In The Name of God

Sharif University of Technology

Electrical Engineering Department

Deep Generative Models

Assignment 3

Fall 2024

Instructor: Dr. S. Amini

Due on Azar 11, 1403 at 23:55

1 Probability Review

Consider the transformation $f : \mathbb{R}^3 \to \mathbb{R}^3$.

$$X = f(Z) = [z_1 \quad z_2 e^{z_1} \quad (z_3 e^{-z_1} + z_1^2)^{\frac{1}{3}}]$$

Is this an invertible transformation? What's the Jacobian of this transformation? Now assume we have a Multivariate Normal base distribution $p(Z) \sim N_3(0, I)$ on the domain of f. What's $p(x)$ at $x = [0 \ 1 \ 1/3]^T$ ($p(X)$ and $p(Z)$ are probability density functions)

2 VP-NFs are not universal

Remember a Normalizing Flow formulated as:

$$p_\theta(\mathbf{x}) = p(z = f_\theta^{-1}(\mathbf{x})) \left| \det \frac{\partial f_\theta^{-1}(\mathbf{x})}{\partial \mathbf{x}} \right|$$

In a Volume-Preserving Normalizing Flow (VP-NF), the determinant of the Jacobian matrix of each transformation is constant, specifically:

$$\left| \det \frac{\partial f_\theta(\mathbf{z})}{\partial \mathbf{z}} \right| = 1$$

This implies that the transformation preserves volume in the latent space, meaning the model can rearrange the probability mass but cannot scale it.

A set of probability distributions $\mathcal{P}$ is called a distributional universal approximator if for every possible target distribution $p(x)$ there is a sequence of distributions $p_n(x) \in \mathcal{P}$ such that $\lim_{n\to\infty} p_n(x) \sim p(x)$ with a chosen type of convergence.

We want to prove: The family of Normalizing Flows with constant Jacobian determinant $\left| \det \frac{\partial f_\theta(\mathbf{z})}{\partial \mathbf{z}} \right| = const$ with latent $z \sim \mathcal{N}(0, 1)$ is not a universal distribution approximator under KL divergence. We do this by finding a lower bound for the KL objective. Use the Pinsker's inequality as below and the counter example given:

$$\delta(p_\theta, q) = \sup_{E\text{: a measurable event}} \{|p_\theta(E) - q(E)|\} \le \sqrt{\frac{1}{2} D_{\mathrm{KL}}(p\|q)}$$

counter example:

$$p(x,y) = \begin{cases} 0.9 & \text{if } (x,y) \in [-0.5, 0.5] \times [-0.5, 0.5], \\ 0.9 - k \cdot (|x| - 0.5) & \text{if } |x| \in [0.5, 0.9k + 0.5] \text{ and } |y| \in [0, |x|], \\ 0.9 - k \cdot (|y| - 0.5) & \text{if } |y| \in [0.5, 0.9k + 0.5] \text{ and } |x| \in [0, |y|], \\ 0 & \text{otherwise.} \end{cases}$$

$$A = \left\{(x,y) \in \mathbb{R}^2 : p_\theta(x,y) \geq 0.9 - \epsilon\right\}$$

$$B = [-0.5, 0.5] \times [-0.5, 0.5]$$

$$\bar{A} = B \setminus A$$

2.1 Part (a)

Prove there exist an event E such that $|p_\theta(E) - p(E)| > 0$ when $A = \emptyset$, and show the lower bound for KL divergence can be positive for some $\epsilon > 0$, using Pinkster's inequality. (hint: prove $\bar{A}$ can be such E)

2.2 Part(b)

Using the change of variables formula, create the event C in latent space resulting from A. Compare its volume with A. Then compute its volume and provide a simple upper bound for it. (hint: volume here is a 2D circle because of the normal distribution)

2.3 Part(c)

Now show that there is an event E such that $|p_\theta(E) - p(E)| > 0$ when $A \neq \emptyset$, and show the lower bound for KL divergence can be positive for some $\epsilon > 0$, using Pinkster's inequality

2.4 Part(d)

Now briefly explain why this means VP-NFs cant approximate every distribution.

3 Optimality of GAN Framework

In class, we covered that the optimal discriminator satisfies the following condition

$$D_G^*(\boldsymbol{x}) = \frac{p_{\text{data}}(\boldsymbol{x})}{p_{\text{data}}(\boldsymbol{x}) + p_g(\boldsymbol{x})} \tag{1}$$

for arbitrary values of $\boldsymbol{x}$. In this problem, you will prove the optimality of this discriminator. To be more specific, for a fixed generator θ, show that $D_\phi^*(\boldsymbol{x})$ minimizes the following loss function:

$$\mathcal{L}(\phi;\theta) = -\mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}}\left[\log D_\phi(\boldsymbol{x})\right] - \mathbb{E}_{\boldsymbol{x}\sim p_\theta(\boldsymbol{x})}\left[\log\left(1 - D_\phi(\boldsymbol{x})\right)\right] \tag{2}$$

4 Perfect Discriminator, Perfect Generation

Recall that when training GAN, the gradient signals for θ come from the term

$$\mathbb{E}_{\boldsymbol{x}\sim p_\theta(\boldsymbol{x})}\left[\log\left(1 - D_\phi(\boldsymbol{x})\right)\right], \tag{3}$$

which can also be written as

$$L_G(\theta;\phi) = \mathbb{E}_{\boldsymbol{z}\sim\mathcal{N}(\mathbf{0},\mathbf{I})}\left[\log\left(1 - \sigma\left(h_\phi\left(G_\theta(\boldsymbol{z})\right)\right)\right)\right], \tag{4}$$

where $\sigma(\cdot)$ denotes the sigmoid function, $h_\phi(\cdot)$ denotes the logits, and G_θ is the generator. Now assume that our discriminator is perfect, i.e. $\nabla_\theta L_G(\theta;\phi) \to 0$, show that the gradient of θ would vanish to zero in this case, i.e. $\nabla_\theta L_G(\theta;\phi) \to 0$.