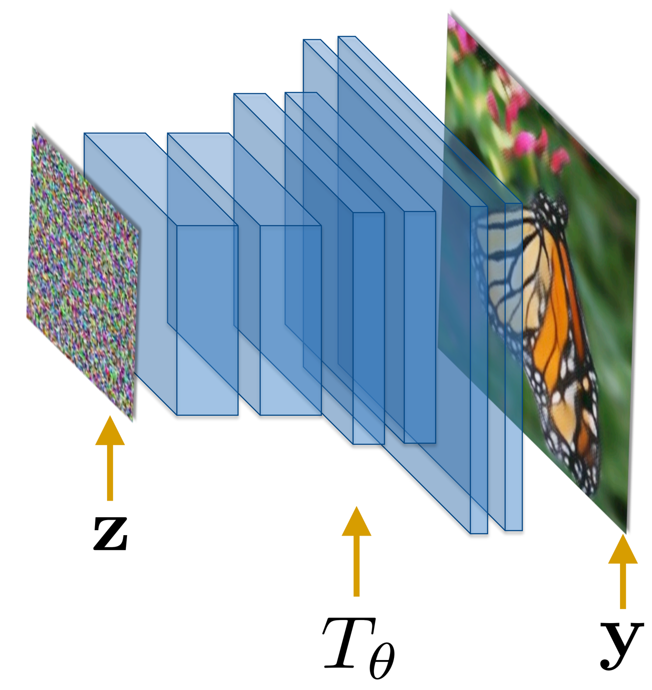
On the Implicit Assumptions of GANs

Berkeley
UNIVERSITY OF CALIFORNIA

Ke Li **Jitendra Malik**
{ke.li,malik}@eecs.berkeley.edu

Introduction

- Implicit probabilistic models: $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \quad \mathbf{y} = T_\theta(\mathbf{z})$
- Standard method for training such models: GANs.
- If we only have a *finite* number of data points, do the theoretical guarantees of GANs still hold (even if the discriminator were infinitely powerful and optimization were not an issue)?



True vs. Empirical Data Distribution

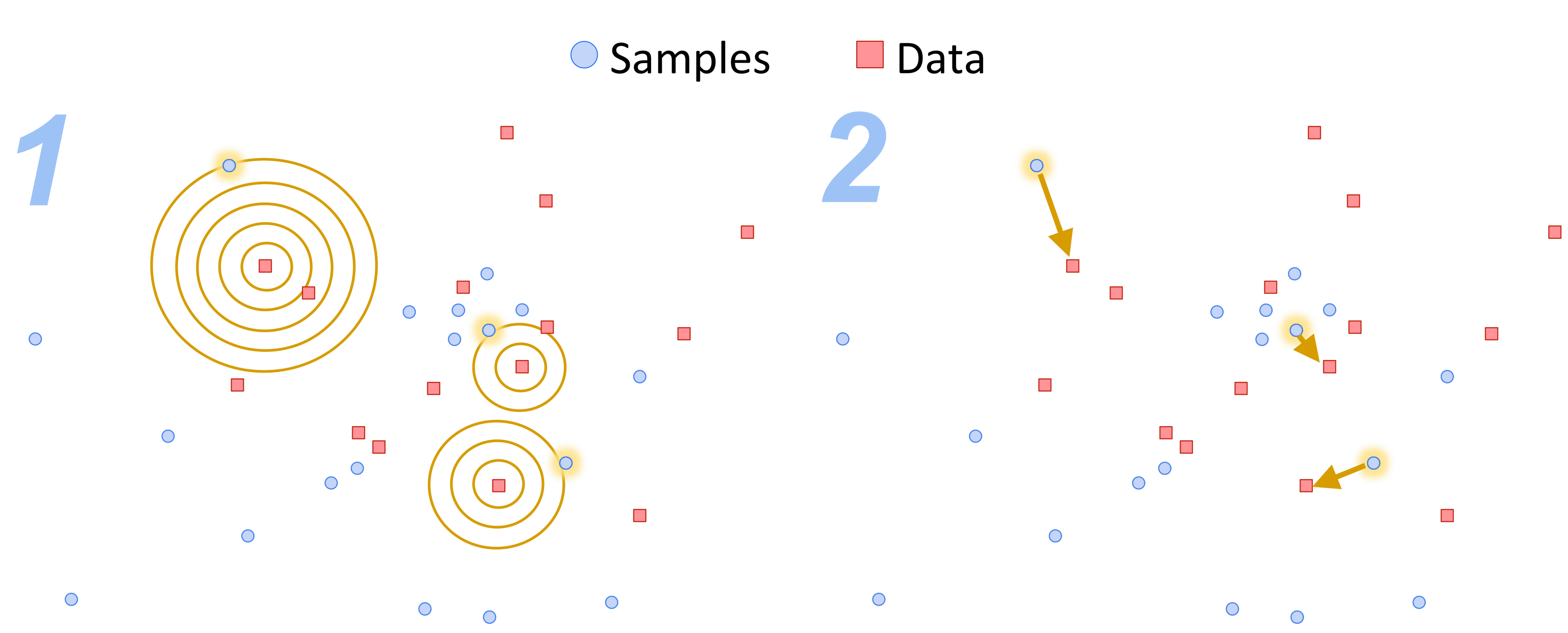
- GAN Objective:

$$\min_{\theta_G} \max_{\theta_D} \mathbb{E}_{\mathbf{x}\sim p_{\text{data}}} [\log D_{\theta_D}(\mathbf{x})] + \mathbb{E}_{\mathbf{z}\sim p_z} [\log (1 - D_{\theta_D}(G_{\theta_D}(\mathbf{z})))]$$

- Optimizing this requires drawing *fresh* samples from the true data distribution p_{data} in every iteration.
- But samples are drawn from a finite training set – this amounts to replacing p_{data} with $\widehat{p_{\text{data}}}$, the *empirical* data distribution.
- Jensen-Shannon divergence is always $\log 2$.
- Reverse KL-divergence $D_{KL}(p_\theta \| \widehat{p_{\text{data}}})$ is undefined.
- Implications:
  - Minimizing JSD or reverse KL does not make sense.
  - GANs are actually *not* asymptotically consistent.

Solution: Implicit Maximum Likelihood Estimation

- Maximum likelihood is consistent – can we maximize likelihood without computing likelihood?



Select the nearest *sample* to each *data point* (*NOT* the nearest data point to each sample).

Pull selected samples towards corresponding data points.

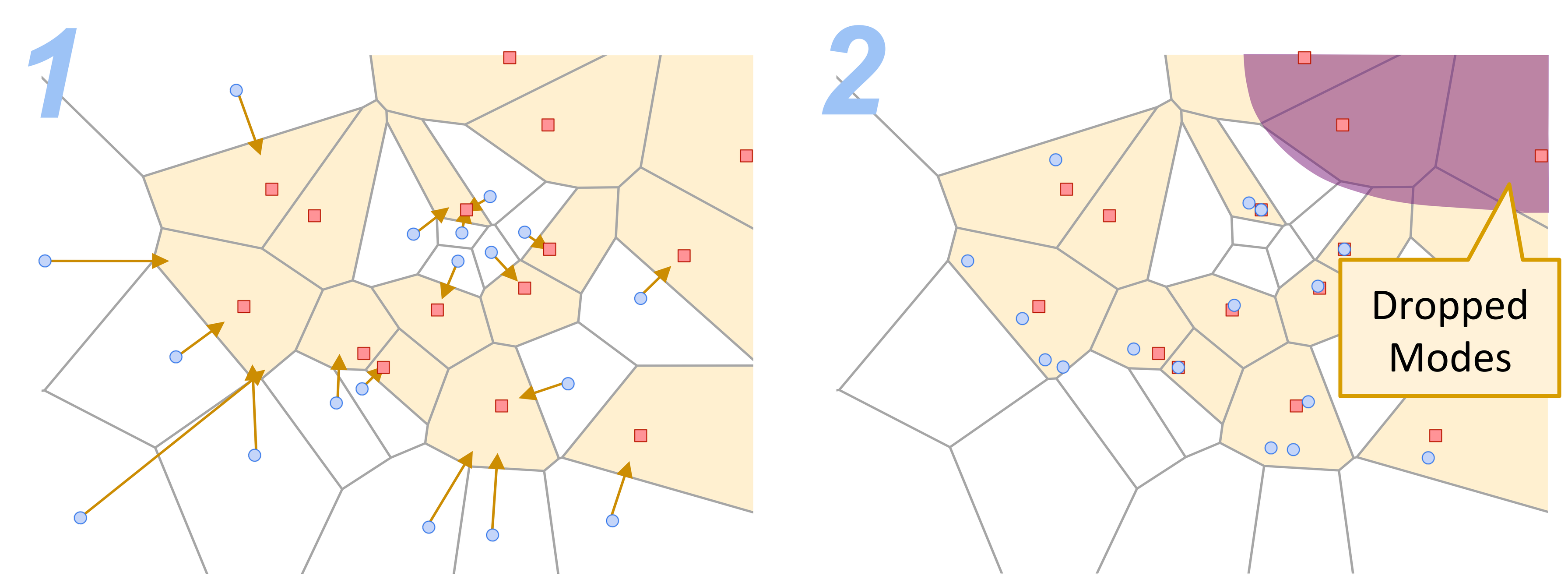
- Why? Maximize likelihood ⇔ High density at each data point ⇔ Samples likely to be near data points (Proof is in the IMLE paper)

~~Mode Dropping~~ ~~Vanishing Gradients~~ ~~Training Instability~~

Comparison to GANs

No More Mode Collapse/Dropping

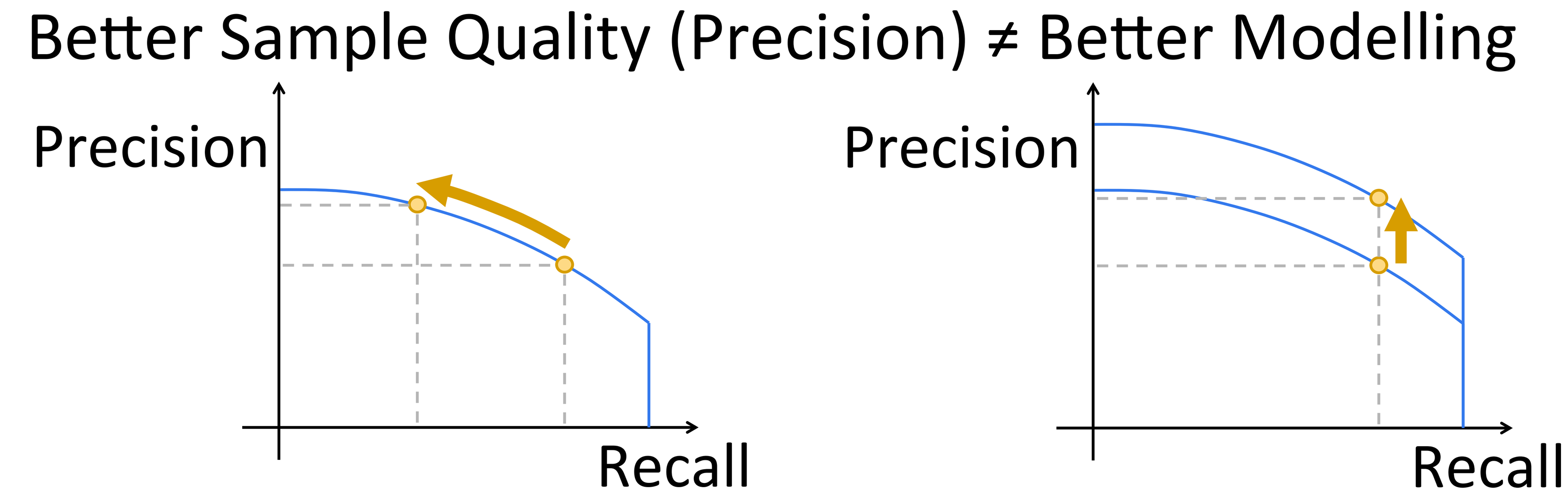
- GAN with a 1-nearest neighbour discriminator:



Push samples towards region containing real data.

Every sample has a nearby data point, but some data points may not have a nearby sample.

- Better Sample Quality (Precision) ≠ Better Modelling



Application: Multimodal Prediction

- Conditional setting: $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \qquad \mathbf{y} = T_\theta(\mathbf{x}, \mathbf{z})$
- Different samples for the same input image:
- *Multimodal Super-Resolution*

- *Multimodal Image Synthesis from Semantic Layout*

References

Ke Li and Jitendra Malik. On the Implicit Assumptions of GANs. *arXiv:1811.12402*, 2018

Ke Li and Jitendra Malik. Implicit Maximum Likelihood Estimation. *arXiv:1809.09087*, 2018

Ke Li*, Shichong Peng* and Jitendra Malik. Super-Resolution via Conditional Implicit Maximum Likelihood Estimation. *arXiv:1810.01406*, 2018

Ke Li*, Tianhao Zhang* and Jitendra Malik. Diverse Image Synthesis from Semantic Layouts via Conditional IMLE. *arXiv:1811.12373*, 2018