HW14

PB20111689 蓝俊玮

T1

1.
$$
\begin{aligned}
p(D,\mu,\lambda) &= p(D|\mu,\lambda)p(\mu,\lambda) \\
&= \prod_{i=1}^{m} p(x_i|\mu,\lambda)p(\mu,\lambda) \\
&= \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi\lambda^{-1}}}\exp\left(-\frac{(x_i-\mu)^2}{2\lambda^{-1}}\right)\cdot\frac{1}{\sqrt{2\pi(k_0\lambda)^{-1}}}\exp\left(-\frac{(\mu-\mu_0)^2}{2(k_0\lambda)^{-1}}\right)\frac{1}{\Gamma(a_0)}b_0^{a_0}\lambda^{a_0-1}\exp\left(-b_0\lambda\right) \\
&= (\frac{1}{\sqrt{2\pi\lambda^{-1}}})^m\exp\left(-\sum_{i=1}^{m}\frac{(x_i-\mu)^2}{2\lambda^{-1}}\right)\cdot\frac{1}{\sqrt{2\pi(k_0\lambda)^{-1}}}s\exp\left(-\frac{(\mu-\mu_0)^2}{2(k_0\lambda)^{-1}}\right)\frac{1}{\Gamma(a_0)}b_0^{a_0}\lambda^{a_0-1}\exp\left(-b_0\lambda\right)
\end{aligned}
$$

2. 证据下界（即变分推断的优化目标）为：

   $$E_q[\log p(x,z)]-E_q[\log q(z)]=E_q[p(x|\mu,\lambda)]+E_q[\log p(\mu|\lambda)]+E_q[\log p(\lambda)]-E_q[\log q(\mu)]-E_q[\log q(\mu)]$$

   变分目标的目标是找到

   $$q^*(z)=\arg\min_{q(z)}\mathrm{KL}(q(z)||p(z|x))$$

   即需要找到一个 $q^*(z)\approx p(z|x)$ 来近似得到 $p(z|x)$，注意到 KL 散度还可以写成：

   $$\mathrm{KL}(q(z)||p(z|x))=E_q[\log q(z)]-E_q[\log p(x,z)]+\log p(x)\geq 0$$

   所以可以得到：

   $$\sum_{i=1}^{m}\log p(x_i)=\log p(x)\geq E_q[\log p(x,z)]-E_q[\log q(z)]$$

   可以发现，不等式的右端即为证据下界，因此证明了数据边际的似然 $\sum\limits_{i=1}^{m}\log p(x_i)$ 的下界为证据下界。

3. 现在需要近似推断得到后验概率 $p(\mu,\lambda|D)$，即 $p(z|x)$，那么如上面所述，需要找到一个 $q^*(z)\approx p(z|x)$ 来近似得到 $p(z|x)$。因此我们需要求解出 KL 散度的最优值。得到：

   $$\frac{\partial L}{\partial q_\lambda(\mu)}=E_\lambda[\log p(\mu|\lambda)]+E_\lambda[\log p(D|\mu,\lambda)]-\log q(\mu)=0$$

   可以得到：

   $$
   \begin{aligned}
   \log q^*(\mu) = & -\frac{E[\lambda]\kappa_0}{2}(\mu-\mu_0)^2-\frac{E[\lambda]}{2}\sum_{i=1}^{m}(x_i-\mu)^2 \\
   = & -\frac{E[\lambda]}{2}\left[(\kappa_0+m)\mu^2+\sum_{i=1}^{m}x_i^2-2\mu(\kappa_0\mu_0+m\bar{x})\right] \\
   = & -\frac{E[\lambda]}{2}\left[(\kappa_0+m)(\mu-\frac{\kappa_0\mu_o+m\bar{x}}{\kappa_0+m})^2+\sum_{i=1}^{m}x_i^2-\frac{(\kappa_0\mu_0+m\bar{x})^2}{\kappa_0+m}\right]
   \end{aligned}
   $$

   后面两项是不影响分布的，因此有：

   $$
   \begin{gathered}
   q^*(\mu)\sim\mathcal{N}(\mu|\mu_m,\lambda_m^{-1}) \\
   \mu_m=\frac{\kappa_0\mu_0+m\bar{x}}{\kappa_0+m},\ \lambda_m=(\kappa_0+m)E[\lambda]
   \end{gathered}
   $$

   同理可得：

   $$\frac{\partial L}{\partial q_\mu(\lambda)}=E_\mu[\log p(D|\mu,\lambda)]+E[\log(\mu|\lambda)]+E_\mu[\log(\mu|\lambda)]-\log q(\lambda)=0$$

   可以得到：

$$
\begin{aligned}
\log q^*(\lambda) =& -\frac{\lambda}{2}E[\mu]\left[\kappa_0(\mu-\mu_0)^2+\sum_{i=1}^{m}(x_i-\mu)^2\right]+(a_0-1)\log\lambda-b_0\lambda+\frac{m+1}{2}\log\lambda \\
=& \log\lambda(a_0+\frac{m-1}{2})-\lambda\left(b_0+\frac{1}{2}E[\mu]\left[\kappa_0(\mu-\mu_0)^2+\sum_{i=1}^{m}(x_i-\mu)^2\right]\right)
\end{aligned}
$$

因此有:

$$
q^*(\lambda)\sim \mathrm{Gam}(\lambda|a_m,b_m)
$$
$$
a_m=a_0+\frac{m-1}{2},\ b_m=b_0+\frac{1}{2}E_\mu\left[\kappa_0(\mu-\mu_0)^2+\sum_{i=1}^{m}(x_i-\mu)^2\right]
$$

无先验条件下:

$$
\mu_0=a_0=b_0=\kappa_0=0
$$
$$
E[\lambda]=\frac{a_m}{b_m},\ E[\mu^2]=\bar{x}^2+\frac{1}{mE[\lambda]},\ E[\mu]=\mu_m=\bar{x}
$$

解得:

$$
E[\lambda]=\frac{1}{\bar{x^2}-\bar{x}^2}=\frac{1}{\mathrm{Var}(x)}
$$

将上述条件代入回，则可以得到:

$$
p(\mu,\lambda|D)\sim\mathcal{N}(\mu|\mu_m,\lambda_m^{-1})\,\mathrm{Gam}(\lambda|a_m,b_m)
$$

T2

使用维比特算法求解条件随机场的预测问题:

1. 初始化

$$
\delta_1(j)=w\cdot F_1(y_0=\text{start},y_1=j,x),\quad j=1,2,\ldots,m
$$

2. 递推。对 $i=2,3,\ldots,n$

$$
\delta_i(l)=\max_{1\le j\le m}\{\delta_{i-1}(j)+w\cdot F_i(y_{i-1}=j,y_i=l,x)\},\quad l=1,2,\ldots,m
$$
$$
\Psi_i(l)=\arg\max_{1\le j\le m}\{\delta_{i-1}(j)+w\cdot F_i(y_{i-1}=j,y_i=l,x)\},\quad l=1,2,\ldots,m
$$

3. 终止

$$
\max_y(w\cdot F(y,x))=\max_{1\le j\le m}\delta_n(j)
$$
$$
y_n^*=\arg\max_{1\le j\le m}\delta_n(j)
$$

4. 返回路径

$$
y_i^*=\Psi_{i+1}(y_{i+1}^*),\quad i=n-1,n-2,\ldots,1
$$

求得最优路径 $y^*=(y_1^*,y_2^*,\ldots,y_n^*)$