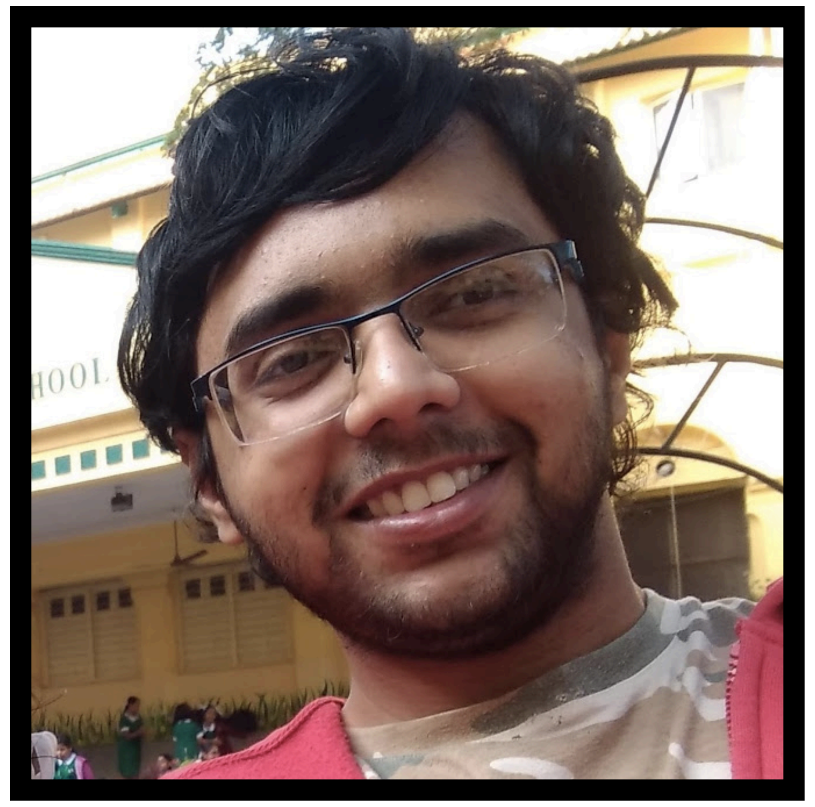
Learning Invariant Representations using Inverse Contrastive Loss

Aditya Akash



Vishnu Lokhande

Sathya Ravi

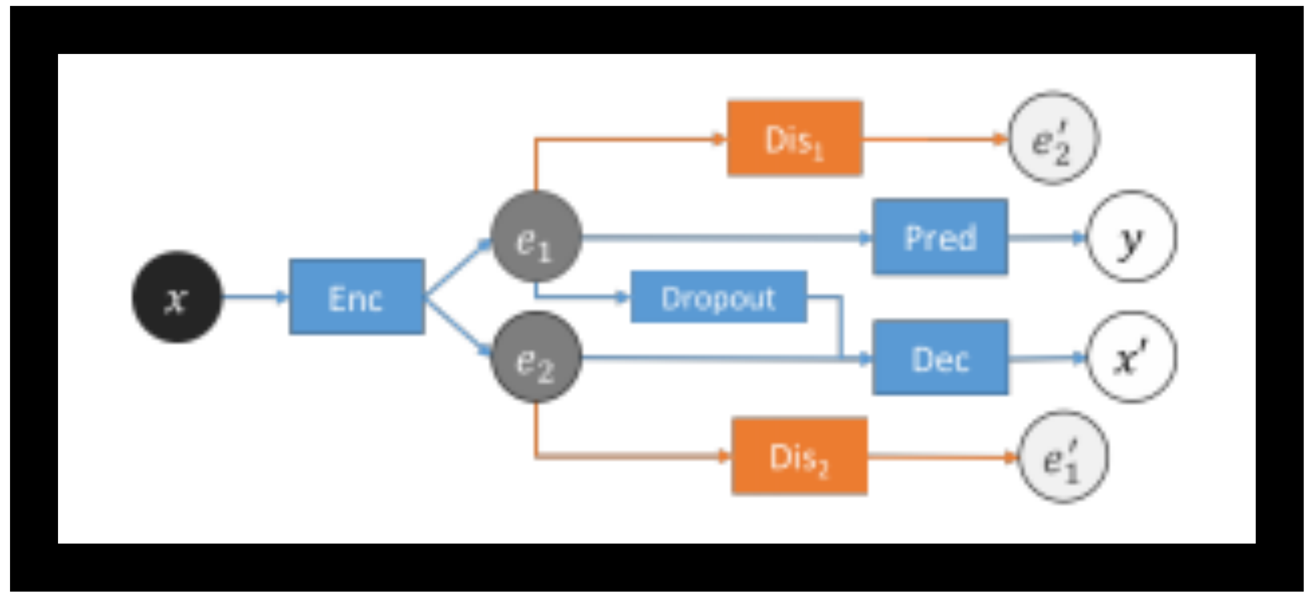
Vikas Singh

WISCONSIN

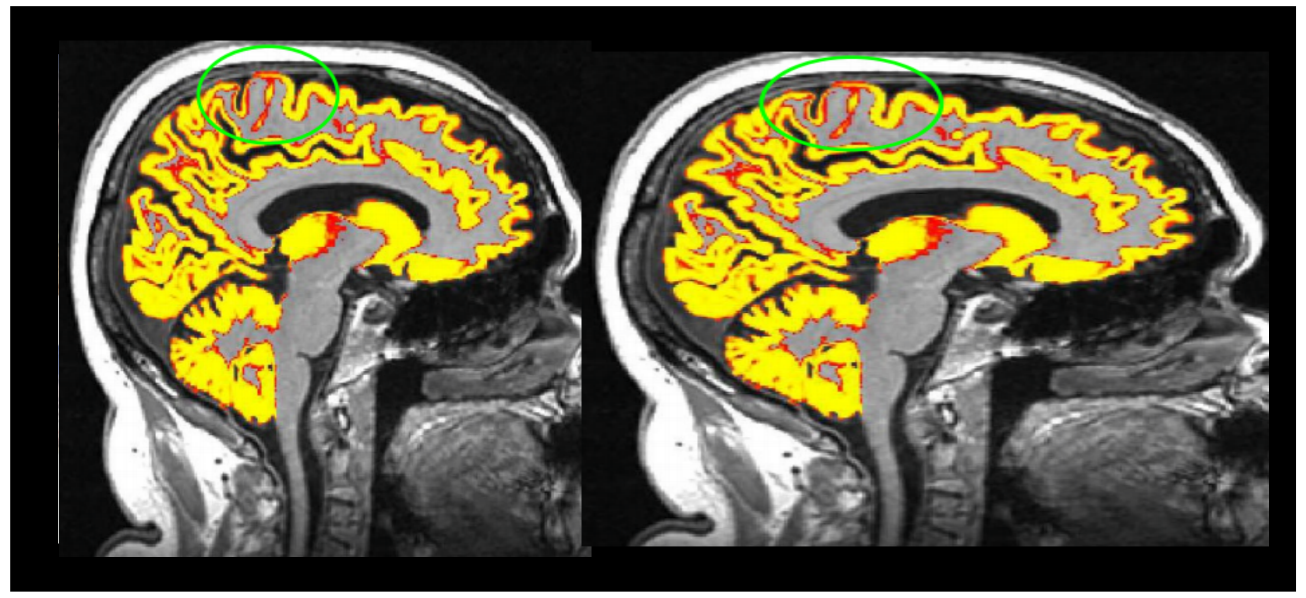
UNIVERSITY OF WISCONSIN–MADISON

Common Approaches

→ **Adversarial Modules**



→ **Augmentation**

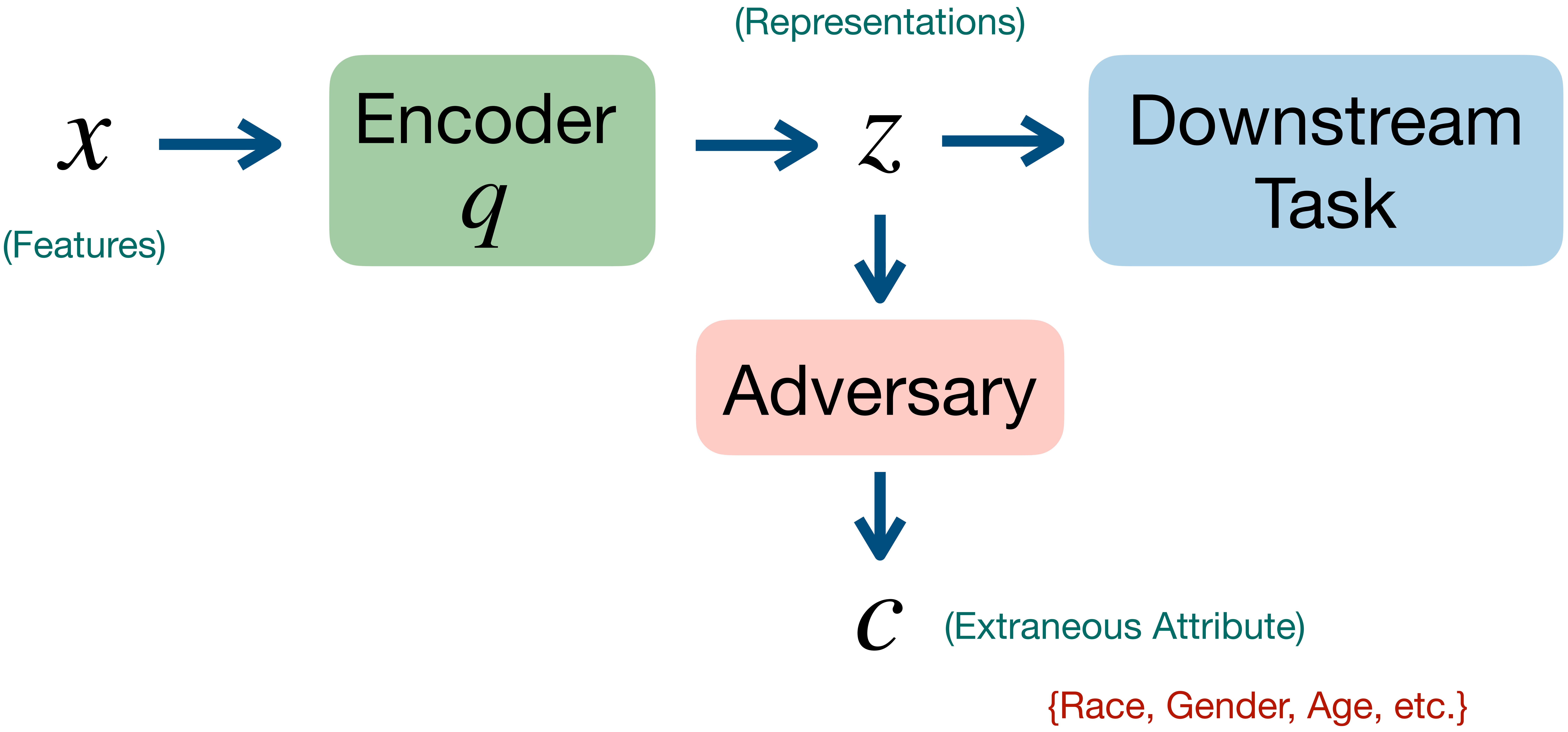


→ **Statistical Independence**

$$z \perp c$$

- **Motivation**
- Key Intuition
    - ICL Formulation
    - Benefits
- Applications
- Contributions

Basic Setting



Modeling Independence $z \perp c$

$$I(z, c) = KL[p(z, c) \,||\, p(z)p(c)]$$

$$I(z, c) \leq \mathbb{E}\Big[KL[q(z|x) \,||\, q(z)]\Big] - \mathbb{E}\Big[\log(p(x|z, c))\Big]$$

Compression **Reconstruction**

Mutual Information Approximation

Moyer, D.; Gao, S.; Brekelmans, R.; Galstyan, A.; and Ver Steeg, G. 2018. Invariant representations without adversarial training. In Advances in Neural Information Processing Systems, 9084–9093

Compression Regularizer in Practice

$$\lambda \quad \mathbb{E}\left[KL[q(z|x)\,||\,q(z)]\right]$$

Gaussian Approximation

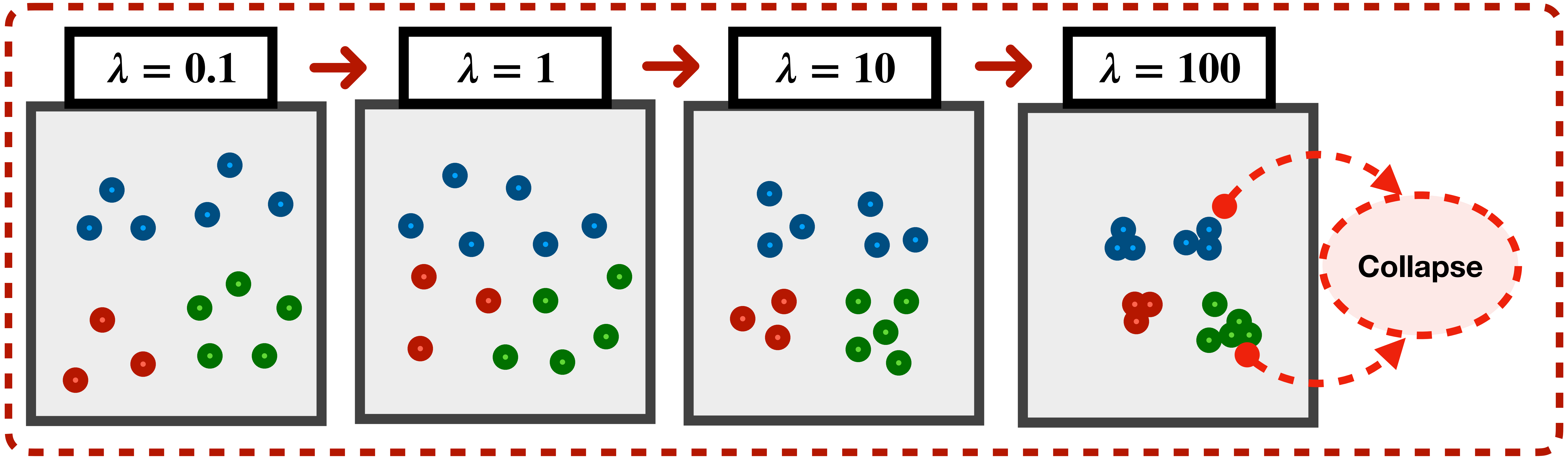
Weight

Compression

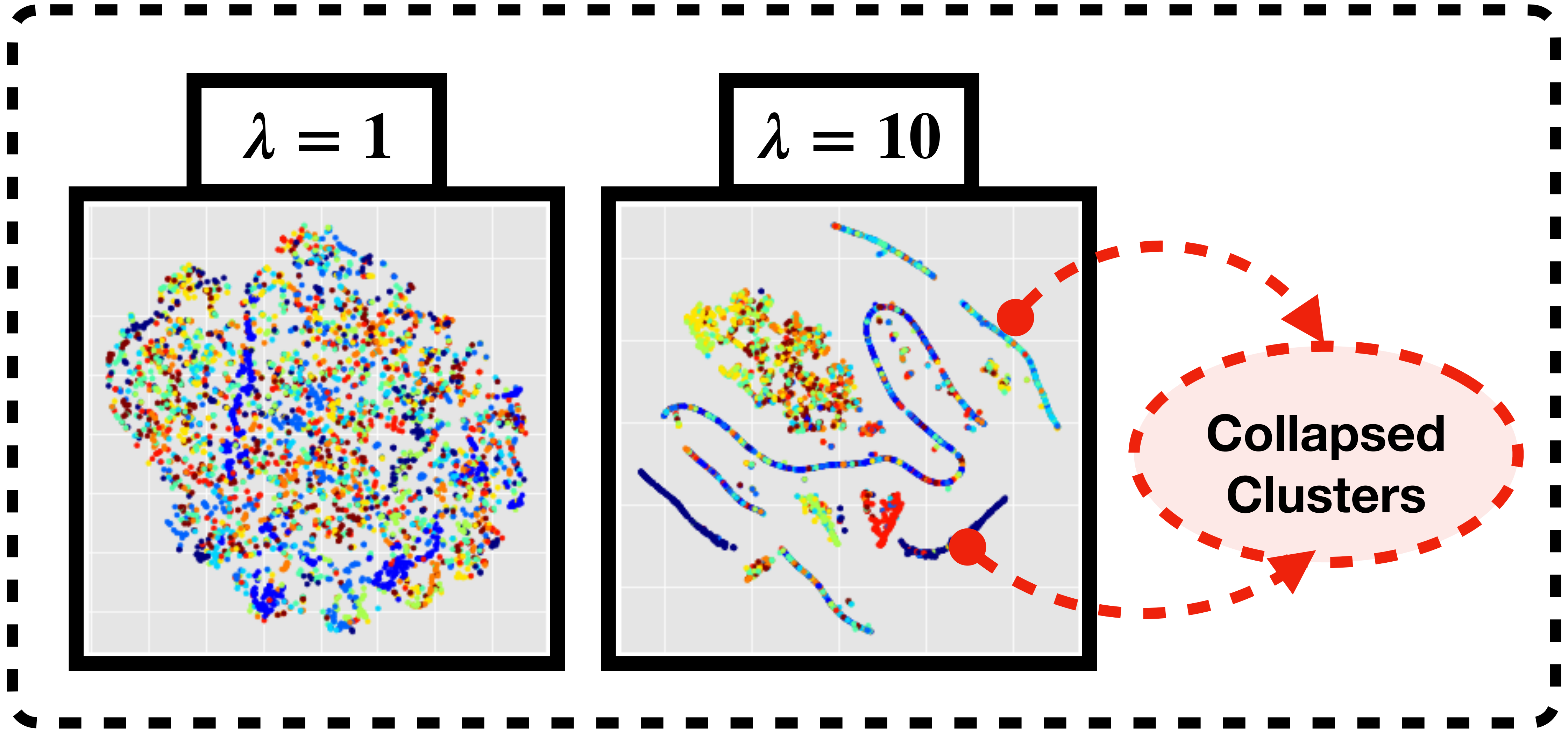
Computationally tractable approximation

$$\approx \sum_{i,j} \mathbb{E}\left[KL[q(z|x_i)\,||\,q(z|x_j)]\right]$$

Example Illustration



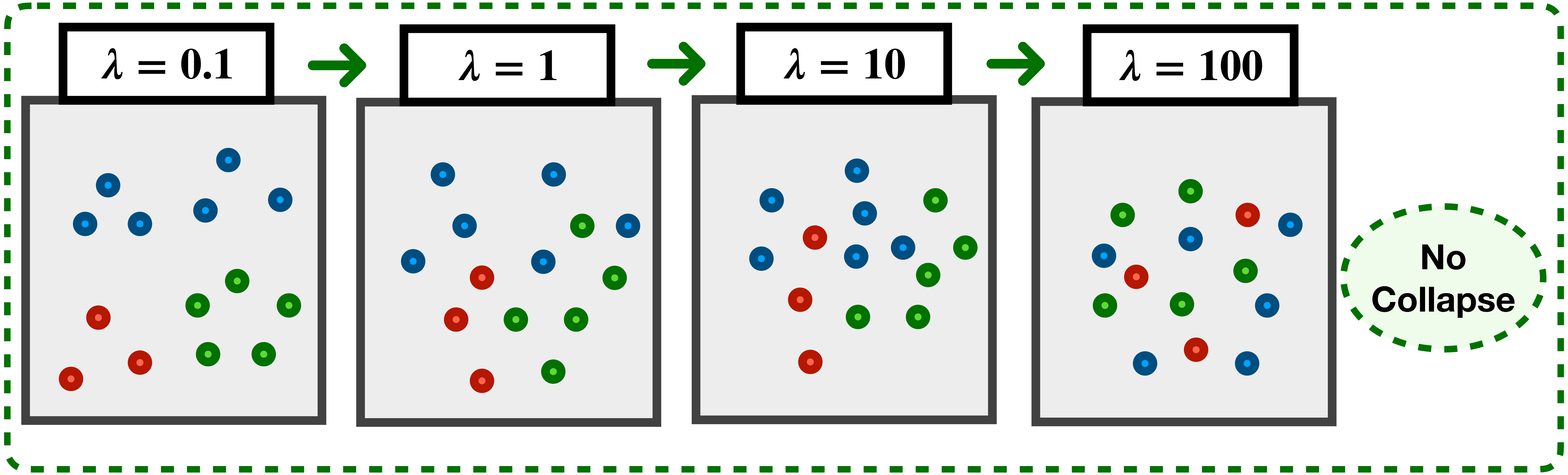
OBSERVED



Existing regularizers do not prevent clustering

- Motivation
- **Key Intuition**
    - ICL Formulation
    - Benefits
- Applications
- Contributions

Intuition



DESIRED

- *Intermix* representations for *different* c's
- *Spread out* representations for *same* c's

- Motivation
- Key Intuition
    - ICL Formulation
    - Benefits
- Applications
- Contributions

Contrastive Loss

Contrastive Loss

High intraclass similarity $z^T z^+$

Low interclass similarity $z^T z^-$

Inverting Contrastive Loss

1. Switch roles of $z^T z^+$ and $z^T z^-$

High interclass similarity

Low intraclass similarity

2a. Apply increasing function (quadratic) on $z^T z^-$

2b. Apply decreasing function (exponential loss) on $z^T z^+$

Definition

Similar C

Dissimilar C

$$\mathrm{ICL}(z,c) = \mathop{\mathbb{E}}_{\substack{(z,c)\sim p(z,c)\\(z',c')\sim p(z',c')}}\left[\mathbf{1}(c' \in \mathcal{N}_\delta(c))\, f(z,z') + \mathbf{1}(c' \notin \mathcal{N}_\delta(c))\, s(z,z')\right]$$

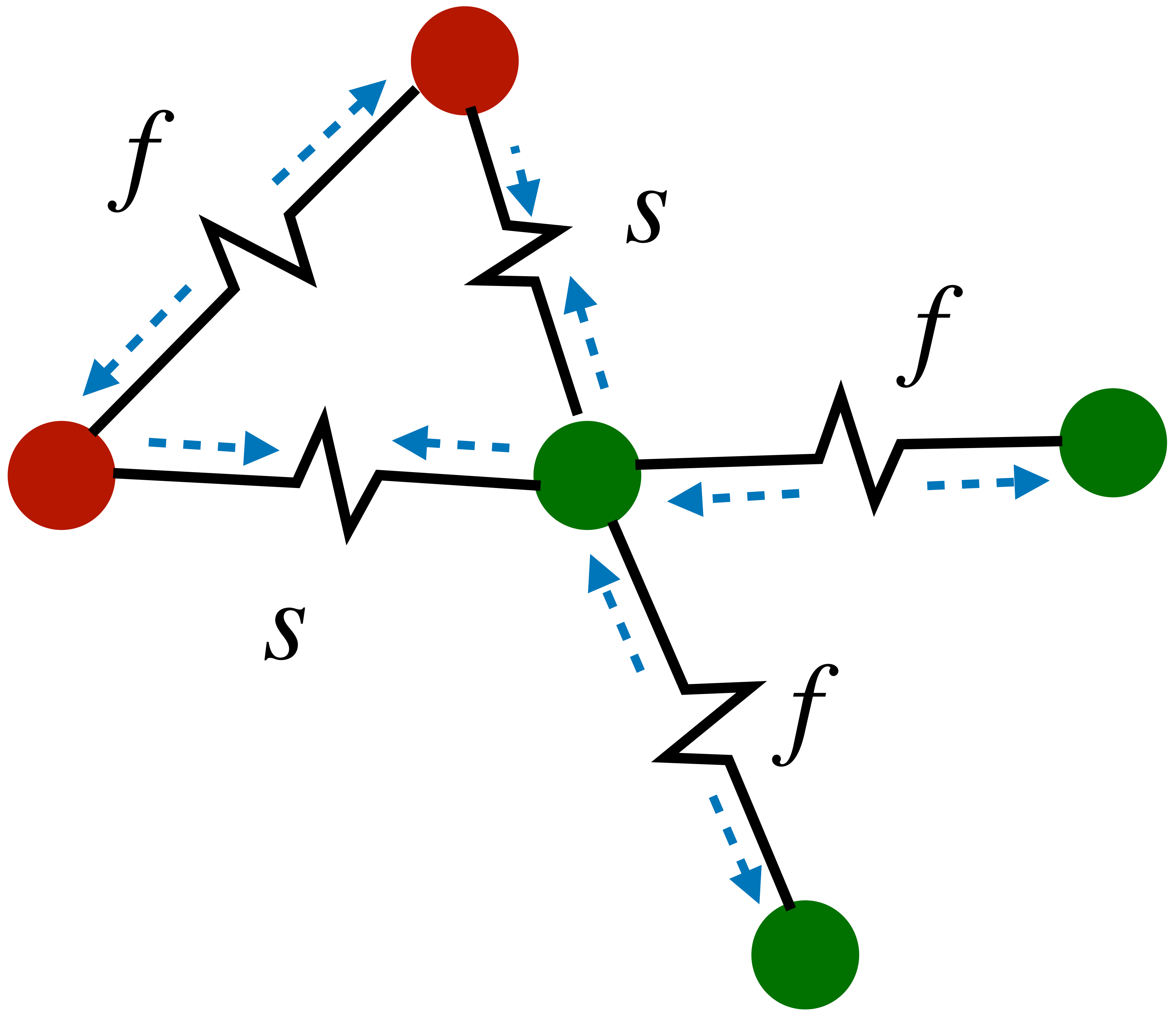
$$f(z,z') = \exp(\alpha - \beta d_Z(z,z'))$$

Decreasing

$$s(z,z') = d_Z^2(z,z')$$

Increasing

Dynamical System Intuition



f : Push spring

s : Pull spring

Optimizing ICL = System Equilibrium

- Motivation
- Key Intuition
  - ICL Formulation
  - **Benefits**
- Applications
- Contributions

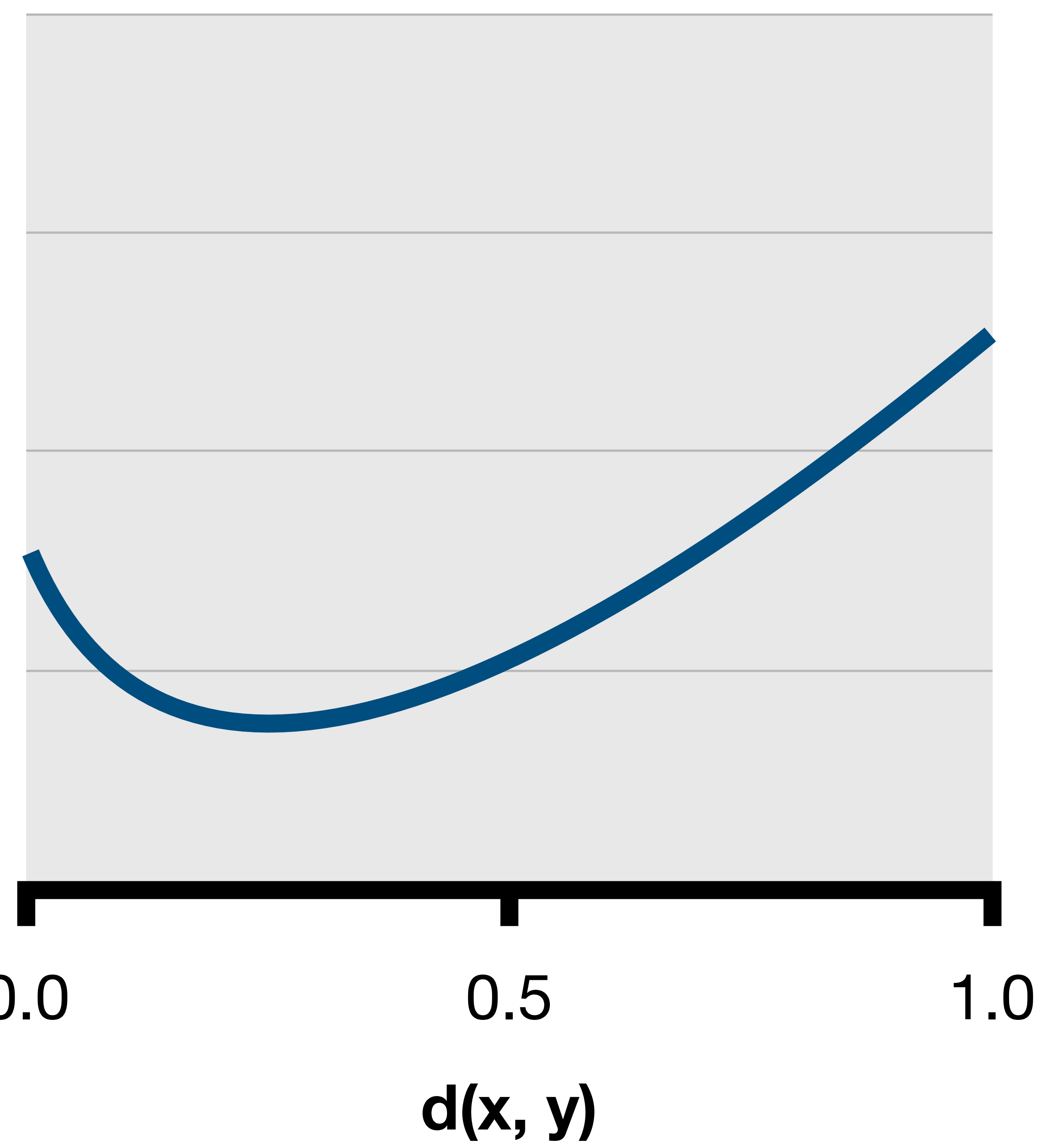
ICL for Discrete Extraneous c

Lemma 1 :

For $c \in \{0,1\}, \ p(c=0) = 1/2$

$$\mathrm{ICL}(Z, C) = \mathrm{MMD}_g(p_0, p_1) + R_w(p_0, p_1)$$

Interaction Potential



$d(x, y) = 0$

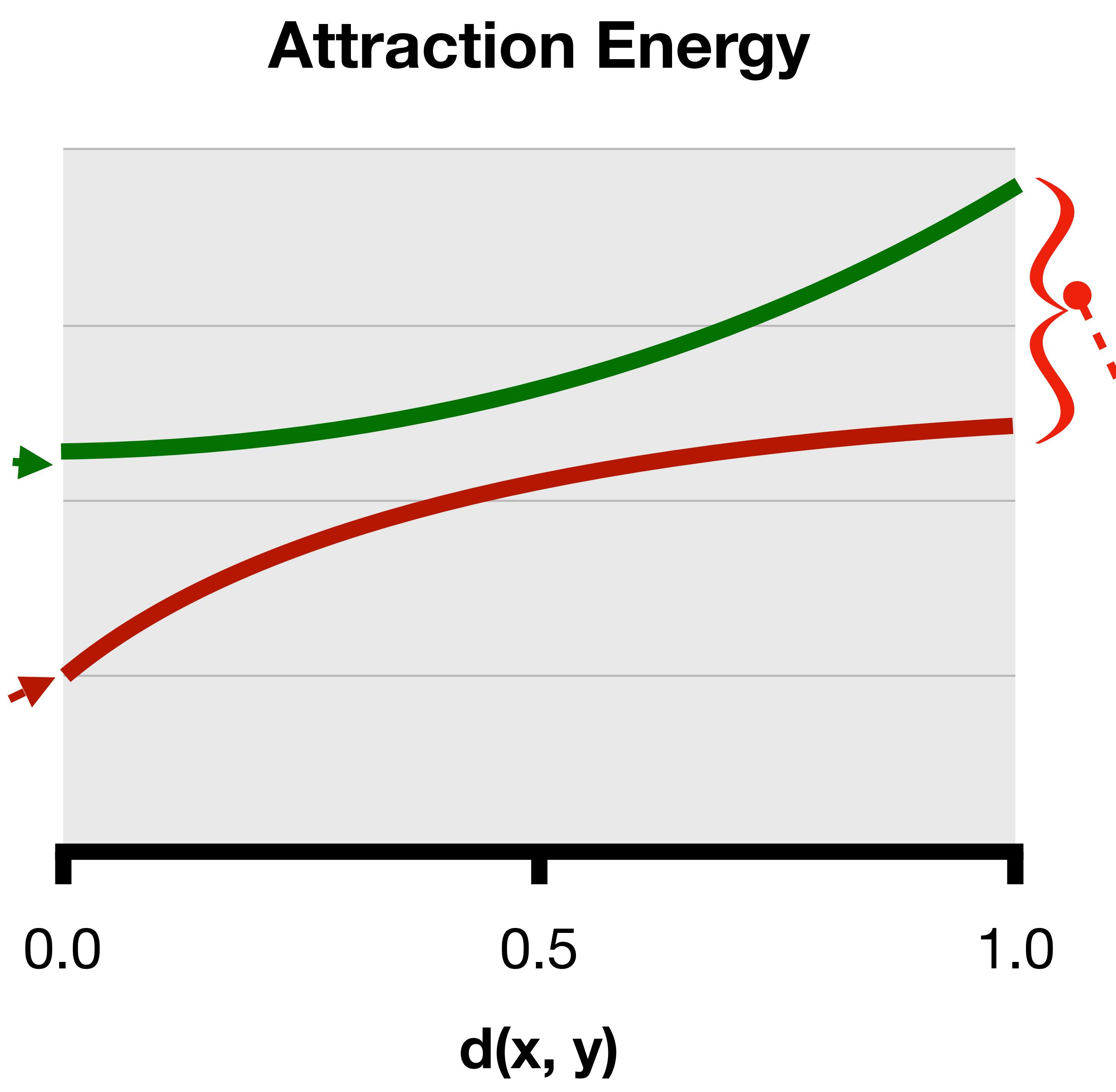
is suboptimal

Suited for First Order Methods

$$\mathrm{ICL}(z, c) = \mathop{\mathbb{E}}_{\substack{x \sim p_0 \\ x' \sim p_0}} \frac{f(x, x')}{4} + \mathop{\mathbb{E}}_{\substack{y \sim p_1 \\ y' \sim p_1}} \frac{f(y, y')}{4} + \mathop{\mathbb{E}}_{\substack{x \sim p_0 \\ y \sim p_1}} \frac{s(x, y)}{2}$$

$$\mathrm{MMD}_f = \mathop{\mathbb{E}}_{\substack{x \sim p_0 \\ x' \sim p_0}} f(x, x') + \mathop{\mathbb{E}}_{\substack{y \sim p_1 \\ y' \sim p_1}} f(y, y') - 2 \mathop{\mathbb{E}}_{\substack{x \sim p_0 \\ y \sim p_1}} f(x, y)$$

Repulsion **Attraction**



Attraction for

$ICL > MMD_f$

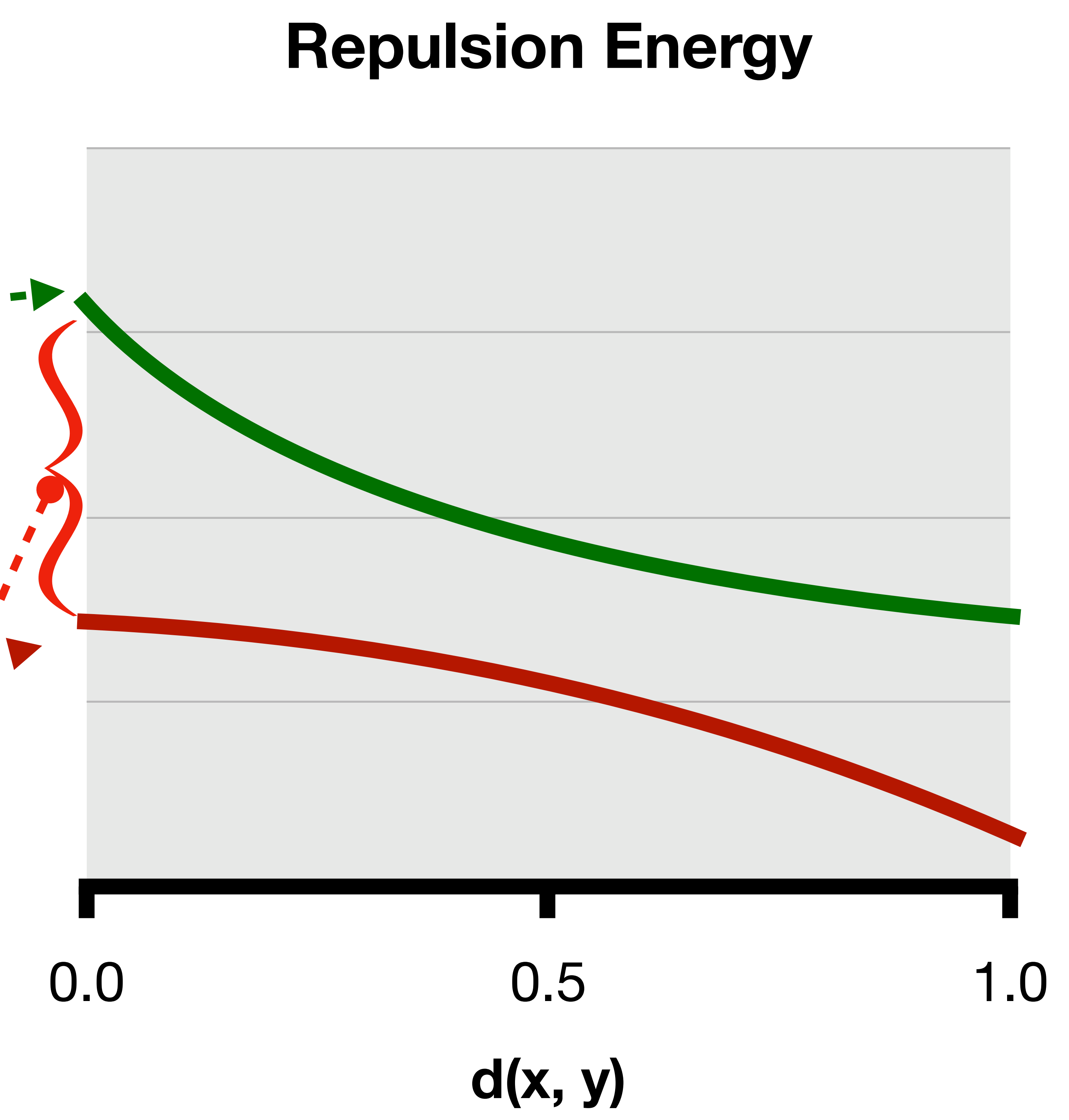
for large d(x,y)

Prevents Collapse

$$\text{ICL}(z,c) = \mathop{\mathbb{E}}_{\substack{x \sim p_0 \\ x' \sim p_0}} \frac{f(x,x')}{4} + \mathop{\mathbb{E}}_{\substack{y \sim p_1 \\ y' \sim p_1}} \frac{f(y,y')}{4} + \mathop{\mathbb{E}}_{\substack{x \sim p_0 \\ y \sim p_1}} \frac{s(x,y)}{2}$$

$$\text{MMD}_{-s} = \mathop{\mathbb{E}}_{\substack{x \sim p_0 \\ x' \sim p_0}} - s(x,x') + \mathop{\mathbb{E}}_{\substack{y \sim p_1 \\ y' \sim p_1}} - s(y,y') + 2 \mathop{\mathbb{E}}_{\substack{x \sim p_0 \\ y \sim p_1}} s(x,y)$$

Repulsion



Repulsion for

$ICL > MMD_{-s}$

for small d(x,y)

Provides Adversarially Invariance

Lemma 2 :

For c continuous, Adversary b L – Lipschitz, $\rho = P_{c,c'}(|c - c'| > \delta)$,

$\exists \alpha$, $\epsilon < \delta^2 \rho^2 / L^2$ such that for $\mathrm{ICL}(z, c) < \epsilon$

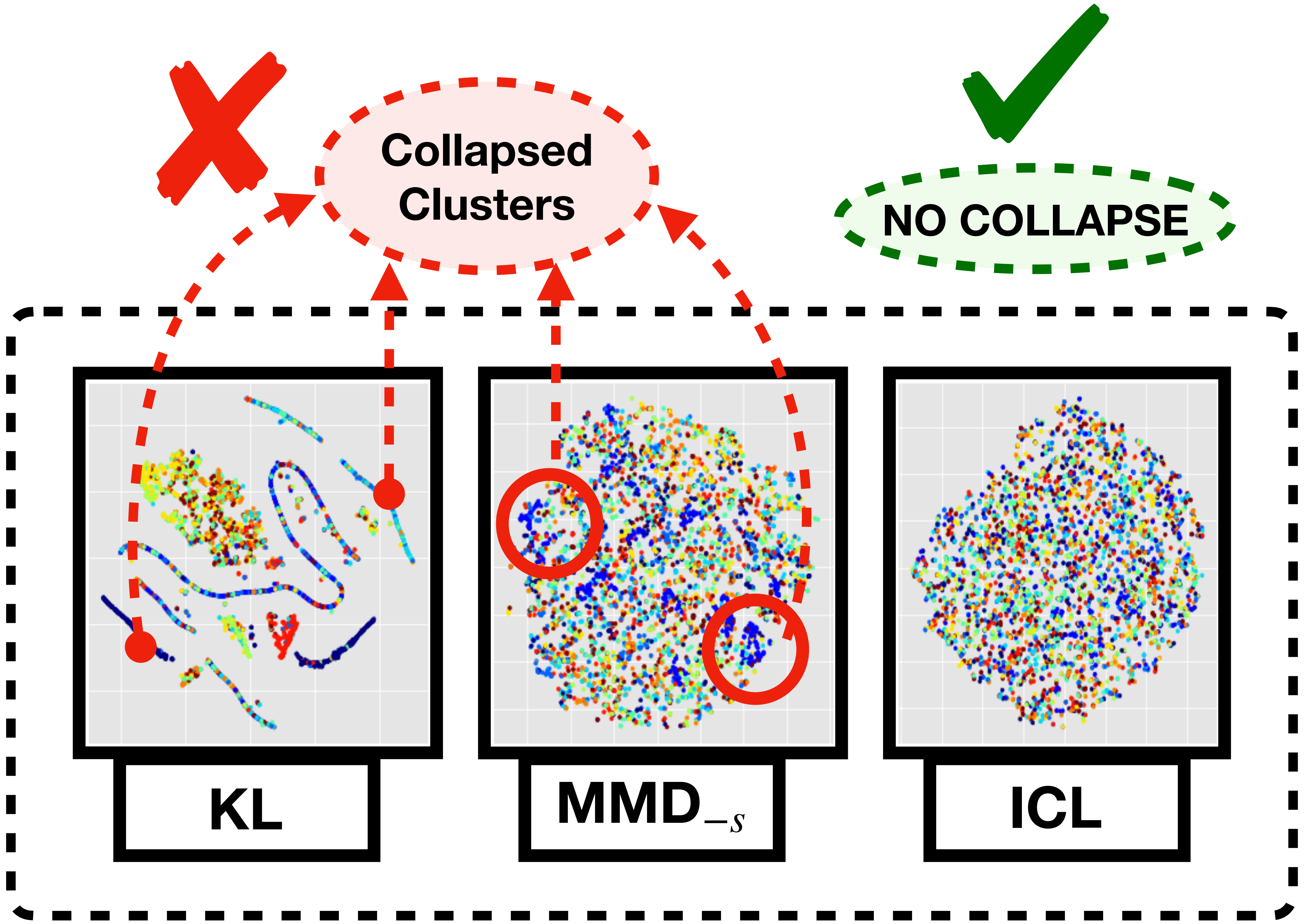
$$\mathbb{E}_z[(b(z) - c)^2] \geq (\delta\rho - L\sqrt{\epsilon})^2 / 4$$

- Motivation
- Key Intuition
  - ICL Formulation
  - Benefits
- **Applications / Experiments**
- Contributions

Generative Unsupervised Setting

$$\min_{p,q} \mathbb{E}_{x,c}\left[\mathbb{E}_{z}[-\log p(x\,|\,z,c)] + \beta\,\mathrm{KL}[q(z\,|\,x)\,||\,p(z)]\right] + \lambda\,\mathrm{ICL}(z,c)$$

Learning Style Information in MNIST



Generative Supervised Setting

VIB $x \rightarrow \boxed{q(z|x)} \rightarrow z \rightarrow \boxed{p(y|z)} \rightarrow y$

$$\min_{p,q} - \mathop{\mathbb{E}}_{x,c} \left[\mathop{\mathbb{E}}_{z,y} \log p(y|z) + \beta \mathop{\mathbb{E}}_{z} \log p(x|z,c) \right] + \lambda \, \mathrm{ICL}(z,c)$$

Learn Invariant Representations For Fairness Datasets

a) Adult

b) German

Discriminative Model Family

$$\min_{f,h} \mathbb{E}_{x,y} [\ell(f(h(x), y))] + \lambda \, \mathrm{ICL}(h(x), c)$$

Predictor **Encoder**

- Rotation Invariance - MNIST-Rot
- Invariance wrt Continuous Attribute - Adult (Age)
- Controlling Scanner Confounds - ADNI Dataset

Results

	MNIST		Adult		German		MNIST-Rot		ADNI	
	R ↓	A ↓	P ↑	A ↓	P ↑	A ↓	P ↑	A ↓	P ↑	A ↓
Unregularized	12.1	46	84	84	73	78	96	42	83	55
MI	13.2	50	84	78	70	76	96	38	-	-
MMD_{-s}	15.8	55	84	82	73	75	96	35	85	49
MMD_f	15.8	50	83	80	74	78	96	34	86	57
OT	14.4	61	83	78	72	75	-	-	-	-
CAI	11.8	48	84	81	73	75	96	38	85	51
UAI	-	-	84	83	73	75	98	34	84	49
ICL (Ours)	16.6	**32**	83	**75**	75	**75**	96	**33**	84	**46**

P: Prediction Accuracy, **R**: Reconstruction Error, **A**: Adversarial Invariance

Results - Continuous Case

	Adult (with Age)	
	P ↑	A ↑
Unregularized	83	112
CAI	82	129
UAI	84	114
ICL (Ours)	83	**161**

P: Prediction Accuracy, **A**: Adversarial Invariance MSE

- Motivation
- Key Intuition
    - ICL Formulation
    - Benefits
- Applications
- **Contributions**

Contributions

- Proposed ICL - computationally efficient
- Interpret as Regularized MMD
- Works for continuous setting

Thank You!

https://github.com/adityakumarakash/ICL