# 从无向图说起

无向图模型通常由一个未归一化的概率分布

$$\phi(x)$$

来定义。即

$$p(x;\theta) = \frac{1}{Z(\theta)}\phi(x;\theta)$$

# 对数似然的梯度

$$\nabla_\theta \log p(x;\theta) = \nabla_\theta \log \phi(x;\theta) - \nabla_\theta \log Z(\theta)$$

注意:

$$\begin{aligned}
\nabla_\theta \log Z(\theta) &= \frac{\nabla_\theta Z}{Z} \\
&= \frac{1}{Z}\nabla_\theta \sum_x \phi(x;\theta) \\
&= \frac{1}{Z}\sum_x \nabla_\theta \phi(x;\theta)
\end{aligned}$$

若 $\phi(x) > 0$ ,则

$$\begin{aligned}
\nabla_\theta \log Z(\theta) &= \frac{1}{Z}\sum_x \nabla_\theta \phi(x;\theta) \\
&= \frac{1}{Z}\sum_x \nabla_\theta \exp \log \phi(x;\theta) \\
&= \frac{1}{Z}\sum_x \phi(x;\theta)\nabla_\theta \log \phi(x;\theta) \\
&= \sum_x p(x;\theta)\nabla_\theta \log \phi(x;\theta) \\
&= E_{\mathrm{x}\sim p(x;\theta)}[\nabla_\theta \log \phi(x;\theta)] \\
&= E_{\text{model}}[\nabla_\theta \log \phi(x;\theta)]
\end{aligned}$$

对 N 个样本梯度下降

$$\begin{aligned}
\nabla_\theta L(\theta) &= \frac{1}{N}\sum_{i=1}^{N} \nabla_\theta \log \phi(x_i;\theta) \\
&= \frac{1}{N}\sum_{i=1}^{N} \nabla_\theta \log \phi(x_i;\theta) - E_{\text{model}}[\nabla_\theta \log \phi(x;\theta)]
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{N}\sum_{i=1} \nabla_\theta \log \phi(x_i;\theta) - E_{\text{mod}}[\nabla_\theta \log \phi(x;\theta)] \\
&= E_{\text{emp}}[\nabla_\theta \log \phi(x;\theta)] - E_{\text{mod}}[\nabla_\theta \log \phi(x;\theta)]
\end{aligned}
$$

**如果有隐变量呢?**

设

$$
p(x,h;\theta) = \frac{1}{Z(\theta)}\phi(x,h;\theta)
$$

则log liklihood

$$
\begin{aligned}
L(\theta) &= \frac{1}{N}\sum_{i=1}^{N} \log p(x_i;\theta) \\
&= \frac{1}{N}\sum_{i=1}^{N} \log \sum_h p(x_i,h;\theta) \\
&= \frac{1}{N}\sum_{i=1}^{N} \log[\frac{1}{Z(\theta)}\sum_h \phi(x_i,h;\theta)]
\end{aligned}
$$

注意

$$
\sum_h \phi(x_i,h;\theta)
$$

是概率分布

$$
\begin{aligned}
q(x_i,h;\theta) &= \frac{1}{\sum_h \phi(x_i,h;\theta)}\phi(x_i,h;\theta) \\
&= \frac{p(x_i,h;\theta)}{p(x_i;\theta)} \\
&= p(h|x_i;\theta)
\end{aligned}
$$

中的配分函数。

所以

$$
\nabla_\theta \log \sum_h \phi(x_i,h;\theta) = E_{h\sim p(h|x_i;\theta)}[\nabla_\theta \log \phi(x_i,h;\theta)]
$$

所以，

$$
\begin{aligned}
\nabla_\theta L(\theta) &= \frac{1}{N}\sum_{i=1}^{N} \log[\frac{1}{Z(\theta)}\sum_h \phi(x_i,h;\theta)] \\
&= \frac{1}{N}\sum_{i} E_{h\sim p(h|x_i;\theta)}[\nabla_\theta \log \phi(x_i,h;\theta)]
\end{aligned}
$$

$$
\begin{aligned}
&\frac{1}{N}\sum_i E_{h\sim p(h|x_i;\theta)}[\nabla_\theta \log \phi(x_i, h;\theta)] \\
&- E_{(x,h)\sim p(x,h;\theta)}[\nabla_\theta \log \phi(x, h;\theta)]
\end{aligned}
$$

## 例子：maxent model

$$
p(x;\theta) = \frac{1}{Z}\exp(\sum_c \theta_c^T \phi_c(x))
$$

则

$$
\nabla_{\theta_c} L(\theta) = E_{\text{emp}}[\phi_c(x)] - E_{\text{mod}}[\phi_c(x)]
$$

含有隐变量时：

$$
\begin{aligned}
\nabla_{\theta_c} L(\theta) = &\frac{1}{N}\sum_i E_{h\sim p(h|x_i;\theta)}[\phi_c(x_i, h;\theta_c)] \\
&- E_{(x,h)\sim p(x,h;\theta)}[\phi_c(x, h;\theta_c)]
\end{aligned}
$$

## 更具体的例子：BernouliRBM

$$
p(v, h|\theta) = \frac{1}{Z(\theta)}\exp(v^T W h + v^T b + h^T c)
$$

则

$$
\begin{aligned}
&\nabla_W(v^T W h + v^T b + h^T c) = \nabla_W(v^T W h) = [\nabla_u(v^T u)]h^T = vh^T \\
&\nabla_b(v^T W h + v^T b + h^T c) = v \\
&\nabla_c(v^T W h + v^T b + h^T c) = h
\end{aligned}
$$

则

$$
\begin{aligned}
\nabla_W L &= \frac{1}{N}\sum_{i=1}^N E_{h\sim p(h|v_i;\theta)}[v_i h^T] - E_{\text{model}}[vh^T] \\
\nabla_b L &= \frac{1}{N}\sum_{i=1}^N E_{h\sim p(h|v_i;\theta)}[v_i] - E_{\text{model}}[v] \\
&= v_i - E_{\text{model}}[v] \\
\nabla_c L &= \frac{1}{N}\sum_{i=1}^N E_{h\sim p(h|v_i;\theta)}[h] - E_{\text{model}}[h]
\end{aligned}
$$

矩阵化：

$$
V = \begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_N^T \end{bmatrix} \quad H = \begin{bmatrix} h_1^T \\ h_2^T \\ \vdots \\ h_N^T \end{bmatrix}
$$

$$\sum_{i=1}^{N} v_i h_i^T = V^T H$$