Minimax Rate of Distribution Estimation On Unknown Submanifolds Under Adversarial Loss

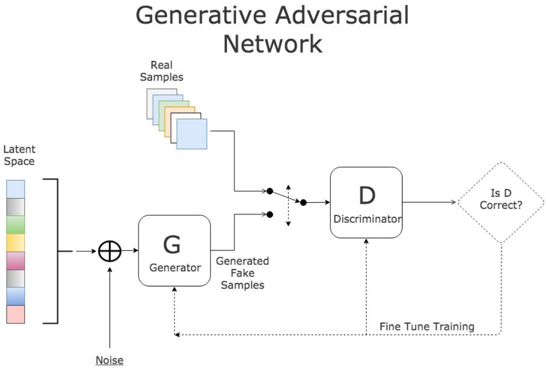
Sharan Sahu, Raghav Ramanujam

University of California, Berkeley
ssahu01@berkeley.edu, *rramanujam@berkeley.edu*

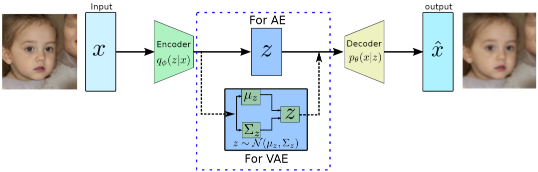
May 4, 2023

Overview

Motivation



Generative Adversarial Network
Real Samples
Latent Space
D
Discriminator
Is D Correct?
G
Generator
Generated Fake Samples
Fine Tune Training
Noise



Input
x
Encoder
$q_\phi(z|x)$
For AE
z
μ_z
Σ_z
z
$z \sim \mathcal{N}(\mu_z, \Sigma_z)$
For VAE
Decoder
$p_\theta(x|z)$
output
$\hat{x}$

Section 2

Mathematical Definitions and Problem Formulation

Generative Models

Definition: Generative Models

A **generative model** is a (ν, G) pair where ν is a distribution on a low-dimensional latent space $\mathcal{Z} \subset \mathbb{R}^d$ called the **generative distribution**, and $G : \mathcal{Z} \to \mathbb{R}^d$ called the **generative map**.

Comments:

1. Generative models decouples distribution estimation into manifold learning (estimation of G) + density estimation on the manifold (estimation of ν)
2. Generating samples from an underlying distribution is more important and useful than estimating the distribution
3. Generative models can capture highly nonlinear structures that may lead to singularities (such as jumps and point mass) in the distribution and are hard to characterize via a density or distribution function

Adversarial Loss

Definition: Adversarial Loss

With respect to **discriminator class** F of bounded and Borel-measurable functions, the **adversarial loss** between two probability measures μ and ν is given by:

$$d_{\mathcal{F}}(\mu, \nu) = \sup_{f \in \mathcal{F}} \left| \int_{\mathbb{R}^D} f(x) d\mu - \int_{\mathbb{R}^D} f(x) d\nu \right|$$

Comments:

1. Useful when candidates distributions have different supports
2. Many statistics can be defined as an integral of f w.r.t underlying measure
3. Computational ease due to empirical sampling of $d_{\mathcal{F}}$

γ-smooth Hölder class

Definition: γ-smooth Hölder class

The γ-**smooth Hölder class** with radius $r > 0$ over input space Ω as

$$C_r^\alpha(\Omega) = \{f : \Omega \to \mathbb{R} \mid \|f\|_{C_r^\alpha(\Omega)} = \sum_{|a| \leq \lfloor \alpha \rfloor} \max_{x \in \Omega} |f^{(a)}(x)|$$

$$+ \sum_{|a| = \lfloor \alpha \rfloor} \max_{x, y \in \Omega, x \neq y} \frac{f^{(a)}(x) - f^{(a)}(y)}{\|x - y\|^{\alpha - \lfloor \alpha \rfloor}} \}$$

The vector valued function space counterpart is given by

$$C_r^\alpha(\Omega; \mathbb{R}^d) = \{f = (f_1, \cdots, f_D) : \Omega \to \mathbb{R}^D \mid \forall j \in [D], f_j \in C_r^\alpha(\Omega)\}$$

Hölder class and Adversarial Loss

Note: Adversarial Loss w.r.t. γ-smooth Hölder class

In this problem, adversarial loss with respect to the γ-smooth Hölder class on the unit ball is used, and is given by

$$d_\gamma(\mu,\nu) = \sup_{f \in C_r^\alpha(\mathbb{R}^d)} \left| \int_{\mathbb{R}^D} f(x) d\mu - \int_{\mathbb{R}^D} f(x) d\nu \right|$$

Comments:

1. d_γ is a valid probability metric (triangle inequlity, Weierstrauss approximation implies $d_\gamma(\mu,\nu) = 0$ iff $\mu = \nu$)
2. When restricted to bounded set, equivalent to Wasserstein-1 when $\gamma = 1$, approaches TV distance as $\gamma \to 0_+$
3. Smaller $\gamma \implies$ higher sensitivity to support misalignment

Smooth Submanifolds

Definition: β-smooth d-dimensional manifold $\mathcal{M}$

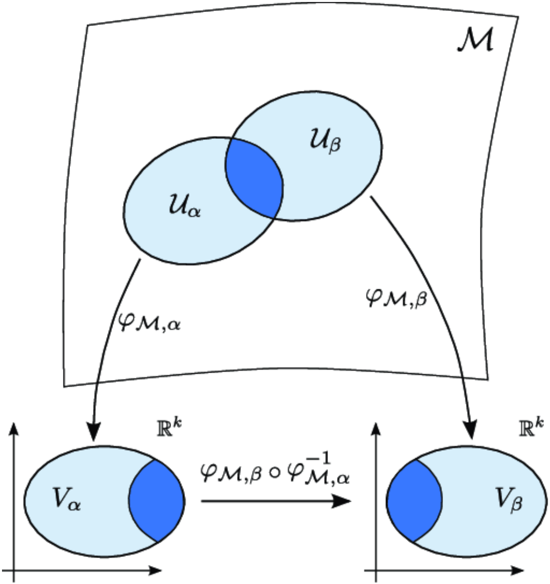
A β**-smooth** d**-dimensional manifold** $\mathcal{M}$ is a topological space satisfying the following properties:

1. There exists an atlas on $\mathcal{M}$ consisting of a collection of d-dimensional charts $\mathscr{A} = \{(U_\lambda, \varphi_\lambda)\}_{\lambda \in \Lambda}$ such that $\mathcal{M} = \bigcup_{\lambda \in \Lambda} U_\lambda$
2. Each chart in (U, φ) in atlas $\mathscr{A}$ consists of a homeomorphism $\varphi : U \to \tilde{U}$ called a coordinate map from the open set $U \subset M$ to an open set $\tilde{U} \subset \mathbb{R}^d$. That is, φ is bijective and both φ and φ^{-1} are continuous
3. Any two charts (U, φ) and (V, ψ) in atlas $\mathscr{A}$ are compatible. That is, the transistion map $\varphi \circ \psi^{-1} : \psi(U \cap V) \to \varphi(U \cap V)$ is C^β-diffeomorphic

Comments:

1. Homeomorphic $\not\Longrightarrow$ Diffeomorphic. Consider $f(x) = x^3$

Smooth Submanifolds Illustration



Problem Formulation

Setup: Minimax estimation on submanifold

In the rest of this presentation, we will establish the minimax rate of convergence for the adversarial risk on $\mu^* \in \mathcal{P}^*$ of distribution estimation on an unknown submanifold with i.i.d examples $X_1, \cdots, X_n \sim u^*$

Section 3

Minimax Rates

Minimax Rate Of Distribution Estimation

Theorem: Minimax Rate of Distribution Estimation

Fix $L^* > 0$, $\gamma \geq 0$, $0 \leq \alpha \leq \beta - 1$, $\beta > 1$, and $D, d \in \mathbb{N}^+$ with $D > d$. Write $\mathcal{P}^* = \mathcal{P}^*(L^*, \gamma, \alpha, \beta, d, D)$. Then,

1. there exists a constant L_0 such that when $L^* \geq L_0$, then

$$\inf_{\hat{\mu} \in \mathcal{P}(\mathbb{R}^{\mathcal{D}})} \sup_{\mu \in \mathcal{P}^*} \mathbb{E}\left[d_\gamma(\hat{\mu}, \mu)\right] \geq C n^{-\frac{1}{2}} \vee n^{\frac{-\alpha+\gamma}{2\alpha+d}} \vee n^{-\frac{\gamma\beta}{d}}$$

2. there exists positive constants L_1, L_2 such that for any $L \geq L_1$ and open cover $\mathscr{O}_M = \{\mathbb{B}_{r_m}(a_m)^\circ\}_{m \in [M]}$ of $\mathbb{B}_L^D$ with $\max\{r_1, r_2, \dots, r_M\} \leq L_2$, it holds that

$$\inf_{\hat{\mu} \in S_{\nu_0}^{\mathrm{ap}}} \sup_{\mu \in \mathcal{P}^*} \mathbb{E}\left[d_\gamma(\hat{\mu}, \mu)\right] \leq C \left(\frac{n}{\log n}\right)^{-\frac{1}{2}} \vee \left(\frac{n}{\log n}\right)^{\frac{-\alpha+\gamma}{2\alpha+d}} \vee \left(\frac{n}{\log n}\right)^{-\frac{\gamma\beta}{d}}$$

Minimax Rate Of Distribution Estimation

Intuition and Proof Insight Sketch For Lower Bound

For obtaining the lower bound, we use the standard Fano's and Le Cam's method by identifying a subset of distributions within the considered distribution family $\mathcal{P}^*(L^*, \gamma, \alpha, \beta, d, D)$ that are statistically hard to distinguish. Here is a breakdown of the methods used to obtain the bound

1. $n^{-\frac{\gamma\beta}{d}}$ reflects the statistical hardness of estimating an unknown β-smooth submanifold [Use Fano's Method]
2. $n^{-\frac{1}{2}} \vee n^{\frac{-\alpha+\gamma}{2\alpha+d}}$ reflects the statistical hardness of estimating an unknown α-smooth density as if the submanifold is known [Use Fano for first, Le Cam for second]

Minimax Rate Of Distribution Estimation

Intuition and Proof Insight Sketch For Upper Bound

For obtaining the upper bound, we make use of wavelets. We create a surrogate loss $\hat{\mathcal{J}} = \hat{\mathcal{J}}_l + \hat{\mathcal{J}}_s + \hat{\mathcal{J}}_h$ to approximate $\mathbb{E}\left[d_\gamma(\hat{\mu}, \mu)\right]$. Now, by using similar techniques we saw in Homework 1 with some clever tricks, we can obtain the upper bound

Comments:

1. To recall the techniques in Homework 1, we upper bounded the L^2 risk of a wavelet smoothing operator by its truncation, discretization, and estimation error. We can then upper-bound each one of these to get our final upper bound
2. The $\frac{n}{\log n}$ terms come from using Bernstein's inequality

Conclusion

1. We established a setup using a generative model structure and adversarial loss which generalizes well and gives us flexibility in our model based on our discriminator class choice
2. Walked through minimax rate of convergence for the aforementioned loss based on i.i.d samples from an unknown submanifold
3. Breakthrough is rate is not dependent on D, and therefore the results break the "curse of dimensionality"

References

1. Rong Tang and Yun Yang. Minimax rate of distribution estimation on unknown submanifold under adversarial losses. *ArXiv:2202.09030*, 2022.
2. Gharakhanian, Al. “Generative Adversarial Network.” Generative Adversarial Networks – Hot Topic in Machine Learning, 2016, https://www.kdnuggets.com/2017/01/generative-adversarial-networks-hot-topic-machine-learning.html. Accessed 1 May 2023.
3. Yadav, Sunil. “Variational Autoencoder.” Variational Autoencoders, 2021, https://data-science-blog.com/blog/2022/04/19/variational-autoencoders/. Accessed 1 May 2023.
4. Mimetic framework on curvilinear quadrilaterals of arbitrary order - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/Coordinate-charts-on-a-manifold_fig_51956714 [accessed 2 May, 2023]