Identifiability of deep generative models without auxiliary information

Bohdan Kivva*, Goutham Rajendran*, Pradeep Ravikumar, Bryon Aragam

Goal: Learn high-level representations Z of low-level data X.

$$U \to Z \to X \qquad X = f(Z) + \varepsilon, \quad (Z|U = v) \text{ is Gaussian}$$

Special cases: Vanilla VAE, Classical ICA

Widely used in practice: SVAE, VaDE, GMVAE, DLGMM, MFC-VAE, …

Widely studied within**:** Nonlinear ICA, Deep Generative models, Latent variable models, Causality, …

Prior work: f and $P(Z)$ are identifiable (up to ~), if auxiliary information U is known

Our work: Identifiability (up to ~) **without auxiliary information**

Theorem: f is invertible (Leaky) ReLU NN => f, $P(Z)$ are identifiable from $P(X)$ up to linear transformation. Under slightly stronger assumptions $P(U, Z)$ is identifiable up to permutation, scaling and translation.

$U \in \{1, 2, 3\}$

$Z \in \mathbb{R}^m$

f

$X \in \mathbb{R}^n$