamma(n_{j..}+\alpha)}$

2) Forming new clusters (1st point in the cluster):

k: γ^K

t: $\Pi_{j=1}^{J}\alpha^{m_{j.}}$

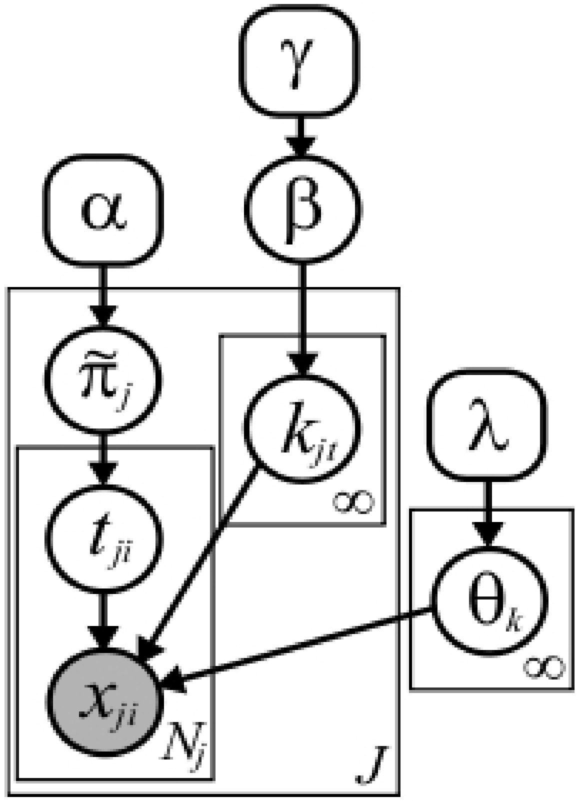
3) Accumulating for each clusters (other points in the cluster):

k: $\Pi_{k=1}^{k=K}(m_{.k}-1)! = \Pi_{k=1}^{k=K}\Gamma(m_{.k})$

t: $\Pi_{j=1}^{j=J}\Pi_{t=1}^{t=m_{j.}}(n_{jt.}-1)! = \Pi_{j=1}^{j=J}\Pi_{t=1}^{t=m_{j.}}\Gamma(n_{jt.})$

So,$log(p(\vec{z}, \vec{k}|\lambda)) = log\{\Pi_{j=1}^{J}[\frac{\Gamma(\alpha)}{\Gamma(n_{j..}+\alpha)}\Pi_{t=1}^{m_{j.}}(\Gamma(n_{jt.}))]\alpha^{\sum_{j=1}^{J} m_{j.}}\} + log\{\frac{\Gamma(\gamma)}{\Gamma(T+\gamma)}\Pi_{k=1}^{K}[\Gamma(m_{.k})]\gamma^K\}$

Graphical Model for HDP


γ
α
β
$\tilde{\pi}_j$
k_{jt}
∞
λ
t_{ji}
θ_k
x_{ji}
∞
N_j
J