Statistical Machine Learning

Homework4

方言

2021210929

2022 年 6 月 8 日

1 PGM

1.1 Problem 1

1.1.1 Write down an expression for $P(S=1|V=1)$ in terms of $\alpha, \beta, \gamma, \delta$.

由图可知:

$$P(S=1|V=1) = \alpha(1-\gamma) + (1-\alpha)(1-\beta)$$

1.1.2 Write down an expression for $P(S=1|V=0)$ in terms of $\alpha, \beta, \gamma, \delta$.

由图可知:

$$P(S=1|V=0) = \alpha(1-\gamma) + (1-\alpha)(1-\beta)$$

对比可知:

$$P(S=1|V=0) = P(S=1|V=1)$$

这是由于在此概率图模型中，S 只由 G 决定，而 V 与 G 相互独立，则 V 与 S 是相互独立的。

1.1.3 Find maximum likelihood estimates of α, β, γ using the following data set.

其 MLE 为:

$$\begin{cases} \alpha = \dfrac{1}{3} \\ \beta = 0 \\ \gamma = 1 \end{cases}$$

2 Approximate Inference

2.1 Problem 2

由于:

$$\mathrm{KL}(p||q) = \int p(x) \cdot \log \frac{p(x)}{q(x)}\, dx$$

当 $\alpha \to 1$ 时，$1-\alpha \to 0$，此时:

$$
\begin{aligned}
D_\alpha(p||q) &= \frac{4}{1-\alpha^2}\left(1-\int p(x)^{(1+\alpha)/2} q(x)^{(1-\alpha)/2}\,dx\right) \\
&= \frac{4}{1-\alpha^2}\left(1-\int p(x)\cdot\left(\frac{q(x)}{p(x)}\right)^{(1-\alpha)/2}\,dx\right) \\
&= \frac{4}{1-\alpha^2}\left(1-\int p(x)\cdot\exp\left(\frac{1-\alpha}{2}\log\frac{q(x)}{p(x)}\right)\,dx\right) \\
&= \frac{4}{1-\alpha^2}\left(1-\int p(x)\left(1+\frac{1-\alpha}{2}\log\frac{q(x)}{p(x)}+O((\frac{1-\alpha}{2})^2)\right)\,dx\right) \\
&\to \frac{4}{1-\alpha^2}\left(1-\int p(x)dx-\int p(x)\frac{1-\alpha}{2}\log\frac{q(x)}{p(x)}\,dx\right) \\
&= \frac{4}{1-\alpha^2}(1-1)-\frac{2}{1+\alpha}\int p(x)\log\frac{q(x)}{p(x)}\,dx \\
&= \int p(x)\log\frac{p(x)}{q(x)}\,dx \\
&= \mathrm{KL}(p||q)
\end{aligned}
$$

因此，当 $\alpha \to 1$ 时，$D_\alpha(p||q) \to \mathrm{KL}(p||q)$

同理可知，当 $\alpha \to -1$ 时，$1+\alpha \to 0$，此时:

$$
\begin{aligned}
D_\alpha(p||q) &= \frac{4}{1-\alpha^2}\left(1-\int p(x)^{(1+\alpha)/2} q(x)^{(1-\alpha)/2}\,dx\right) \\
&= \frac{4}{1-\alpha^2}\left(1-\int q(x)\cdot\left(\frac{p(x)}{q(x)}\right)^{(1+\alpha)/2}\,dx\right) \\
&= \frac{4}{1-\alpha^2}\left(1-\int q(x)\cdot\exp\left(\frac{1+\alpha}{2}\log\frac{p(x)}{q(x)}\right)\,dx\right) \\
&= \frac{4}{1-\alpha^2}\left(1-\int q(x)\left(1+\frac{1+\alpha}{2}\log\frac{p(x)}{q(x)}+O((\frac{1+\alpha}{2})^2)\right)\,dx\right) \\
&\to \frac{4}{1-\alpha^2}\left(1-\int q(x)dx-\int q(x)\frac{1+\alpha}{2}\log\frac{p(x)}{q(x)}\,dx\right) \\
&= \frac{4}{1-\alpha^2}(1-1)-\frac{2}{1-\alpha}\int q(x)\log\frac{p(x)}{q(x)}\,dx \\
&= \int q(x)\log\frac{q(x)}{p(x)}\,dx \\
&= \mathrm{KL}(q||p)
\end{aligned}
$$

因此，当 $\alpha \to -1$ 时，$D_\alpha(p||q) \to \mathrm{KL}(q||p)$

3 Deep Generative Models

3.1 Problem 3

3.1.1 Show that $\text{IWAE}(1) \le \text{IWAE}(K) \le \log p(x;\theta)$

由于从 q 分布中采样 K 个，再从中等概率采样一个，等价于直接从 q 分布中采样一个，因此，借助琴森不等式:

$$\begin{aligned}
\text{IWAE}(K) &= \mathbb{E}_{q(z^K|x;\phi)}\left[\log\left(\frac{1}{K}\sum_{k=1}^{K}\frac{p(z^{(k)},x;\theta)}{q(z^{(k)}|x;\phi)}\right)\right] \\
&\ge \mathbb{E}_{q(z^K|x;\phi)}\left[\frac{1}{K}\sum_{k=1}^{K}\log\left(\frac{p(z^{(k)},x;\theta)}{q(z^{(k)}|x;\phi)}\right)\right] \\
&= \mathbb{E}_{q(z|x;\phi)}\left[\log\left(\frac{p(z,x;\theta)}{q(z|x;\phi)}\right)\right] \\
&= \text{IWAE}(1)
\end{aligned}$$

进一步，由琴森不等式:

$$\begin{aligned}
\text{IWAE}(K) &= \mathbb{E}_{q(z^K|x;\phi)}\left[\log\left(\frac{1}{K}\sum_{k=1}^{K}\frac{p(z^{(k)},x;\theta)}{q(z^{(k)}|x;\phi)}\right)\right] \\
&\le \log\mathbb{E}_{q(z^K|x;\phi)}\left[\frac{1}{K}\sum_{k=1}^{K}\frac{p(z^{(k)},x;\theta)}{q(z^{(k)}|x;\phi)}\right] \\
&= \log\mathbb{E}_{q(z|x;\phi)}\left[\frac{p(z,x;\theta)}{q(z|x;\phi)}\right] \\
&= \log p(x;\theta)
\end{aligned}$$

综上:

$$\text{IWAE}(1) \le \text{IWAE}(K) \le \log p(x;\theta)$$

3.1.2 Show that $\lim_{K\to\infty}\text{IWAE}(K) = \log p(x;\theta)$

当 K 很大，趋于无穷时，由大数定律，$\frac{1}{K}\sum_{k=1}^{K}\frac{p(z^{(k)},x;\theta)}{q(z^{(k)}|x;\phi)}$ 会收敛到其均值，即其期望:

$$\frac{1}{K}\sum_{k=1}^{K}\frac{p(z^{(k)},x;\theta)}{q(z^{(k)}|x;\phi)} \to \mathbb{E}_{q(z|x;\phi)}\left[\frac{p(z,x;\theta)}{q(z|x;\phi)}\right] = \log p(x;\theta)$$

因此:

$$\text{IWAE}(K) = \mathbb{E}_{q(z^k|x;\phi)}\left[\frac{1}{K}\sum_{k=1}^{K}\frac{p(z^{(k)},x;\theta)}{q(z^{(k)}|x;\phi)}\right] \to \mathbb{E}\left[\log p(x;\theta)\right] = \log p(x;\theta)$$

3.1.3 Show that if $K < L$, $\text{IWAE}(K) \le \text{IWAE}(L)$

从 $\{1,2,..,L\}$ 等概率采样 K 个，构成集合 $I \subset \{1,2,..,L\}$。显然，此时:

$$\mathbb{E}_{I=\{i_1,i_2,\ldots,i_K\}}\left[\frac{1}{K}\sum_{k=1}^{K}f(z^{(i_k)})\right] = \frac{1}{L}\sum_{l=1}^{L}f(z^{(i_l)})$$

由此，借助琴森不等式，可得:

$$
\begin{aligned}
\mathrm{IWAE}(L) &= \mathbb{E}_{q(z^L|x;\phi)}\left[\log\left(\frac{1}{L}\sum_{l=1}^{L}\frac{p(z^{(l)},x;\theta)}{q(z^{(l)}|x;\phi)}\right)\right] \\
&= \mathbb{E}_{q(z^L|x;\phi)}\left[\log\left(\mathbb{E}_{I=\{i_1,i_2,\dots,i_K\}}\left[\frac{1}{K}\sum_{k=1}^{K}\frac{p(z^{(i_k)},x;\theta)}{q(z^{(i_k)}|x;\phi)}\right]\right)\right] \\
&\geq \mathbb{E}_{q(z^L|x;\phi)}\left[\mathbb{E}_{I=\{i_1,i_2,\dots,i_K\}}\left[\log\left(\frac{1}{K}\sum_{k=1}^{K}\frac{p(z^{(i_k)},x;\theta)}{q(z^{(i_k)}|x;\phi)}\right)\right]\right] \\
&= \mathbb{E}_{q(z^K|x;\phi)}\left[\log\left(\frac{1}{K}\sum_{k=1}^{K}\frac{p(z^{(k)},x;\theta)}{q(z^{(k)}|x;\phi)}\right)\right] \\
&= \mathrm{IWAE}(K)
\end{aligned}
$$

3.2 Problem 4

$$
\begin{aligned}
\mathbb{E}_{p(n)}\left[\mathrm{KL}(q(z|n)||p(z))\right] &= \mathbb{E}_{p(n)}\left[\int q(z|n)\log\frac{q(z|n)}{p(z)}dz\right] \\
&= \frac{1}{N}\sum_{n=1}^{N}\int q(z|n)\log\frac{q(z|n)}{p(z)}dz \\
&= \frac{1}{N}\sum_{n=1}^{N}\int q(z|n)\left(\log\frac{q(z|n)}{q(z)}+\log\frac{q(z)}{p(z)}\right)dz \\
&= \frac{1}{N}\sum_{n=1}^{N}\int q(z|n)\log\frac{q(z|n)}{q(z)}dz + \int\frac{1}{N}\sum_{n=1}^{N}q(z|n)\log\frac{q(z)}{p(z)}dz \\
&= \sum_{n=1}^{N}\int q(z,n)\log\frac{q(z,n)}{q(z)p(n)}dz + \int q(z)\log\frac{q(z)}{p(z)}dz \\
&= \mathrm{KL}\left(q(z,n)||q(z)p(n)\right) + \int q(z)\left(\log\frac{q(z)}{\prod_j q(z_j)}+\log\frac{\prod_j q(z_j)}{\prod_j p(z_j)}\right)dz \\
&= \mathrm{KL}\left(q(z,n)||q(z)p(n)\right) + \int q(z)\log\frac{q(z)}{\prod_j q(z_j)}dz_1,\dots dz_J + \int q(z)\log\frac{\prod_j q(z_j)}{\prod_j p(z_j)}dz_1,\dots dz_J \\
&= \mathrm{KL}\left(q(z,n)||q(z)p(n)\right) + \mathrm{KL}\left(q(z)||\prod_j q(z_j)\right) + \int q(z_1,\dots,z_J)\log\frac{\prod_j q(z_j)}{\prod_j p(z_j)}dz_1,\dots dz_J \\
&= \mathrm{KL}\left(q(z,n)||q(z)p(n)\right) + \mathrm{KL}\left(q(z)||\prod_j q(z_j)\right) + \sum_j\int q(z_j)\log\frac{q(z_j)}{p(z_j)}dz_j \\
&= \mathrm{KL}\left(q(z,n)||q(z)p(n)\right) + \mathrm{KL}\left(q(z)||\prod_j q(z_j)\right) + \sum_j \mathrm{KL}\left(q(z_j)||p(z_j)\right)
\end{aligned}
$$

3.3 Problem 5

首先求解:

$$
\max_y f(y) = \max_y \; a\log(y) + b\log(1-y), \quad a,b\in[0,1]
$$

上式对 y 求导，并令其等于 0 可得:

$$
f'(y) = \frac{a}{y} - \frac{b}{1-y} = 0
$$

求解得:

$$y^\star = \frac{a}{a+b}$$

由于固定 G，因此转变为求解:

$$\begin{aligned}
&\max_D \left(\mathbb{E}_{p_{\text{data}}(x)}[\log D(x)] + \mathbb{E}_{p(z)}[\log(1 - D(G(z)))]\right) \\
=&\max_D \left(\mathbb{E}_{p_{\text{data}}(x)}[\log D(x)] + \mathbb{E}_{p_{\text{model}}(x)}[\log(1 - D(x))]\right) \\
=&\max_D \left(\int p_{\text{data}}(x)\log D(x)dx + \int p_{\text{model}}(x)\log(1 - D(x))dx\right) \\
=&\int \max_D \left(p_{\text{data}}(x)\log D(x) + p_{\text{model}}(x)\log(1 - D(x))\right) dx
\end{aligned}$$

因此，D 的最优解为:

$$D^\star(x) = \frac{p_{\text{data}}(x)}{p_{\text{data}}(x) + p_{\text{model}}(x)}$$