机器学习 第九次作业

题目 1：设玻尔兹曼分布$P(x)=\frac{1}{Z}exp\big(-\beta H(x)\big)$的能量函数为$\mathrm{H(x)}=-\sum_{\mathrm{i<j<k}}\mathrm{J_{i,j,k}x_ix_jx_k}$。根据指数族(exponential family)定义，指出自然参数与充分统计学量分别是什么？

假设已知训练集合$D=\{x^{(1)},x^{(2)},\dots,x^{(L)}\}$，计算对数似然函数关于$J_{i,j,k}$的导数的表达式，指明化简后的表达式中每一项的含义。

参考链接：

1. 玻尔兹曼因数与配分函数 Boltzmann Factor & Partition Function. https://zhuanlan.zhihu.com/p/103512926
2. 指数族分布|机器学习推导系列（九） https://blog.csdn.net/weixin_42431920/article/details/107960335

根据指数族分布的定义$p(x;\eta)=b(x)\exp\big(\eta^T T(x)-a(\eta)\big)=\frac{b(x)}{e^{a(\eta)}}\exp\left(-\eta^T T(x)\right)$，再观察玻尔兹曼分布的式子，对比二者，不难得出：

自然参数η为β，而充分统计学量$T(x)$应为$-H(x)$。

对数似然函数为

$$\ln\mathcal{L}_{\theta,D}=\ln P(x)=\ln\frac{1}{Z}\exp\big(-\beta H(x)\big)=-\ln Z-\beta H(x)$$

其中

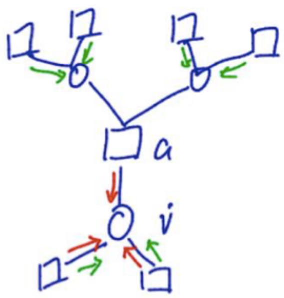
$$Z=\sum_x e^{-\beta H(x)}=\sum_x e^{\beta\left(\sum_{i<j<k}J_{i,j,k}x_ix_jx_k\right)}$$

故

$$\begin{aligned}
\frac{\partial\ln\mathcal{L}_{\theta,D}}{\partial J_{i,j,k}}&=-\frac{\partial\ln Z}{\partial J_{i,j,k}}-\frac{\partial\beta H(x)}{\partial J_{i,j,k}}\\
&=-\frac{\partial\ln Z}{\partial Z}\left(\frac{\partial Z}{\partial\big(-\beta H(x)\big)}\frac{\partial\big(-\beta H(x)\big)}{\partial J_{i,j,k}}\right)+\beta\mathbb{E}_{x\sim D,i<j<k}\left(x_ix_jx_k\right)\\
&=-\frac{1}{Z}\sum_x e^{\beta\left(\sum_{i<j<k}J_{i,j,k}x_ix_jx_k\right)}\beta\left(x_ix_jx_k\right)_{i<j<k}+\beta\mathbb{E}_{x\sim D,i<j<k}\left(x_ix_jx_k\right)\\
&=-\beta\left(\mathbb{E}_{x\sim P_\theta,i<j<k}\left(x_ix_jx_k\right)-\mathbb{E}_{x\sim D,i<j<k}\left(x_ix_jx_k\right)\right)
\end{aligned}$$

第一项为服从假设的分布的均值，第二项为服从真实的分布的均值。

题目 2: 参考下图，并利用边缘概率相融条件$b_i(x_i) = \sum_{x_j, j\in N(a)\setminus i} b_a\left(x_j \middle| j \in N(a)\right)$推导消息传递方程$m_{a\to i}(x_i)$的更新表达式。



参考链接：

1. 13 : Variational Inference: Loopy Belief Propagation. https://www.cs.cmu.edu/~epxing/Class/10708-14/scribe_notes/scribe_note_lecture13.pdf

解答：

易知

$$b_i(x_i) = \sum_{x_j, j\in N(a)\setminus i} b_a\left(x_j \middle| h \in N(a)\right) \propto f_i(x_i) \prod_{a\in N(i)} m_{a\to i}(x_i)$$

$$b_a(X_a) \propto f_a(X_a) \prod_{i\in N(a)} \prod_{c\in N(i)\setminus a} m_{c\to i}(x_i)$$

且有$m_{a\to i}(x_i) = \sum_{X_a\setminus x_i} b_a(X_a)$

那么可以推出消息传递更新公式为：

$$m_{a\to i}(x_i) = \sum_{X_a\setminus x_i} f_a(X_a) \prod_{j\in N(a)\setminus i} \prod_{(b\in N(j)\setminus a)} m_{b\to j}(x_j)$$

题目 3:简述张量网络与因子图模型的相同点与不同点

1. 因子图模型可以转化为张量网络。
2. 因子图模型的主要任务在于推断、估计边缘分布，在经典统计力学中应用较多。
3. 因子图模型与玻尔兹曼分布关系密切，例子便是 Ising 模型。
4. 张量网络和因子图都可以用图来表示。不同的是两者图中节点和边的含义不同。因子图代表的是和概率相关的概念。而张量网络中，每一个节点代表了一个张量，边代表的是指标，而每一个指标连接了一个或两个张量。

5. 张量网络可以看成一种特殊的因子图模型。每一个因子都是一个张量，每一个变量就是一个指标。变量仅连接一或两个张量。
6. 张量网络和因子图模型可以进行平均量和边缘概率分布计算。