Homework 3

Deep Learning 2024 Spring

Due on 2024/4/13

1 True or False

Problem 1. Stochastic Gradient MCMC is designed to solve the optimization problem $\arg\max_\theta \mathbb{P}(\theta|\mathsf{X})$, where θ is the collection of parameters and X represents data.

Problem 2. The convolution used in GLOW and WaveNet should be both invertible (i.e., the convolution kernels W should be invertible) since they are all flow models.

2 Q&A

Problem 3. (Importance Sampling) x is a random variable. Given target distribution $p(\mathsf{x})$ and target random variable $\mathsf{y} = f(\mathsf{x})$, importance sampling gives an estimator of $\mathbb{E}[\mathsf{y}]$ from a proposal distribution $q(\mathsf{x})$:

$$\mathbb{E}_{\mathsf{x}\sim p}\left[f(\mathsf{x})\right] = \mathbb{E}_{\mathsf{x}\sim q}\left[\frac{p(\mathsf{x})}{q(\mathsf{x})}f(\mathsf{x})\right] \approx \frac{1}{N}\sum_{x\sim q(\cdot)}\frac{p(x)}{q(x)}f(x).$$

Prove that when q has the following form,

$$q^\star(\mathsf{x}) \propto p(\mathsf{x})|f(\mathsf{x})|$$

the variance of this estimator can be minimized.

Problem 4. (Markov Chain Monte Carlo)

1. Prove random-walk Metropolis-Hasting sampling (i.e., $\mathsf{s}' \leftarrow \mathsf{s} +$ Gaussian noise) is a valid MCMC algorithm, i.e., it constructs a Markov chain which is ergodic and satisfies the detailed balance property.

2. Prove that Gibbs sampling is a special case of Metropolis-Hasting sampling, and that the acceptance rate of Gibbs sampling (i.e., $\alpha(\mathsf{s} \to \mathsf{s}')$) is 1.

   Here we consider the following 2-step Gibbs proposal: (1) randomly sample a coordinate index i; (2) sample coordinate s_i from the coordinate proposal $q(\mathsf{s}_i \to \mathsf{s}'_i) = p(\mathsf{s}'_i|\mathsf{s}_{j\neq i})$.

3. (**Bonus Question**) In fact, Gibbs sampling is typically implemented in a *cyclic fashion*, i.e., running posterior sampling in a fixed order over all the dimensions. Prove that cyclic Gibbs sampling yields the same stationary distribution as random-order Gibbs sampling in the above question, as long as the Markov chain can access all states under the fixed ordering.

Problem 5. (Directed Probabilistic Model)

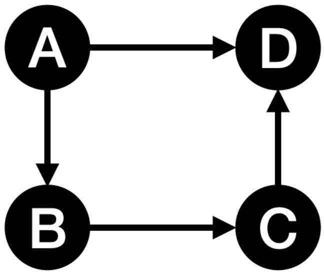


Figure 1: An example of directed graphical model.

Continuing Problem 6 in the previous homework, we consider **directed probabilistic model** in this problem. In a directed graphical model, an edge (arrow) implies dependency of the downstream random variable to the upstream one. Figure 1 shows an example, where the joint probability distribution can be factorized as

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

1. Judge whether the following statements regarding independence are true or not. $\mathsf{X} \perp \mathsf{Y} \,|\, \mathsf{Z}$ denotes that X and Y are independent given Z.

   (a) $\mathsf{A} \perp \mathsf{C}$

   (b) $\mathsf{A} \perp \mathsf{C} \,|\, \mathsf{B}$

   (c) $\mathsf{A} \perp \mathsf{C} \,|\, \mathsf{D}$

   (d) $\mathsf{A} \perp \mathsf{C} \,|\, \mathsf{B}, \mathsf{D}$

   (e) $\mathsf{B} \perp \mathsf{D}$

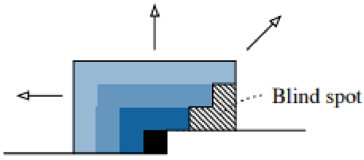
   (f) $\mathsf{B} \perp \mathsf{D} \,|\, \mathsf{A}$

   (g) $\mathsf{B} \perp \mathsf{D} \,|\, \mathsf{C}$

   (h) $\mathsf{B} \perp \mathsf{D} \,|\, \mathsf{A}, \mathsf{C}$

2. Suppose the conditional distributions are given by

$$\mathbb{P}(\mathsf{B}|\mathsf{A} = a) = \mathcal{N}(a, 1)$$
$$\mathbb{P}(\mathsf{C}|\mathsf{B} = b) = \mathcal{N}(b, 1)$$
$$\mathbb{P}(\mathsf{D}|\mathsf{A} = a, \mathsf{C} = c) = \mathcal{N}(c + a, 1).$$



(a) A sawtooth-shaped receptive field.

1	1	1	1	1
1	1	1	1	1
1	1	0	0	0
0	0	0	0	0
0	0	0	0	0

(b) A PixelCNN causal mask.

The prior distribution of A is a Gaussian distribution with zero mean and unit variance. Derive the likelihood $\mathbb{P}(\mathsf{B}, \mathsf{C}, \mathsf{D}|\mathsf{A})$ and posterior $\mathbb{P}(\mathsf{A}|\mathsf{B}, \mathsf{C}, \mathsf{D})$ in closed form.

Problem 6. (PixelCNN) PixelCNN[1] is an auto-regressive generative model based on masked convolution kernels. Please answer the following questions:

1. Show that masked convolution kernels induce sawtooth-shaped receptive fields, as shown in Figure 2a.
2. An obvious flaw of PixelCNN is that the generated pixel can not condition on all the left and upper pixels due to the sawtooth-shaped receptive field, which is called the issue of *blind spot*. Denote the unmasked kernel as w, the corresponding mask as m (as shown in Figure 2b), and the image/feature map as x. The computation of a PixelCNN layer can be represented as

$$y = \mathrm{conv}(m \cdot w, x)$$

Propose a minimal modification on PixelCNN to remove the blind spot while maintaining the auto-regressive property. Write down your per-layer computation process.

[1] `https://arxiv.org/abs/1601.06759`