Homework 6

Deep Learning 2025 Spring

Due on 2025/4/14

1 Q&A

Problem 1. Let $\hat{p}(x)$ be an uncalibrated Gaussian autoregressive model defined over $x \in \mathbb{R}^d$. The true density function for discrete data is constructed by integrating $\hat{p}(x)$ over the hypercube $[0,1)^d$, i.e.,

$$p(x) = \int_{[0,1)^d} \hat{p}(x+u)\, du$$

where x lies on a discrete lattice (e.g., integer grid). To train $\hat{p}(x)$, a data augmentation scheme is applied: for each data point x sampled from the empirical data distribution $p_{\text{data}}(x)$, a noise vector u is drawn uniformly from $[0,1)^d$, and the perturbed sample $x' = x+u$ is generated. The training objective maximizes the expected log-likelihood of the augmented data:

$$\mathcal{L}(\theta) = \mathbb{E}_{x'} \left[\log \hat{p}(x')\right],$$

where the expectation is taken over the augmented distribution induced by $p_{\text{data}}(x)$ and $u \sim [0,1)^d$.
Prove that:

$$\mathcal{L}(\theta) \leq \mathbb{E}_{x \sim p_{\text{data}}(x)} \left[\log p(x)\right].$$

Problem 2. (Autoregressive Flow)
Consider a generative model for sequential data $\mathbf{x} = [x_1, x_2, \ldots, x_L]$ of fixed length L defined by the following autoregressive factorization:

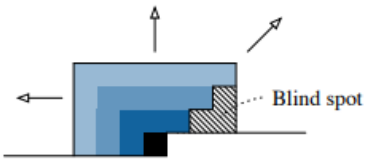
$$p(\mathbf{x}) = \prod_{i=1}^{L} p(x_i | x_1, \ldots, x_{i-1}),$$

where each conditional $p(x_i|x_{<i})$ is a Gaussian:

$$p(x_i|x_{<i}) = \mathcal{N}\left(x_i \,|\, \mu_i(x_{<i}),\, \exp\left(\alpha_i(x_{<i})\right)^2\right).$$

Assume that μ_i and α_i are computed by a causal convolutional neural network, such as WaveNet.
Let the base distribution $p_{\mathbf{U}}(\mathbf{u})$ be standard Gaussian, i.e., $\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. Define a transformation $T : \mathbb{R}^L \to \mathbb{R}^L$ such that:

$$x_i = \mu_i(x_{<i}) + \exp\left(\alpha_i(x_{<i})\right) \cdot u_i \quad \text{for } i = 1, 2, \ldots, L.$$



(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. This mask is used in the first convolution layer in PixelCNN. The rest layers use a similar mask, with the only difference being that it does not mask the center pixel.

Please prove that this transformation T is a valid normalizing flow.

Problem 3. (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 1a.

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 1b), and the image/feature map as x. The computation of a PixelCNN layer can be represented as

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

   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`