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

$$abla_{ heta} \log p(x; heta) =
abla_{ heta} \log \phi(x; heta) -
abla_{ heta} \log Z(heta)$$

注意:

$$egin{aligned}
abla_{ heta} \log Z(heta) &= rac{
abla_{ heta} Z}{Z} \ &= rac{1}{Z}
abla_{ heta} \sum_{x} \phi(x; heta) \ &= rac{1}{Z} \sum_{x}
abla_{ heta} \phi(x; heta) \end{aligned}$$

若 $\phi(x) > 0$,则

$$egin{aligned}
abla_{ heta} \log Z(heta) &= rac{1}{Z} \sum_{x}
abla_{ heta} \phi(x; heta) \ &= rac{1}{Z} \sum_{x}
abla_{ heta} \exp \log \phi(x; heta) \ &= rac{1}{Z} \sum_{x} \phi(x; heta)
abla_{ heta} \log \phi(x; heta) \ &= \sum_{x} p(x; heta)
abla_{ heta} \log \phi(x; heta) \ &= E_{x \sim p(x; heta)} [
abla_{ heta} \log \phi(x; heta)] \ &= E_{ ext{model}} [
abla_{ heta} \log \phi(x; heta)] \end{aligned}$$

对 N 个样本梯度下降

$$egin{aligned}
abla_{ heta} L(heta) &= rac{1}{N} \sum_{i=1}^N
abla_{ heta} \log \phi(x_i; heta) \ &= rac{1}{N} \sum_{i=1}^N
abla_{ heta} \log \phi(x_i; heta) \end{aligned}$$

$$egin{aligned} &= \overline{N} \sum_{i=1}^{N} \mathbf{v}_{\, heta} \log \varphi(x_i; \sigma) - \mathbf{E}_{\mathrm{mod}\,[\, \mathbf{v}_{\, heta} \log \varphi(x; \sigma)]} \ &= E_{\mathrm{emp}}[
abla_{ heta} \log \phi(x; heta)] - E_{\mathrm{mod}}[
abla_{ heta} \log \phi(x; heta)] \end{aligned}$$

如果有隐变量呢?

设

$$p(x,h; heta) = rac{1}{Z(heta)}\phi(x,h; heta)$$

则log liklihood

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

注意

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

是概率分布

$$egin{aligned} q(x_i,h; heta) &= rac{1}{\sum_h \phi(x_i,h; heta)} \phi(x_i,h; heta) \ &= rac{p(x_i,h; heta)}{p(x_i; heta)} \ &= p(h|x_i; heta) \end{aligned}$$

中的配分函数。

所以

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

所以,

$$egin{aligned}
abla_{ heta} L(heta) &= rac{1}{N} \sum_{i=1}^N \log[rac{1}{Z(heta)} \sum_h \phi(x_i, h; heta)] \ &= rac{1}{N} \sum_{i=1}^N E_{h \sim p(h|x_i: heta)} [
abla_{ heta} \log \phi(x_i, h; heta)] \end{aligned}$$

$$N \stackrel{\textstyle extstyle }{\sim} i$$
 $-E_{(x,h)\sim p(x,h; heta)}[
abla_{ heta}\log\phi(x,h; heta)]$

例子: maxent model

$$p(x; heta) = rac{1}{Z} \exp(\sum_c heta_c^T \phi_c(x))$$

则

$$abla_{ heta_c} L(heta) = E_{ ext{emp}}[\phi_c(x)] - E_{ ext{mod}}[\phi_c(x)]$$

含有隐变量时:

$$egin{aligned}
abla_{ heta_c} L(heta) &= rac{1}{N} \sum_i E_{h \sim p(h|x_i; heta)} [\phi_c(x_i,h; heta_c)] \ &- E_{(x,h) \sim p(x,h; heta)} [\phi_c(x,h; heta_c)] \end{aligned}$$

更具体的例子: BernouliRBM

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

则

$$egin{aligned}
abla_W(v^TWh + v^Tb + h^Tc) &=
abla_W(v^TWh) &= [
abla_u(v^Tu)]h^T &= vh^T \
abla_b(v^TWh + v^Tb + h^Tc) &= v \
abla_c(v^TWh + v^Tb + h^Tc) &= h \end{aligned}$$

则

$$egin{aligned}
abla_W L &= rac{1}{N} \sum_{i=1}^N E_{h \sim p(h|v_i; heta)}[v_i h^T] - E_{ ext{model}}[v h^T] \
abla_b L &= rac{1}{N} \sum_{i=1}^N E_{h \sim p(h|v_i; heta)}[v_i] - E_{ ext{model}}[v] \ &= v_i - E_{ ext{model}}[v] \
abla_c L &= rac{1}{N} \sum_{i=1}^N E_{h \sim p(h|v_i; heta)}[h] - E_{ ext{model}}[h] \end{aligned}$$

矩阵化:

$$V = egin{bmatrix} v_1^T \ v_2^T \ dots \ v_N^T \end{bmatrix} \ H = egin{bmatrix} h_1^T \ h_2^T \ dots \ h_N^T \end{bmatrix}$$

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