Rare Event Modeling

Tony Tran

UNIVERSITY OF HOUSTON

Introduction - Rare Event (Post-Event)

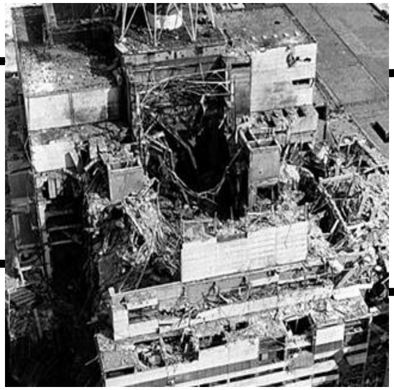
A rare event is an observation that occurs with a low probability under a given distribution

Rare events can be expensive

$700 billion USD

30 immediate deaths

Chernobyl 1986



Understand underlying cause

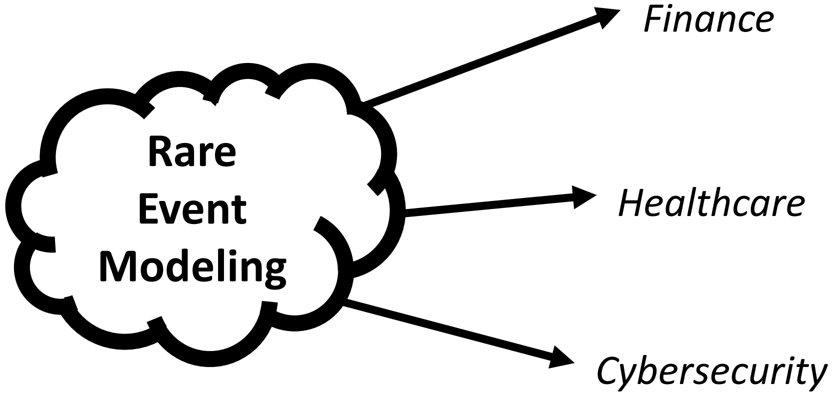
design flaw?

operator error?

Key in safety-critical systems

Collision Warning

Broad relevance across domains



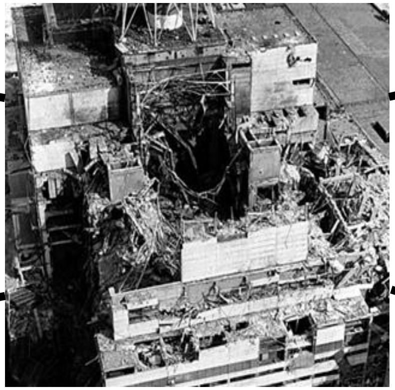
Rare Event Modeling - Posterior Learning

UNIVERSITY OF HOUSTON

Question: What can we learn given the rare event has occurred?

Design Flaw?

Test Mismanagement?



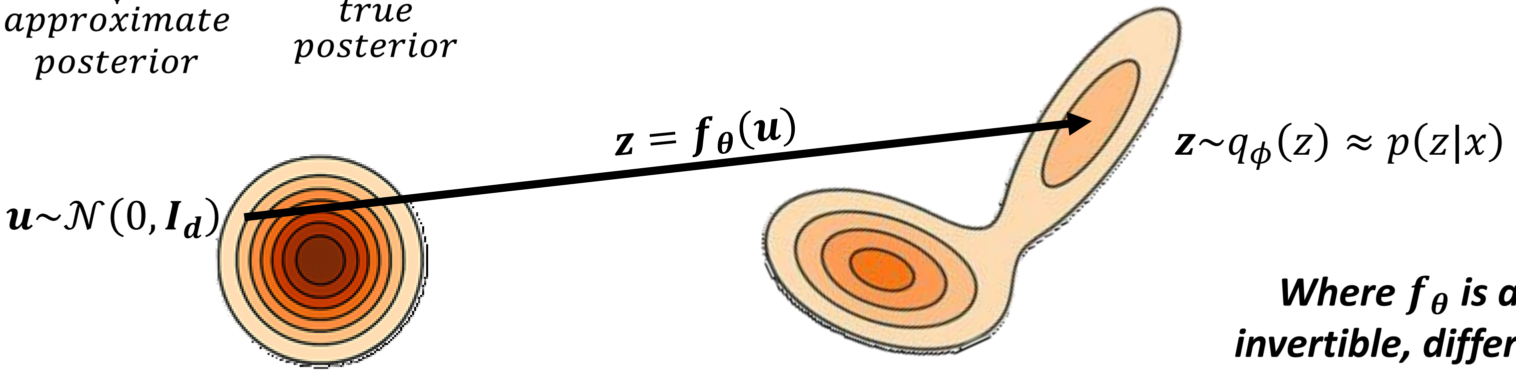
Operator Error?

Flawed Shutdown?

Task: Model Posterior Over Latent Variables $\boldsymbol{p(z|x)}$

$$\underbrace{q_\phi(z)}_{\substack{\text{approximate}\\ \text{posterior}}} \approx \underbrace{p(z|x)}_{\substack{\text{true}\\ \text{posterior}}} \quad \text{with } q_\phi(z) \text{ modeled via normalizing flow}$$

Where z is the latent variable representing hidden causes and x is the observed rare event



Simple Distribution
$\mathcal{N}(0, \boldsymbol{I}_d)$

Estimated Distribution
$q_\phi(z) \approx p(z|x)$

Where $\boldsymbol{f}_{\boldsymbol{\theta}}$ is a composition of invertible, differentiable functions estimated by a neural network

Rare Event Modeling - Variational Inference

UNIVERSITY OF HOUSTON

Inverse Bayesian Problem

Posterior is intractable

$$\underbrace{p(z|\mathcal{D})}_{ground\ truth} = \frac{p(\mathcal{D}|z)p(z)}{p(\mathcal{D})}$$

Because denominator term is intractable

Where $\mathcal{D}$ is the dataset

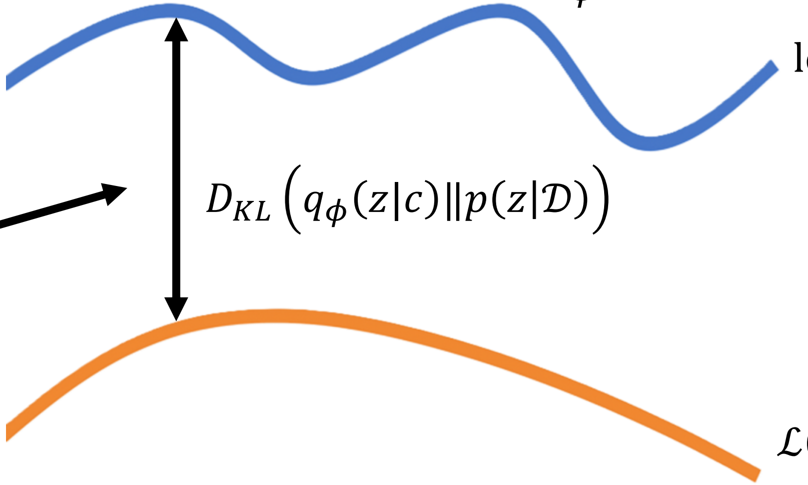
Use Variational Inference!

Conditional Evidence Lower BOund (ELBO) Objective

$$\mathcal{L}(\phi, c, \mathcal{D}) = \mathbb{E}_{(x,y)\in\mathcal{D}, z\sim q_\phi(z|c)}\left[\log p(x, z|y) - \log q_\phi(z|c)\right]$$

$$\phi^* = \arg\max_\phi \mathcal{L}(\phi, c, \mathcal{D})$$

c is the conditional (guidance) vector



$\log p(x|y)$

Maximum Likelihood Estimation maximizes likelihood or "evidence" directly

Posterior Learning aims to minimize KL divergence between learned and true posterior

$\mathcal{L}(\phi, c, \mathcal{D})$

Variational Inference maximizes "ELBO"

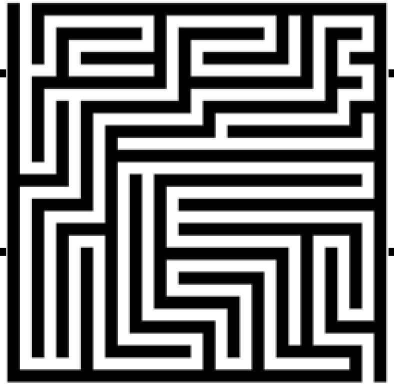
Challenges - Data Constraint

Collection of data is challenging and expensive in some domains

Data Scarcity

Imbalanced Dataset

Environment



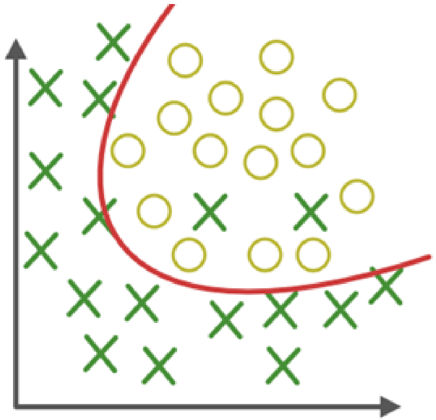
< 0.1% *chance of occurrence*

few rare event samples

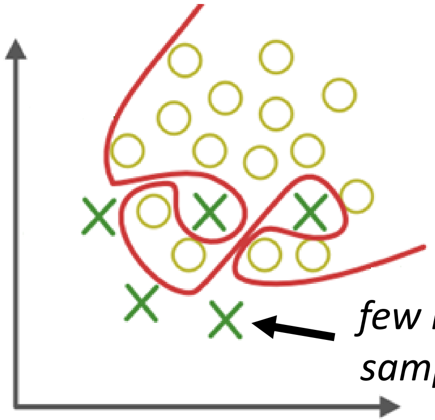
> 99.9% *chance of occurrence*

abundant nominal samples

Overfitting



Good Fit



few rare samples

Over Fit

Distribution Shift

Simulated

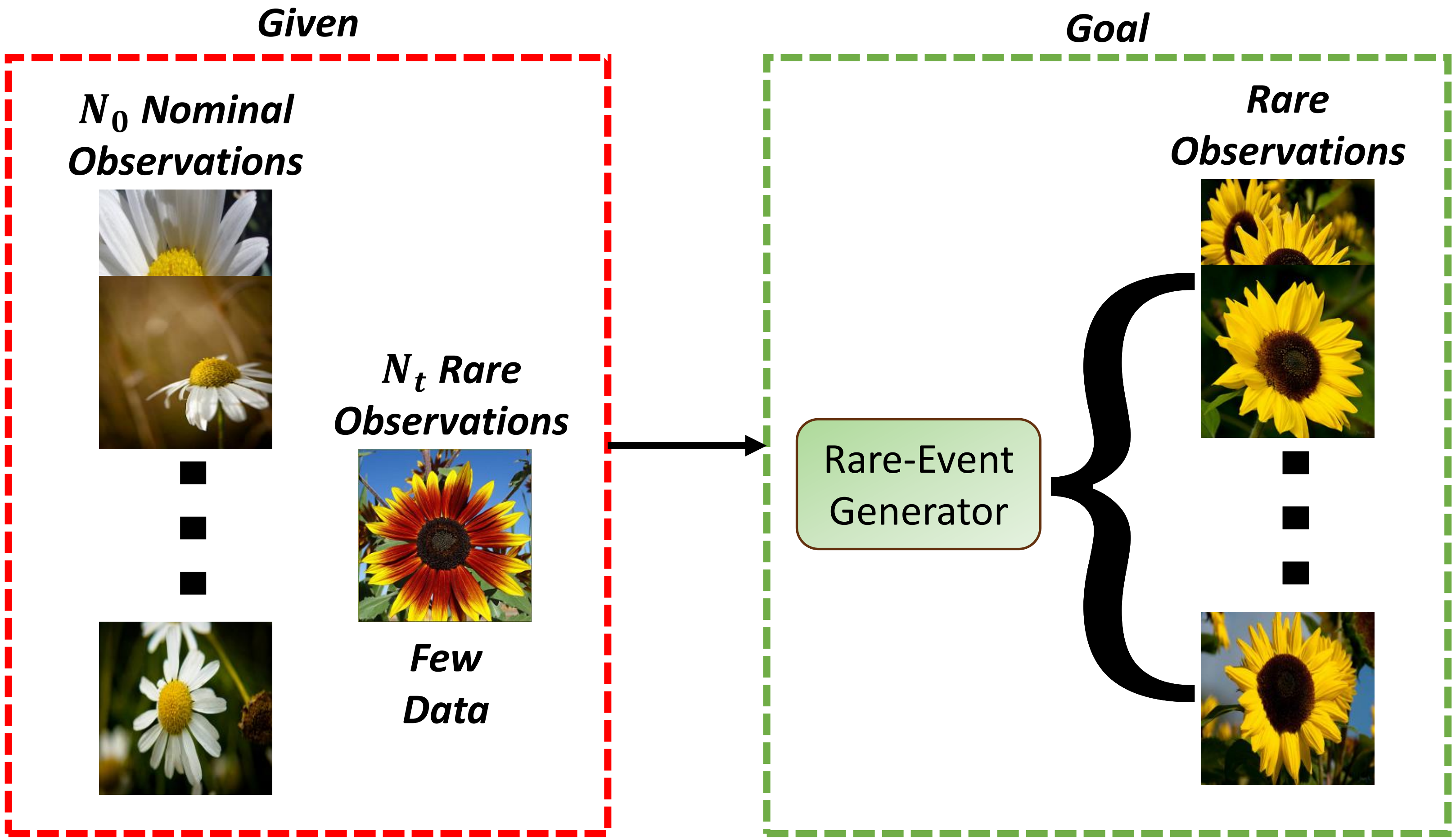
Domain Shift

Real

few rare samples

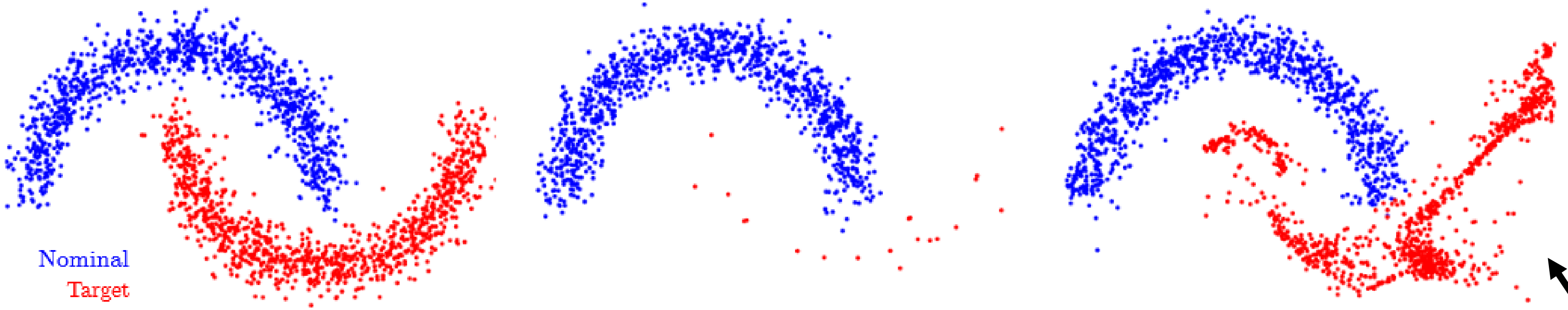
Conditional Generative Modeling Problem

$$N_t \ll N_0$$



Assumption: Nominal and rare observations are from the same domain

Rare Event Modeling - Overfitting



True Distribution

Few Rare Events

Overfit Distribution

Few Data Causes Overfitting

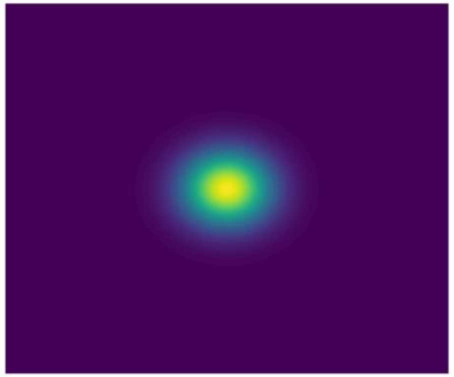
Regularizes

Solution:

1. Train a single flow model that represents both nominal and target distributions, guided by a guidance vector

2. Regularize the target distribution using the nominal distribution

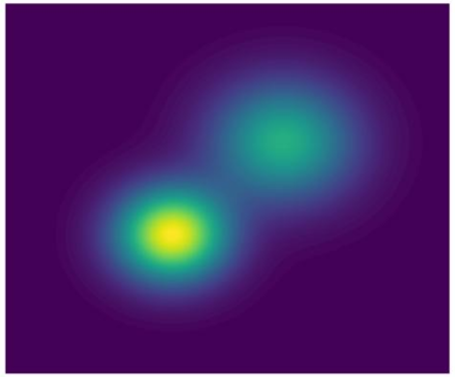
$c = [0, 0, 0, 0, 0]$



Nominal Distribution q_{ϕ_0}

Guidance Vector c

$c = [1, 1, 1, 1, 1]$



Target Distribution q_{ϕ}

Rare Event Modeling - Prior Regularization

UNIVERSITY OF HOUSTON

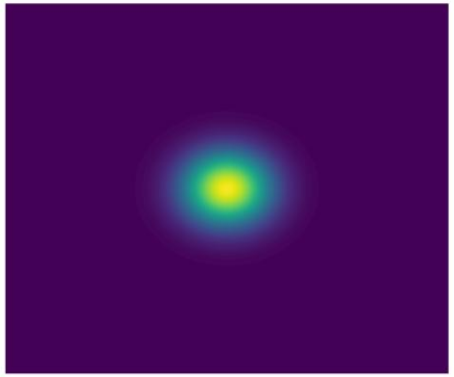
Method 1: KL Divergence

$$J(\phi) = \underbrace{\mathcal{L}(\phi, \mathbf{1}, \mathcal{D}_t)}_{target\ distribution} + \underbrace{\mathcal{L}(\phi, \mathbf{0}, \mathcal{D}_0)}_{nominal\ distribution} - \underbrace{\beta D_{KL}\left(q_{\phi_0}, q_\phi\right)}_{KL\ divergence\ penalty}$$

***β* difficult to tune**

Method 2: Wasserstein Divergence

$$J(\phi) = \underbrace{\mathcal{L}(\phi, \mathbf{1}, \mathcal{D}_t)}_{target\ distribution} + \underbrace{\mathcal{L}(\phi, \mathbf{0}, \mathcal{D}_0)}_{nominal\ distribution} - \underbrace{\beta W_2^2\left(q_{\phi_0}, q_\phi\right)}_{Wasserstein\ distance\ penalty}$$

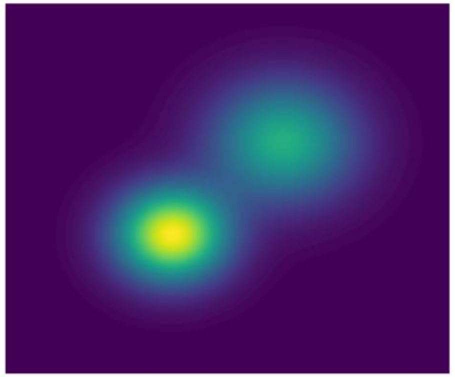


Nominal Distribution q_{ϕ_0}

Measures difference between nominal and target distribution

$$D_{KL}\left(q_{\phi_0}, q_\phi\right)$$

$$W_2^2\left(q_{\phi_0}, q_\phi\right)$$



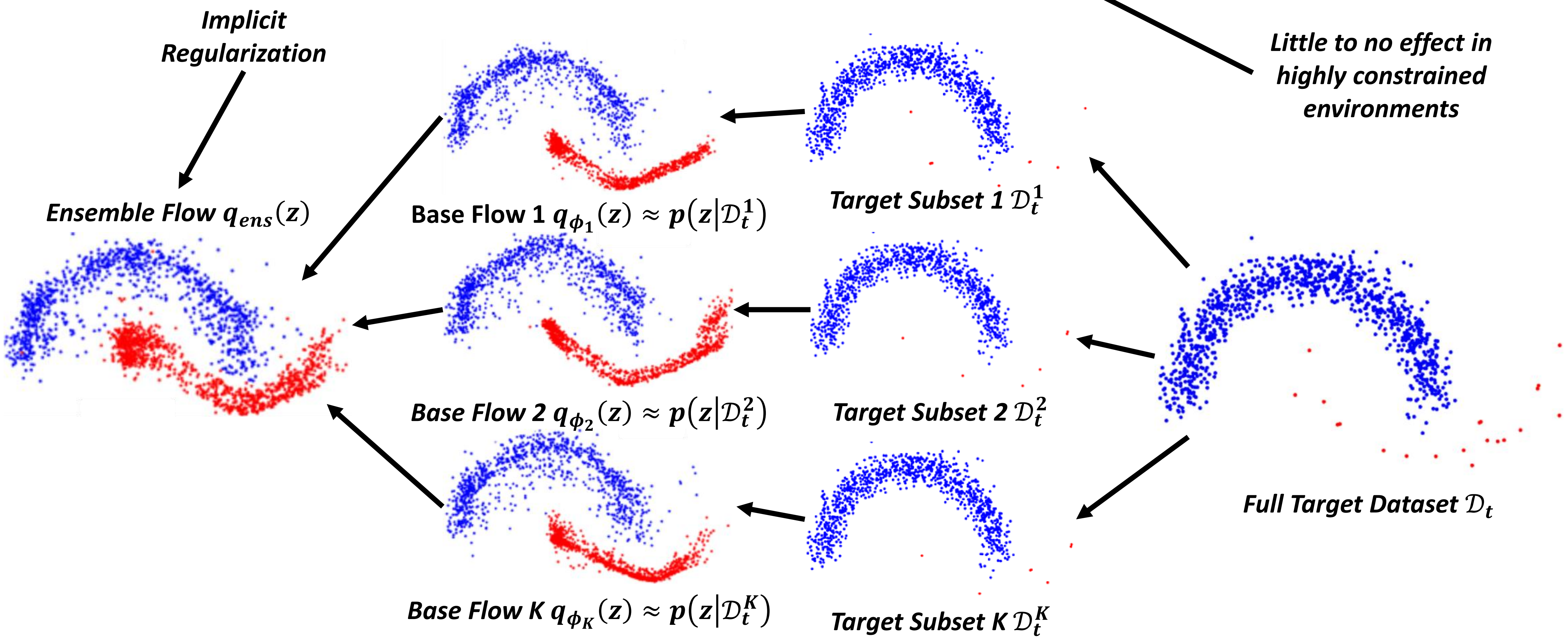
Target Distribution q_ϕ

Rare Event Modeling - Bootstrapping

UNIVERSITY OF HOUSTON

Method 3: Ensemble Method

$$\underbrace{q_{ens}(z)}_{ensemble\ flow} = \frac{1}{K}\sum_{i=1}^{K} \underbrace{q_{\phi_i}(z)}_{base\ flows}$$



Rare Event Modeling - Self Regularization

Question:

1. How to adaptively choose regularization strength β?

$$J(\phi) = \underbrace{\mathcal{L}(\phi, \mathbf{1}, \mathcal{D}_t)}_{target\ distribution} + \underbrace{\mathcal{L}(\phi, \mathbf{0}, \mathcal{D}_0)}_{nominal\ distribution} - \underbrace{\beta D_{KL}\left(q_{\phi_0}, q_\phi\right)}_{KL\ divergence\ penalty}$$

$$J(\phi) = \underbrace{\mathcal{L}(\phi, \mathbf{1}, \mathcal{D}_t)}_{target\ distribution} + \underbrace{\mathcal{L}(\phi, \mathbf{0}, \mathcal{D}_0)}_{nominal\ distribution} - \beta W_2^2 \underbrace{\left(q_{\phi_0}, q_\phi\right)}_{Wasserstein\ distance\ penalty}$$

2. How to share information between flows to learn robustly?

$$\underbrace{q_{ens}(z)}_{ensemble\ flow} = \frac{1}{K}\sum_{i=1}^{K} \underbrace{q_{\phi_i}(z)}_{base\ flows}$$

Method 4: Self-Regularization

$$\mathrm{J}(\phi, c) = \frac{1}{K}\sum_{i=1}^{K} \underbrace{\mathcal{L}(\phi, \mathbf{1}_i, \mathcal{D}_t^i)}_{base\ flows} + \underbrace{\mathcal{L}(\phi, \mathbf{0}_K, \mathcal{D}_0)}_{nominal\ flow} + \underbrace{\mathcal{L}(\phi, c, \mathcal{D}_t)}_{ensemble\ flow} - \beta \sum_{i \neq j}^{K} \underbrace{D_{KL}\left(q_\phi(\cdot; \mathbf{1}_i), q_\phi(\cdot; \mathbf{1}_j)\right)}_{penalty\ between\ base\ flows}$$

Rare Event Modeling - Nominal Flow

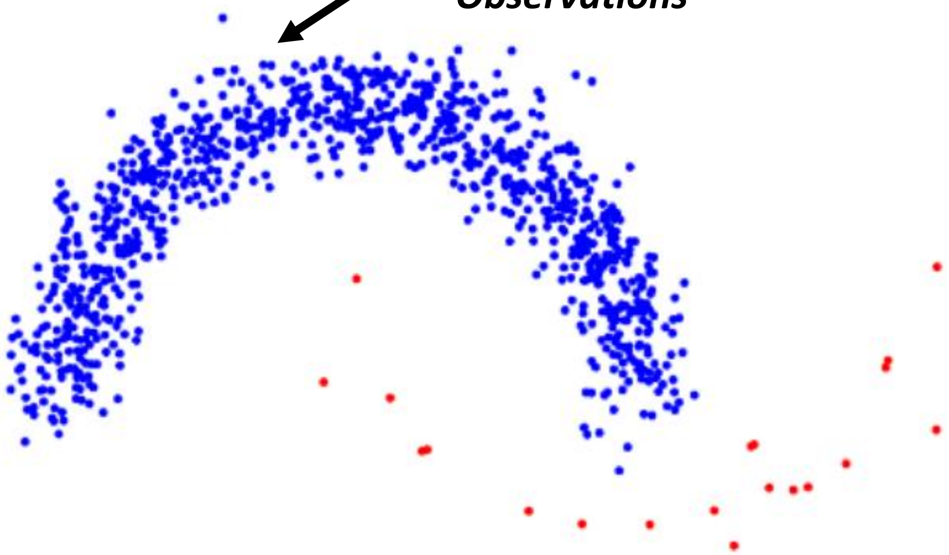
UNIVERSITY OF HOUSTON

$$J(\phi, c) = \frac{1}{K}\sum_{i=1}^{K} \underbrace{\mathcal{L}(\phi, \mathbf{1}_i, \mathcal{D}_t^i)}_{base\ flows} + \underbrace{\mathcal{L}(\phi, \mathbf{0}_K, \mathcal{D}_0)}_{nominal\ flow} + \underbrace{\mathcal{L}(\phi, c, \mathcal{D}_t)}_{ensemble\ flow} - \beta \sum_{i \neq j}^{K} \underbrace{D_{KL}\left(q_\phi(\cdot; \mathbf{1}_i), q_\phi(\cdot; \mathbf{1}_j)\right)}_{penalty\ between\ base\ flows}$$

$$[0, 0, 0, 0, 0]$$

Guidance vector

Nominal Observations



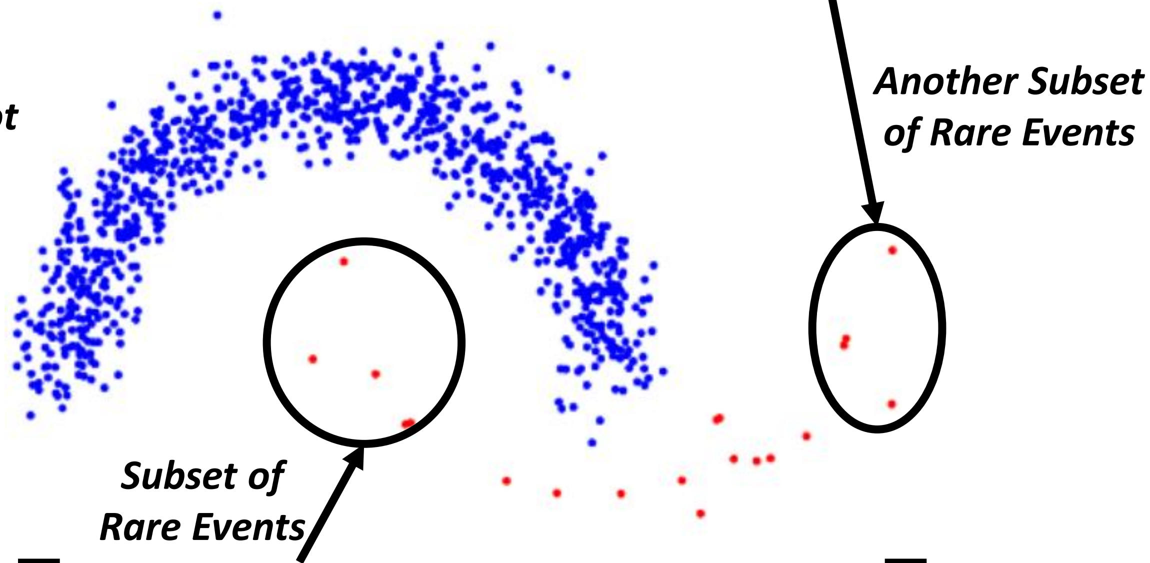
Rare Event Modeling - Base Flows

UNIVERSITY OF HOUSTON

$$\mathrm{J}(\phi, c) = \frac{1}{K}\sum_{i=1}^{K} \underbrace{\mathcal{L}(\phi, \mathbf{1}_i, \mathcal{D}_t^i)}_{base\ flows} + \underbrace{\mathcal{L}(\phi, \mathbf{0}_K, \mathcal{D}_0)}_{nominal\ flow} + \underbrace{\mathcal{L}(\phi, c, \mathcal{D}_t)}_{ensemble\ flow} - \beta \sum_{i \neq j}^{K} \underbrace{D_{KL}\left(q_\phi(\cdot; \mathbf{1}_i), q_\phi(\cdot; \mathbf{1}_j)\right)}_{penalty\ between\ base\ flows}$$

$$[0, 0, 0, 0, 1]$$

Guidance vector is one-hot encoded to represent unique target subset



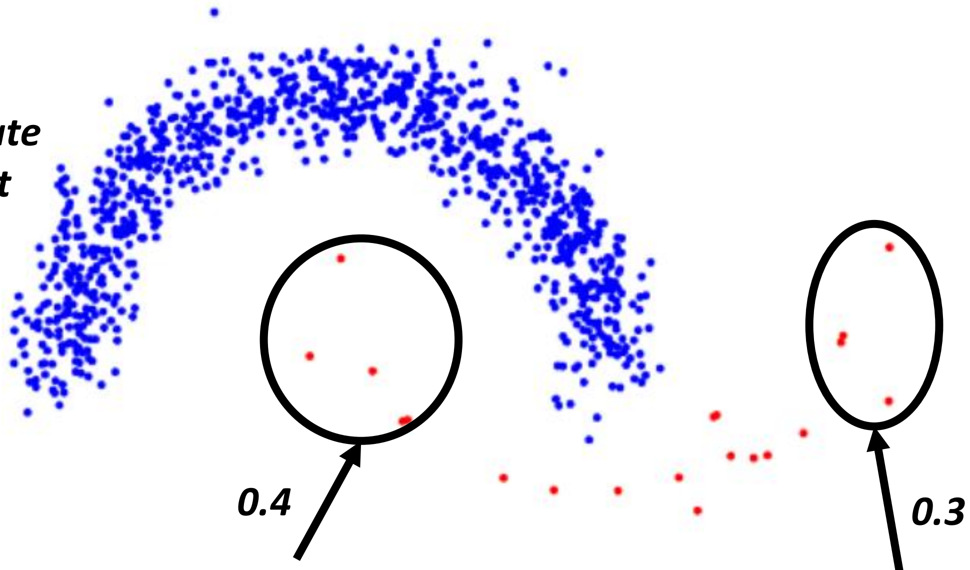
$$[0, 1, 0, 0, 0]$$

Rare Event Modeling - Ensemble Flow

UNIVERSITY OF HOUSTON

$$\mathrm{J}(\phi, c) = \frac{1}{K}\sum_{i=1}^{K}\underbrace{\mathcal{L}(\phi, \mathbf{1}_i, \mathcal{D}_t^i)}_{base\ flows} + \underbrace{\mathcal{L}(\phi, \mathbf{0}_K, \mathcal{D}_0)}_{nominal\ flow} + \underbrace{\mathcal{L}(\phi, c, \mathcal{D}_t)}_{ensemble\ flow} - \beta\sum_{i\neq j}^{K}\underbrace{D_{KL}\Big(q_\phi(\cdot;\mathbf{1}_i), q_\phi\big(\cdot;\mathbf{1}_j\big)\Big)}_{penalty\ between\ base\ flows}$$

Ensemble flow learns intermediate flows between weighted target subsets of target data

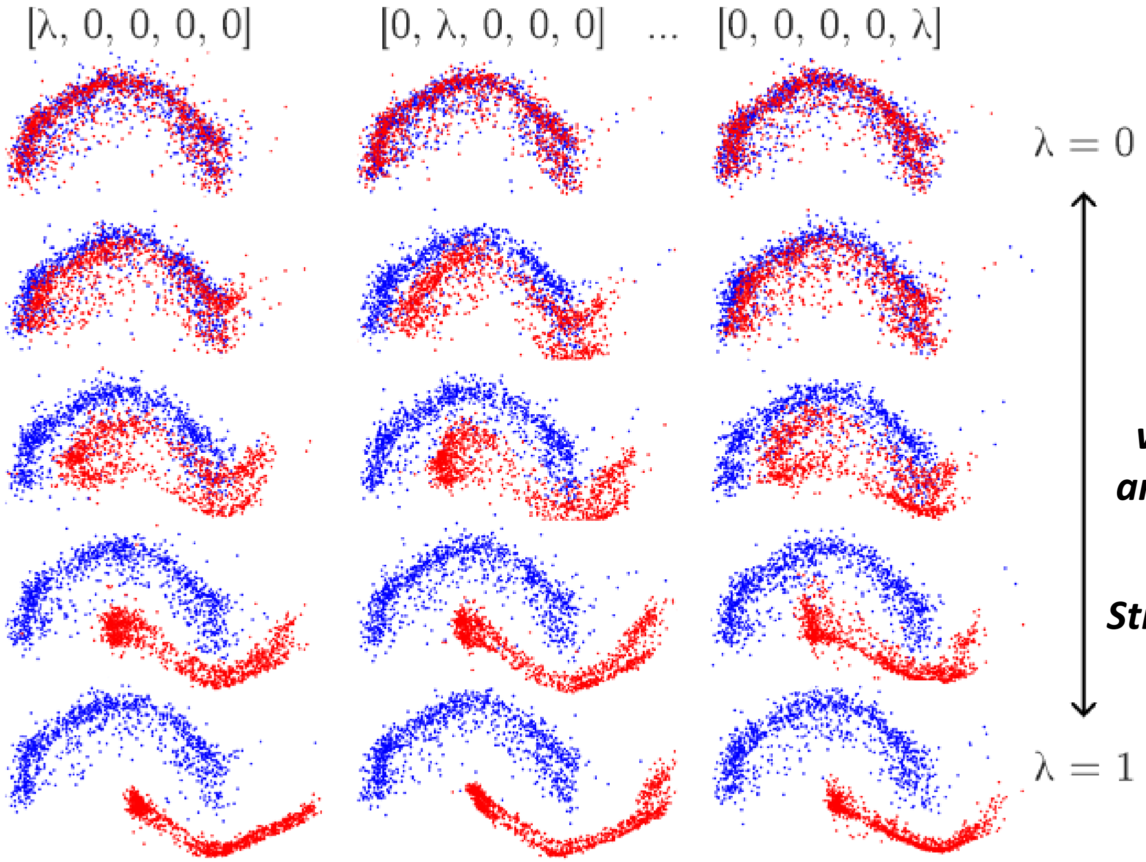


$$[0, 0.4, 0, 0, 0.3]$$

Rare Event Modeling - Final Flow Model

$$\mathrm{J}(\phi, c) = \frac{1}{K}\sum_{i=1}^{K} \underbrace{\mathcal{L}(\phi, \mathbf{1}_i, \mathcal{D}_t^i)}_{base\ flows} + \underbrace{\mathcal{L}(\phi, \mathbf{0}_K, \mathcal{D}_0)}_{nominal\ flow} + \underbrace{\mathcal{L}(\phi, c, \mathcal{D}_t)}_{ensemble\ flow} - \beta \sum_{i \neq j}^{K} \underbrace{D_{KL}\left(q_\phi(\cdot; \mathbf{1}_i), q_\phi(\cdot; \mathbf{1}_j)\right)}_{penalty\ between\ base\ flows}$$

Step 1: Yield a family of flows



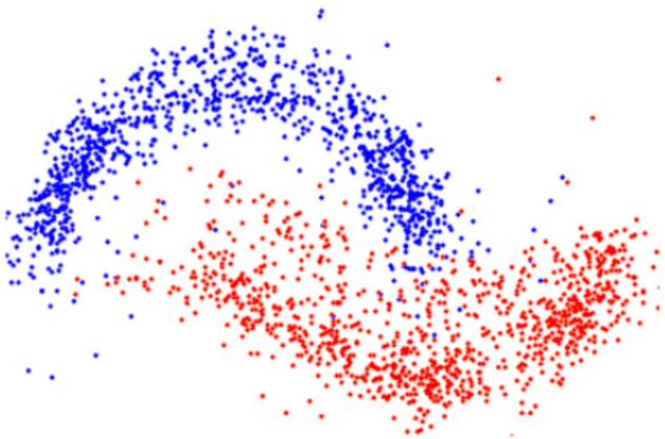
λ represents regularization coefficient which weights the nominal and target distribution

Bigger λ Stronger Regularization

Step 2:

Solve optimal Guidance Vector

$$c^* = \arg\max_c \mathcal{L}(\phi^*, c, \mathcal{D}_t)$$

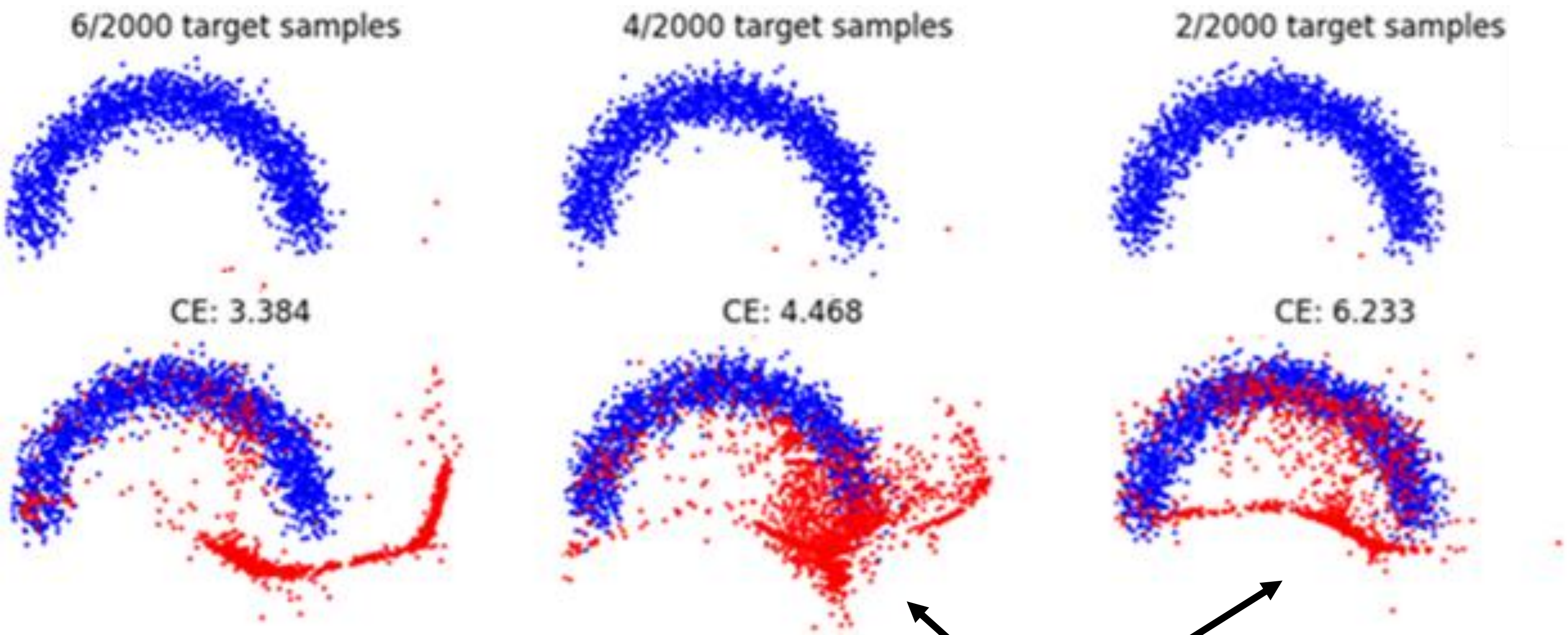


Limitations?

Hypothesis - Overfitting to Small N_t

UNIVERSITY OF HOUSTON

Hypothesis: Self-regularization may still overfit when rare event observations are extremely limited.

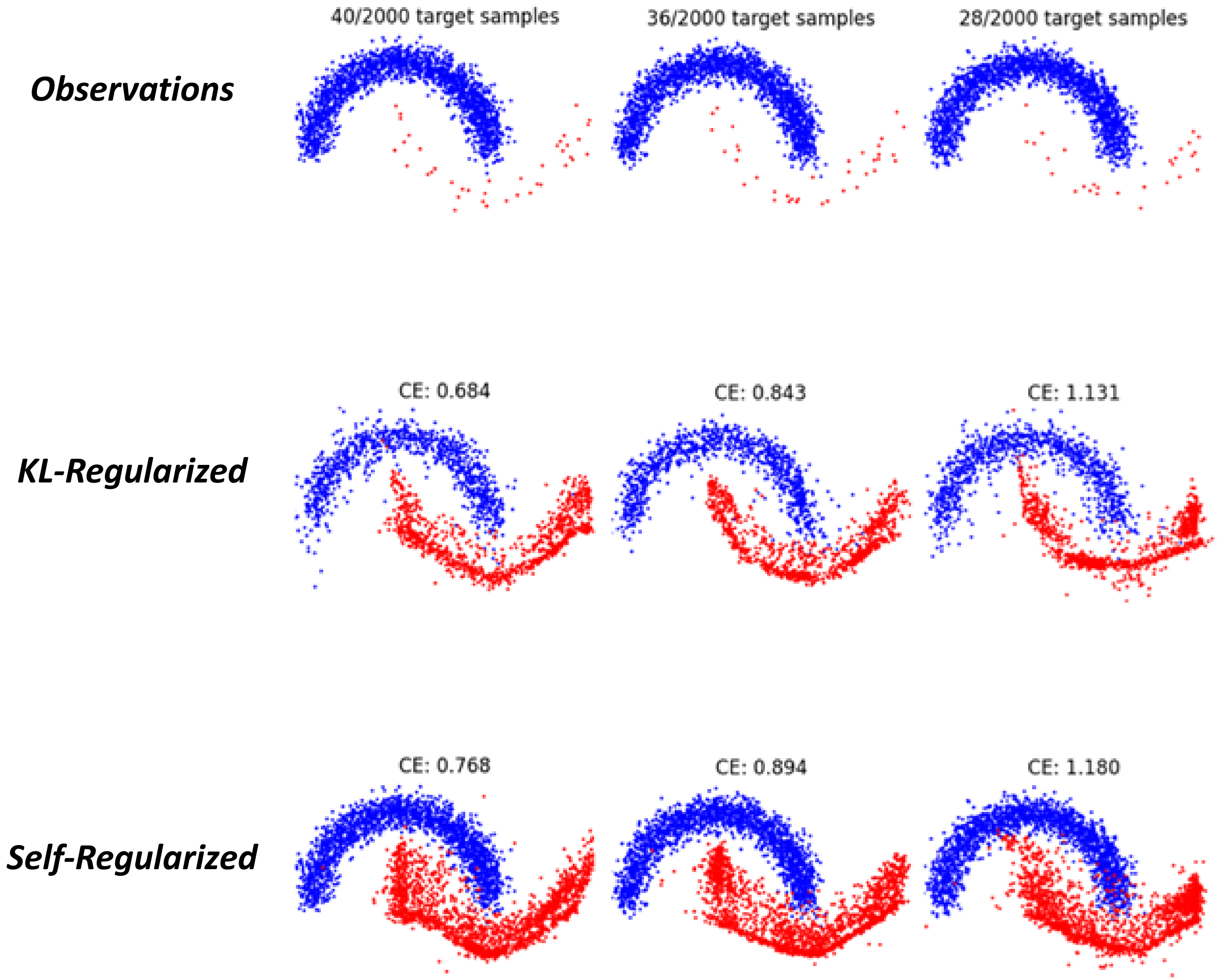


Strong Resistance to Overfitting

Hypothesis - Less Effective with Bigger N_t

UNIVERSITY OF HOUSTON

Hypothesis: Simple prior regularization methods can outperform self-regularization with more rare event samples.



Cross Entropy Test:

$$CE\left(p(z|x), q_\phi(z)\right) = -\sum p(z|x)\log q_\phi(z)$$

Lower cross entropy is better.

There exist instances in which simple KL-regularized paradigms outperform CalNF.

Hypothesis - Objective Tension

UNIVERSITY OF HOUSTON

$$\mathrm{J}(\phi, c) = \frac{1}{K}\sum_{i=1}^{K} \underbrace{\mathcal{L}(\phi, \mathbf{1}_i, \mathcal{D}_t^i)}_{\text{specialization term}} + \mathcal{L}(\phi, \mathbf{0}_K, \mathcal{D}_0) + \underbrace{\mathcal{L}(\phi, c, \mathcal{D}_t)}_{\text{generalization term}} - \beta \sum_{i \neq j}^{K} D_{KL}\left(q_\phi(\cdot; \mathbf{1}_i), q_\phi(\cdot; \mathbf{1}_j)\right)$$

Pushes ϕ to learn flows to specialize per task

Hypothesis: There exists inherent conflict between the specialization and generalization terms in the objective within parameters.

Pushes ϕ to learn flows to generalize across task

Cosine Similarity Test between Gradients: $S_C(\boldsymbol{a}, \boldsymbol{b}) = \frac{\boldsymbol{a} \cdot \boldsymbol{b}}{\|\boldsymbol{a}\|\|\boldsymbol{b}\|}$

$$K = 2$$

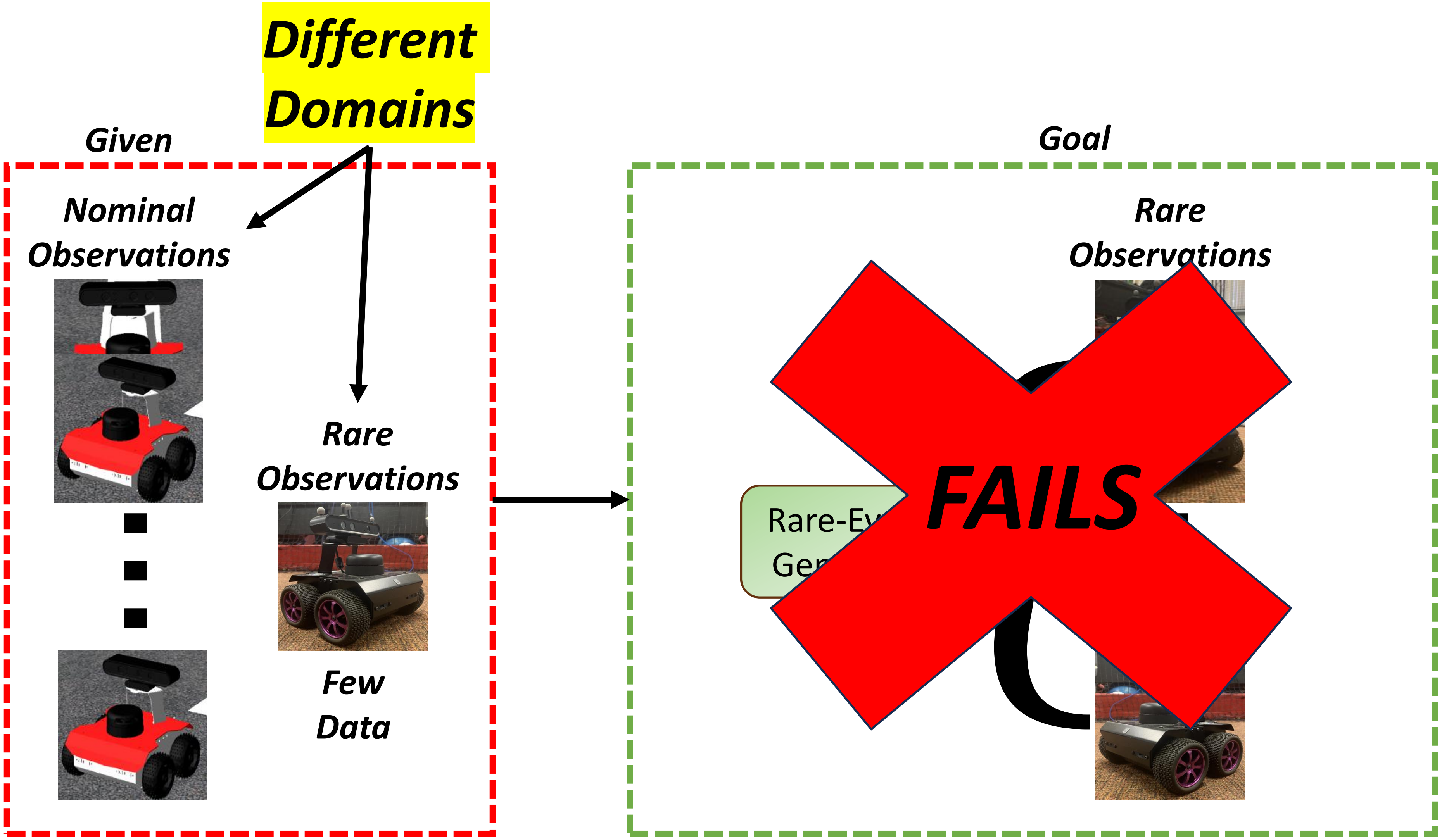
$$S_C\left(\nabla_\theta \mathcal{L}(\phi, \mathbf{1}_1, \mathcal{D}_t^1), \nabla_\theta \mathcal{L}(\phi, c^*, \mathcal{D}_t)\right) = +0.8799$$

$$S_C\left(\nabla_\theta \mathcal{L}(\phi, \mathbf{1}_2, \mathcal{D}_t^2), \nabla_\theta \mathcal{L}(\phi, c^*, \mathcal{D}_t)\right) = +0.6994$$

Since S_C is positive, gradients roughly align

Limitation - Cross Domain Adapt.

UNIVERSITY OF HOUSTON



Limitation: Nominal and rare observations MUST originate from same domain

References

[1] Dawson, C., Tran, V., Li, M. Z., & Fan, C. (2025). *Rare event modeling with self-regularized normalizing flows: What can we learn from a single failure?* (arXiv:2502.21110). arXiv. https://doi.org/10.48550/arXiv.2502.21110

[2] Abdollahzadeh, Milad, Touba Malekzadeh, Christopher T. H. Teo, Keshigeyan Chandrasegaran, Guimeng Liu, and Ngai-Man Cheung. "A Survey on Generative Modeling with Limited Data, Few Shots, and Zero Shot." arXiv, July 26, 2023. https://doi.org/10.48550/arXiv.2307.14397.