Deep Learning 2023040163 Zhao Han Hong

Homework 3

1 True or False Questions

Problem 1

False.

Problem 2

True.

2 Q & A

Problem 3

We have the variance being

$$\mathrm{Var}\left[\frac{1}{N}\sum_{x\sim q}\frac{p(x)}{q(x)}f(x)\right]=\frac{1}{N^2}\cdot N\cdot \mathrm{Var}_{x\sim q}\left[\frac{p(x)}{q(x)}f(x)\right]$$
$$=\frac{1}{N}\left(\int\frac{p(x)^2f(x)^2}{q(x)}dx-\left(\int p(x)f(x)dx\right)^2\right).$$

Since

$$\int\frac{p(x)^2f(x)^2}{q(x)}dx\int q(x)dx\geq\left(\int p(x)|f(x)|dx\right)^2,$$

and the equality holds if and only if $q(x)\propto p(x)|f(x)|$, we know that the variance is minimized when $q(x)\propto p(x)|f(x)|$.

Problem 4

(1) Suppose that the sampling gives

$$T(s\to s')=c\exp\left(-\frac{(s-s')^2}{\sigma^2}\right),$$

we can immediately find that the Markov Chain satisfies the detailed balance property since we can choose $\pi(s)$ such that

$$\frac{T(s \to s')}{T(s' \to s)} = \frac{\pi(s')}{\pi(s)}.$$

Moreover, we can check the ergodicity of the Markov Chain by checking that

$$\min_z \min_{\pi(z')>0} \frac{T(z \to z')}{\pi(z')} = c' \exp\left(-\frac{(s-s')^2}{\sigma^2}\right) > 0.$$

Thus, it is a valid Markov chain.

(2) Since the way of updating is

$$q(s_i \to s_i') = p(s_i'|s_{j\neq i}), \alpha(s_i \to s_i') = \min\left(1, \frac{p(s_i')q(s_i' \to s_i)}{p(s_i)q(s_i \to s_i')}\right),$$

and the other parts of the algorithms are the same, we know that Gibbs sampling is a case of Metropolis-Hasting sampling.

Now, we only have to calculate the acceptance rate. In fact, we have

$$\frac{p(s_i')q(s_i' \to s_i)}{p(s_i)q(s_i \to s_i')} = \frac{p(s_i')}{p(s_i)}\frac{p(s_i|s_{j\neq i}')}{p(s_i'|s_{j\neq i})} = \frac{p(s_i')}{p(s_i)}\frac{p(s_i)}{p(s_i')} = 1.$$

Thus, we know that Gibbs sampling is a case of Metropolis-Hasting sampling, and the acceptance rate is always 1.

(3)

Problem 5

(1)

(a) False. Since $\mathbb{P}(A,C) = \mathbb{P}(A)\int_B \mathbb{P}(C|B)\mathbb{P}(B|A)dB$, we have $\mathbb{P}(C|A) = \int_B \mathbb{P}(C|B)\mathbb{P}(B|A)dB$, which may depend on A.

(b) True. Given B,

$$\mathbb{P}(A,C|B) = \frac{\mathbb{P}(A,B,C)}{\mathbb{P}(B)} = \frac{1}{\mathbb{P}(B)}\mathbb{P}(A)\mathbb{P}(B|A)\mathbb{P}(C|B) = \mathbb{P}(A|B)\mathbb{P}(C|B),$$

where $\mathbb{P}(A|B), \mathbb{P}(C|B)$ are functions only depending on A, C given B.

(c) False. We have

$$\mathbb{P}(A, C|D) = \frac{\mathbb{P}(A, C, D)}{\mathbb{P}(D)} = \frac{1}{\mathbb{P}(D)}\mathbb{P}(A)\mathbb{P}(D|A, C)\int_B \mathbb{P}(B|A)\mathbb{P}(C|B)dB,$$

this can't be written as $\mathbb{P}(A)\mathbb{P}(C)$ in general.

(d) True. Since A, C are independent given B, they must also be independent given B, D.

(e) False. In general, we have

$$\mathbb{P}(B, D) = \int_A \mathbb{P}(A)\mathbb{P}(B|A)dA \int_C \mathbb{P}(D|A, C)\mathbb{P}(C|B)dC,$$

which may not be able to be rewritten as $\mathbb{P}(B)\mathbb{P}(D)$.

(f) False. Given A,

$$\mathbb{P}(B, D|A) = \frac{\mathbb{P}(A, B, D)}{\mathbb{P}(A)} = \mathbb{P}(B|A)\int_C \mathbb{P}(D|A, C)\mathbb{P}(C|B)dC,$$

which may not be able to written as $\mathbb{P}(B|A)\mathbb{P}(D|A)$, in general.

(g) False. Given C,

$$\mathbb{P}(B, D|C) = \frac{1}{\mathbb{P}(C)}\mathbb{P}(C|B)\int_A \mathbb{P}(A)\mathbb{P}(B|A)\mathbb{P}(D|A, C)dA,$$

which may not be able to written as $\mathbb{P}(B|A)\mathbb{P}(D|A)$, in general.

(h) True. We have

$$\mathbb{P}(B, D|A, C) = \frac{1}{\mathbb{P}(A, C)}\mathbb{P}(A)\mathbb{P}(B|A)\mathbb{P}(D|A, C) = \mathbb{P}(B|A, C)\mathbb{P}(D|A, C),$$

where $\mathbb{P}(B|A, C), \mathbb{P}(D|A, C)$ are functions only depending on B, D given A, C.

(2) The likelihood is

$$\begin{aligned}\mathbb{P}(B, C, D|A) &= \mathbb{P}(B|A)\mathbb{P}(C|B)\mathbb{P}(D|A, C) \\ &= \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(B-A)^2}{2}\right)\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(C-B)^2}{2}\right)\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(D-(C+A))^2}{2}\right) \\ &= \frac{1}{(2\pi)^{\frac{3}{2}}}\exp\left(-\frac{2A^2+2B^2+2C^2+D^2}{2} + (D+B)(C+A) - AC\right).\end{aligned}$$

For the posterior, we have

$$\begin{aligned}
\mathbb{P}(A|B,C,D) &= \frac{\mathbb{P}(A,B,C,D)}{\mathbb{P}(B,C,D)} \\
&= \frac{1}{\mathbb{P}(B,C,D)}\frac{1}{4\pi^2}\exp\left(-\frac{3A^2+2B^2+2C^2+D^2}{2}+(D+B)(C+A)-AC\right),
\end{aligned}$$

and we can calculate

$$\begin{aligned}
\mathbb{P}(B,C,D) &= \int_A \mathbb{P}(B,C,D|A)\mathbb{P}(A)dA \\
&= \int_A \frac{1}{4\pi^2}\exp\left(-\frac{3A^2+2B^2+2C^2+D^2}{2}+(D+B)(C+A)-AC\right)dA \\
&= \frac{1}{\sqrt{3}(2\pi)^{\frac{3}{2}}}\exp\left(-\frac{5}{6}(B^2+C^2)-\frac{1}{3}D^2+\frac{1}{3}BD+\frac{2}{3}C(B+D)\right).
\end{aligned}$$

Thus, we have

$$\mathbb{P}(A|B,C,D) = \sqrt{\frac{3}{2\pi}}\exp\left(-\frac{3}{2}\left(A-\frac{D+B-C}{3}\right)^2\right).$$

Problem 6

(1) Let the kernel be $(2k+1)\times(2k+1)$. Then the pixel at (x_0, y_0) can be only influenced by the pixel lowest as (x_0+k, y_0+1). Thus, $(x_0+1, y_0), (x_0+2, y_0), \cdots, (x_0+k, y_0)$ can not influence (x_0, y_0). We can then replace x_0, y_0 by x_0+k, y_0+1 and find that the pixel at $(x_0+k+1, y_0+1), \cdots, (x_0+2k, y_0+1)$ can't influence (x_0, y_0). Notice that $(x_0+k+1, y_0), \cdots, (x_0+2k, y_0)$ also can't influence (x_0, y_0). We may repeat this process for times and conclude that for y_0+l, the pixels $(x_0+kl+1, y_0+l), \cdots, (x_0+(l+1)k, y_0+l)$ can't influence (x_0, y_0). We then get a sawtooth-shaped receptive field.

(2) We can mimic the method of Gated PixelCNN, which uses both a vertical stack and a horizontal stack to calculate the generating pixels and avoid blind spots.

For each layer computation, the first step is to calculate the vertical stack. On that stack, we define a kernel of size $2k\times(k+1)$ such that the vertical stack value $z(x, y)$ of pixel (x, y) depends on $z([x-k : x+k], [y : y+k])$. The next step is letting the final value $f(x, y)$ of pixel (x, y) depend horizontally on $f(x-1, y), \cdots, f(x-k, y)$ and the vertical stack value $z(x, y+1)$. In summary, our per-layer computation process is:

$$z(x,y) = \sum_{i=-k}^{k} \sum_{j=0}^{k} z(x+i, y+j) w_{ij} \qquad (w_{00} = 0);$$
$$f(x,y) = \sum_{i=1}^{k} f(x-i, y) w'_{ij} + z(x, y+1).$$

We now demonstrate that this computation process will not lead to blind spots while maintaining the autoregressive property. In fact, we can find that $z(x, y)$ depends on all $z(x_1, y_1)$ where $y_1 \geq y$. Thus, $f(x, y)$ depends on all $z(x_1, y_1)$ where $y_1 > y$. Moreover, we can notice that $f(x, y)$ depends on $f(x-1, y), \cdots, f(x-k, y)$. This makes sure that $f(x, y)$ depends on all the previous values. Moreover, we can clearly see that it maintains the autoregressive property since there is no way for $f(x, y)$ to depend on the values at pixel (x_1, y_1) where $x_1 > x, y_1 = y$ or $y_1 > y$.