AI – 395T – HOMEWORK – 4

EID: nk24865 – 4 (Q/A).

Q1. LIKELIHOOD VS. LIKELIHOOD FREE TRAINING

Why GANs are "Likelihood-Free"

GANs define an implicit distribution through a generator $x = T_\theta(\xi)$, with $\xi \sim P_0$ (e.g. Gaussian).

Unlike VAEs & normalizing flows, GANs dont provide a tractable likelihood $P_\theta(x)$ because T_θ is generally non-invertible and may map to a lower dimensional manifold.

Thus, maximum likelihood estimation (MLE) cannot be applied directly

$$P_\theta(x) = P_0\left(T_\theta^{-1}(x)\right) \left| \det \nabla T_\theta^{-1}(x) \right|$$

requires an invertible T_θ. When T_θ collapses dimensions this density is undefined w.r.t. Lebesgue measure – all generated samples lie on a surface, not in the full data space.

Density of $X = \dfrac{\xi}{\|\xi\|}$

Here $T(\xi) = \xi / \|\xi\|$ maps any

$\xi = R^d$ onto ~~any~~ the unit sphere S^{d-1}

- X has no density in R^d (since it support has measure zero)

- It does ~~not~~ have a density on the sphere (uniform over directions if ξ is isotopic)

SUMMARY : GANs avoid likelihood estimation by using adversarial training to compare samples instead of computing densities.

(Q3) DISCRIMINATIVE FUNCTION CLASS

Definition: A function class $\mathcal{H}$ is discriminative if

$$\sup_{h \in H} \left| E_p[h] - E_Q[h] \right| = 0$$

$$\Rightarrow \quad P = Q$$

This means the class is rich enough to distinguish any two different distributions.

Example :

- Discriminative : All bounded continuous functions - forms the basis of the total variation distance
- Non- Discriminative: Constant functions $h(x) = c$.

  here $E_p[h] - E_q[h] = 0$ for all P, Q; it cannot distinguish distributions

INTUTION :

The more expressive $\mathcal{H}$ is, the more sesitive the metric is to distributional differences. GAN critic aim to make $\mathcal{H}$ (parameterized by neural nets) sufficiently discriminative while still learnable.

Q4. Single-Neuron Test Function

Let $h(x) = \sigma(w^T x + b)$.

A single neuron can serve as a test function in an IPM or GAN critic.

Why Nonlinearity is Needed

If σ is linear ($\sigma(z)) = z$);

$$h(x) = w^T x + b \Rightarrow E_P[h] - E_Q[h] = w^T(E_P[x] - E_Q[x])$$

This measures only mean differences - insesitive to shape, variance, or multimodal structure.

When σ is nonlinear (eg. ReLU, tanh), $h(x)$ responds differently to various regions of the input space, allowing the critic to detect higher-order mismatches between P and Q

if σ = Identity

The model collapses to a linear test, equivaleent to matching only the first moment - making it non-discriminative.

Thus, nonlinearity gives richer feature tests, essential for expressive discriminators.

Q5 Integral Probability Metric (IPM)

Definition,

$$IPM_{\mathcal{H}}(P, Q) = \sup_{h \in \mathcal{H}} \left| E_P[h] - E_Q[h] \right|$$

it measures maximum discrepancy between expectations over a function class $\mathcal{H}$.

why Constrain h,

without constraints, scaling $h \to ch$ would make Supreme infinite.

To ensure a finite, meaningful distance, $\mathcal{H}$ is restricted.

- Lipschitz constraint ($\|\nabla h\| \leq 1$): gives Wasserstein distance.
- Bounded functions ($\|h\|_\infty \leq 1$): gives Total Variation,
- RKHS unit ball: gives Maximum Mean Discrepancy (MMD).

Interpretation,

IPM generalizes many distances. In GANs:

- Choosing a neural network for h_β approximates IPM with a learned fuction class.
- Regulization (like gradient penalty) enforces the smoothness constraint.