DL 2023040165 Zhicheng Jiang

Assignment 3

Problem 1

False.

Problem 2

True.

Problem 3

For simplicity, denote $E = \mathbb{E}_{x\sim p} f(x) = \mathbb{E}_{x\sim q} \frac{p(x)f(x)}{q(x)}$ Then, we have

$$\mathbb{V}_{x\sim q} \frac{p(x)f(x)}{q(x)} = \mathbb{E}_{x\sim q}[(\frac{p(x)f(x)}{q(x)} - E)^2]$$

$$= \mathbb{E}_{x\sim q}[(\frac{p(x)f(x)}{q(x)})^2 - 2E \cdot E + E^2]$$

$$= \mathbb{E}_{x\sim q}[(\frac{p(x)f(x)}{q(x)})^2] - E^2$$

$$= \int q(x) \frac{p(x)f(x)}{q(x)} \frac{p(x)f(x)}{q(x)} dx - E^2$$

$$= \int \frac{p^2(x)f^2(x)}{q(x)} dx \int q(x) dx - E^2$$

$$\geq (\int p(x)|f(x)|dx)^2 - E^2$$

The Cauchy-Schwarz inequality is used in the last step, and equality holds if and only if $\frac{p(x)|f(x)|}{q(x)}$ is a constant with probability 1.

Problem 4

1.

We check that $p(z)$ is a stationary distribution of the random walk Metropolis Hasting Algorithm:

Detailed balance: We know that the transition probability of the random walk Metropolis Hasting Algorithm is

$$T(s \to s') = \frac{e^{-(s'-s)^2}}{Z} min(1, \frac{p(s')}{p(s)})$$

so, we can easily check that the detailed balance holds:

$$p(s)T(s \to s') = \frac{e^{-(s'-s)^2}}{Z} p(s) min(1, \frac{p(s')}{p(s)})$$

$$= \frac{e^{-(s'-s)^2}}{Z} min(p(s), p(s')) = \frac{e^{-(s'-s)^2}}{Z} p(s') min(1, \frac{p(s)}{p(s')}) = p(s')T(s' \to s)$$

Ergodic:

$$\forall z, z'(p(z') > 0),$$

$$\frac{T(z \to z')}{p(z')} = \frac{e^{-(s'-s)^2}}{Z} \frac{\min(\frac{p(z')}{p(z)}, 1)}{p(z')} = \frac{e^{-(s'-s)^2}}{Z} \min(\frac{1}{p(z)}, \frac{1}{p(z')}) \geq \frac{R}{P}$$

where $P = \max_z p(z), R = \frac{e^{-(Range)^2}}{Z}$

2. It's easy to see that this is a vaild Markov chain, so it's a special case of the Metropolis-Hastings algorithm.

The acceptance probability is

$$\alpha(s \to s') = \min(1, \frac{p(s')q(s' \to s)}{p(s)q(s \to s')})$$

However, by the definition of Gibbs sampling, we have

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

So, the acceptance rate is 1.

3. WLOG suppose there're n dimensions, and a "round" is defined as n Gibbs sampling steps among all the n dimensions. By the condition, we know that after several rounds, the transition probability is positive for every pair of states. Let these rounds be a "Big Round". Consider a Markov chain that uses every "Big Round" as a step, we only need to prove that this Markov chain has a stable distribution $p(z)$, as we can start from any step of the "Big Round", to get the desired distribution.

For the detailed balance part, by the equation in 2, we know that during each transition, the detailed balance holds. So, it also holds for the "Big Round".

For the ergodic part, we know that the transition probability is positive for every pair of states, since there're a finite number of states, thus, the ergodicity holds.

Problem 5

1. The joint probability distribution is

$$\mathbb{P}(A,B,C,D) = \mathbb{P}(A)\mathbb{P}(B|A)\mathbb{P}(C|B)\mathbb{P}(D|A,C) \tag{*}$$

TLDR: the answer for the following questions is Dep;Ind;Dep;Dep;Dep;Dep;Dep;Ind
(a)
A and C can be dependent(e.g. A=B=C=D)
(b)
A and C are independent given B, because by (*),

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

(c)
A and C can be dependent given D(e.g. A=B=C, D random)
(d)
A and C can be dependent given B,D(e.g. A,B,C are i.i.d, D=[A==C])
(e)
B and D can be dependent(e.g. A=B=C=D)
(f)
B and D can be dependent given A(e.g. B=C=D, and independent of A)
(g)
B and D can be dependent given C(e.f. A=B=D, and independent of C)
(h)
B and D are independent given A and C, because by (*),

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

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

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

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

(2)

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

$$= \frac{\exp(\frac{-(B-A)^2-(C-B)^2-(D-C-A)^2}{2})}{\sqrt{8\pi^3}}$$

$$\mathbb{P}(A|B,C,D) = \frac{e^{-[a^2+(b-a)^2+(c-b)^2+(d-a-c)^2]/2}}{\int_a e^{-[a^2+(b-a)^2+(c-b)^2+(d-a-c)^2]/2}}$$
$$= \frac{e^{-[\frac{3}{2}(a+(c-d-b)/3)^2]}}{\sqrt{2\pi/3}}$$

Problem 6

1.

We induct on i*n+j to prove that the final receptive field of (i,j) is $R_{i,j}$={(i,[1:j]),(i-1,[1:j+1]),(i-2,[1:j+2]),...,(i+j-n,[1:n]),(i+j-n-1,[1:n]),...,(1,[1:n])}.

It's easy to check that the base case holds, and $R_{i,j} = \cup_v R_{(i,j)-v} + (i,j)$, where $v \in \{(0,[1:k]),([1:k],[-k:k])\}$, and denote the R_v to be empty set if v is out of bound.

2.

We modify the mask by getting the middle line to be all zero, denote this new mask to be m'. Also, we add a new "horizontal stack", which has a mask m'' of all zero, except for the first half part of the middle line(including the middle element).

The convolutional layer is represented by

$$x' = conv(m' \cdot w, x) + z$$

And z is the horizontal stack, which is intialized as x, and updated by

$$z = conv(m'' \cdot w, z) + z$$