Implicit Maximum Likelihood Estimation

Ke Li **Jitendra Malik**

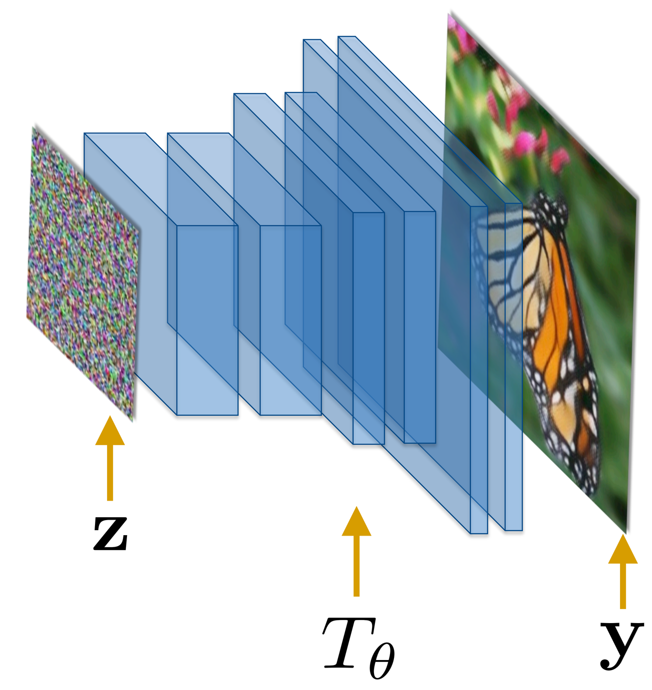
{ke.li,malik}@eecs.berkeley.edu

Berkeley UNIVERSITY OF CALIFORNIA

Introduction

- Implicit probabilistic models:

$$\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$
$$\mathbf{y} = T_\theta(\mathbf{z})$$



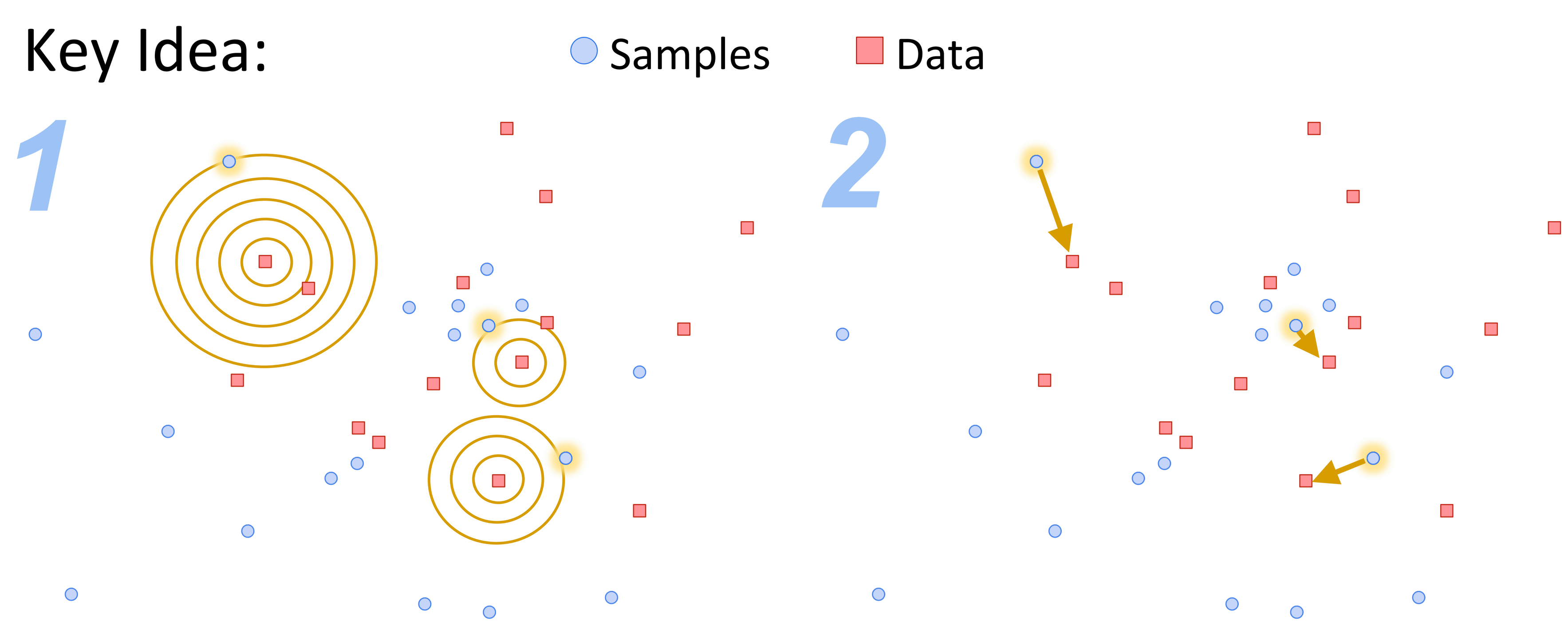
- Challenge: Likelihood function cannot be expressed analytically or computed numerically.
- Question: How to train such models?
- Existing Approach: Generative adversarial nets (GANs)

Mode Dropping | Vanishing Gradients | Training Instability

- Consequence: Unable to learn the data distribution.

Implicit Maximum Likelihood Estimation

- Can we maximize likelihood without computing likelihood?
- Key Idea:



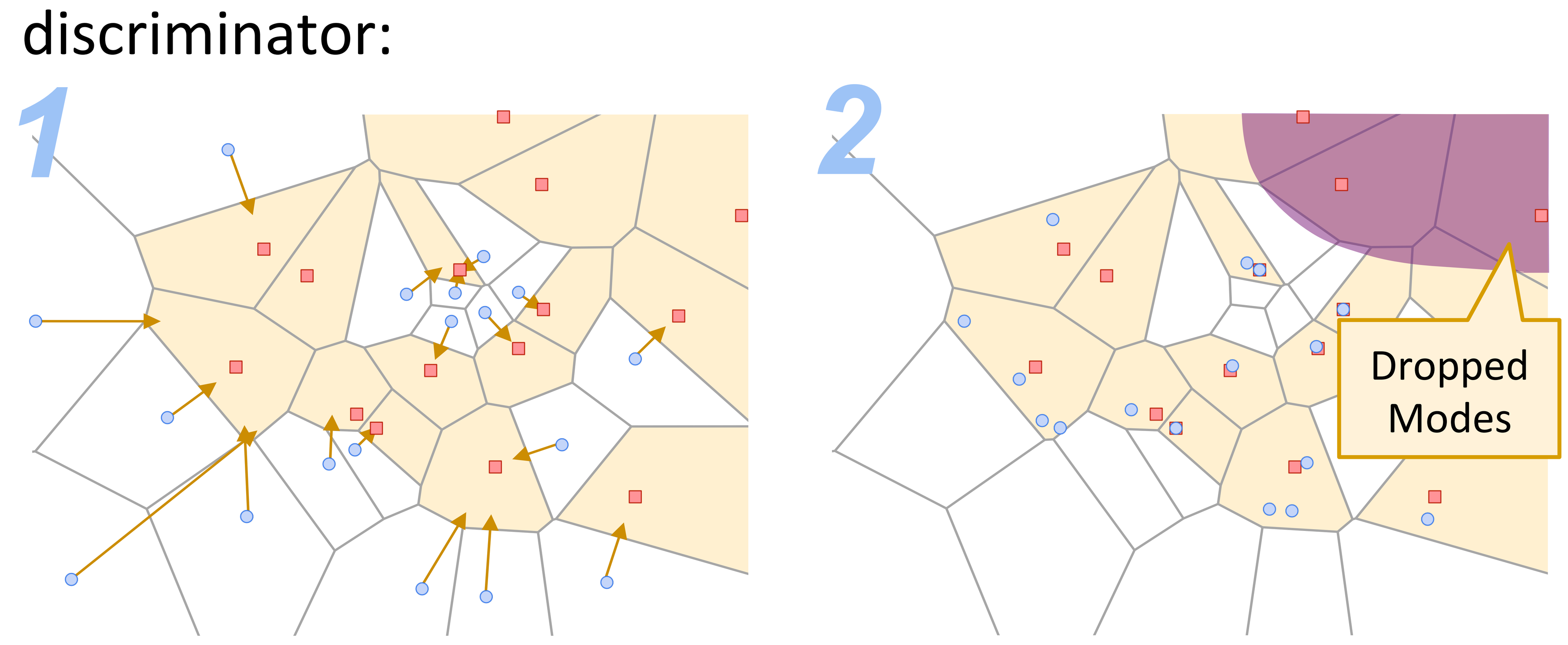
Select the nearest *sample* to each *data point* (*NOT* the nearest data point to each sample).

Pull selected samples towards corresponding data points.

$\mathbf{x}_1, \ldots, \mathbf{x}_n$: data points $\tilde{\mathbf{x}}_1^\theta, \ldots, \tilde{\mathbf{x}}_m^\theta$: i.i.d. samples

$$\hat{\theta}_{\text{IMLE}} := \arg\min_\theta \mathbb{E}_{\tilde{\mathbf{x}}_1^\theta, \ldots, \tilde{\mathbf{x}}_m^\theta}\left[\sum_{i=1}^{n} \min_{j\in[m]} \left\|\tilde{\mathbf{x}}_j^\theta - \mathbf{x}_i\right\|_2^2\right]$$

- Why? Maximize likelihood ⇔ High density at each data point ⇔ Samples likely to be near data points (Proof is in the paper)
- Difference from a GAN with a 1-nearest neighbour discriminator:



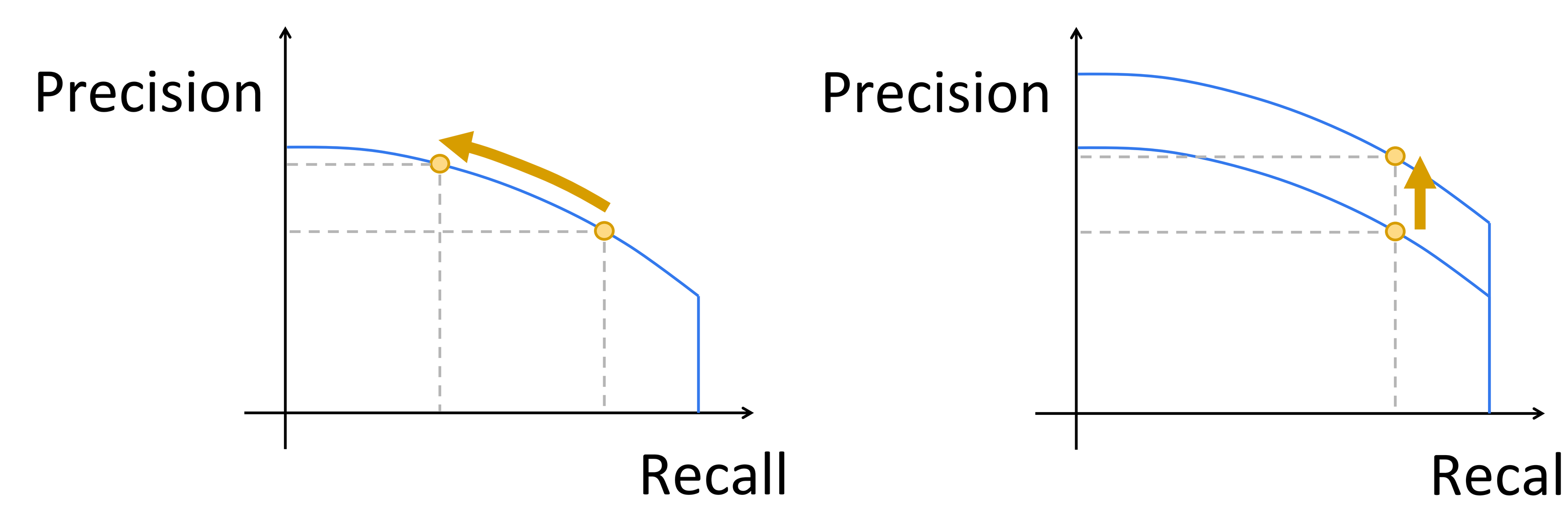
Push samples towards region containing real data.

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

Advantages

No More Mode Collapse/Dropping

- Why is this important? This allows control over recall.
- Otherwise better sample quality (precision) does not necessarily mean better modelling.
  - Wouldn't be able to tell apart the following cases:



- Overcomes all three problems of GANs:

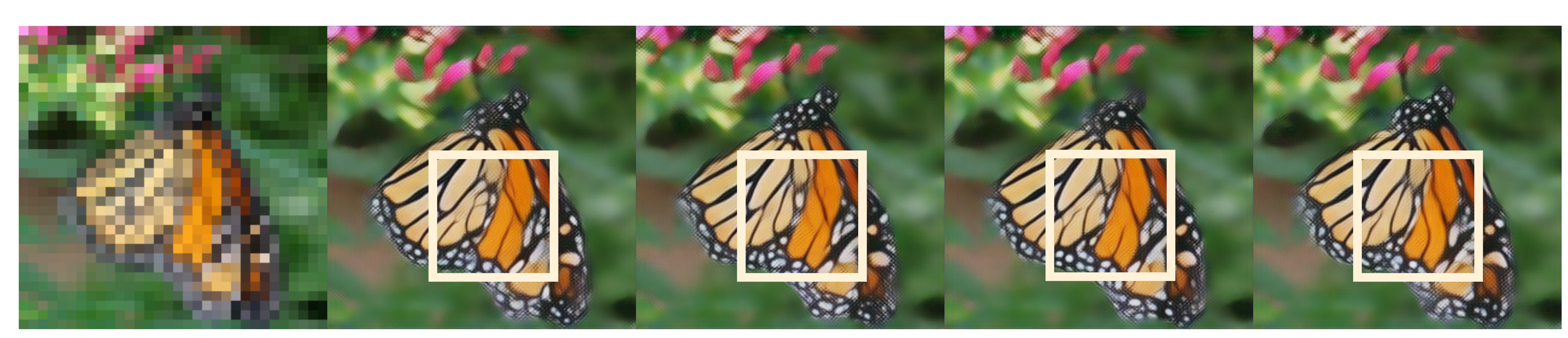
~~Mode Dropping~~ | ~~Vanishing Gradients~~ | ~~Training Instability~~

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. 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