Variational Autoencoders in the Presence of Low-Dimensional Data: Landscape and Implicit Bias

Frederic Koehler*[1], Viraj Mehta*[2], Chenghui Zhou*[3], Andrej Risteski[3]

[1]Simons Institute, UC Berkeley
[2]Robotics Institute, Carnegie Mellon University
[3]Machine Learning Department, Carnegie Mellon University

Motivation

- VAE has common applications in image and language data;
  - e.g. Oord et al., 2017; Razavi et al., 2019; Vahdat & Kautz, 2020, …
- These data is hypothesized to be supported on a lower-dimensional manifold
  - Goodfellow et al., 2016
- Prior work shed insights into VAE in lower-dimensional data setting
  - Dai & Wipf, 2019

Background

Dai & Wipf proposed a 2-stage VAE based on the following hypothesis:

- In standard VAE training with tunable decoder variance, the decoder will learn the correct manifold, but not the correct density under lower-dimensional data setting.

The proposed method has good performance,

but is the hypothesis true?

Vanila VAE

2-stage VAE

Our findings

Through empirical and theoretic study, we show that:

- In linear case, the VAE recovers the support of the manifold by an implicit bias of training dynamic, but the density over the manifold is not recovered

- In nonlinear case, *neither the manifold nor the density* is recovered

Problem Setup

Synthetic data generation from **linear** and **nonlinear** manifolds:

$x = f(z)$ and $z \sim \mathcal{N}(0, I_{r*})$ and $f(\cdot)$ could be linear or nonlinear

Why do we use synthetic dataset?

- The ground-truth distribution is *accessible*
- The intrinsic dimension and ambient dimension are *known*

Lower-Dimensional Linear Data

Empirical Results:

Data Generation:

$$x = (Az, \overrightarrow{0})$$

Observations:

1. The rank of the decoder matches the intrinsic dimension; ✓
2. The model does not recover the correct density over the manifold. ✗

Informal Theorem:

To achieve the asymptotic global minima, the learnt parameter $\tilde{A}$ can have higher or equal rank to A.

However....

The gradient flow run on the VAE loss has an *implicit bias* to converge to the lowest-rank global optima

- i.e. recovering the correct subspace

Lower-Dimensional Nonlinear Data

Empirical Results:

Data Generation:

$$x = (z, \sigma(\langle a^*, z\rangle), \overrightarrow{0})$$

Observations:

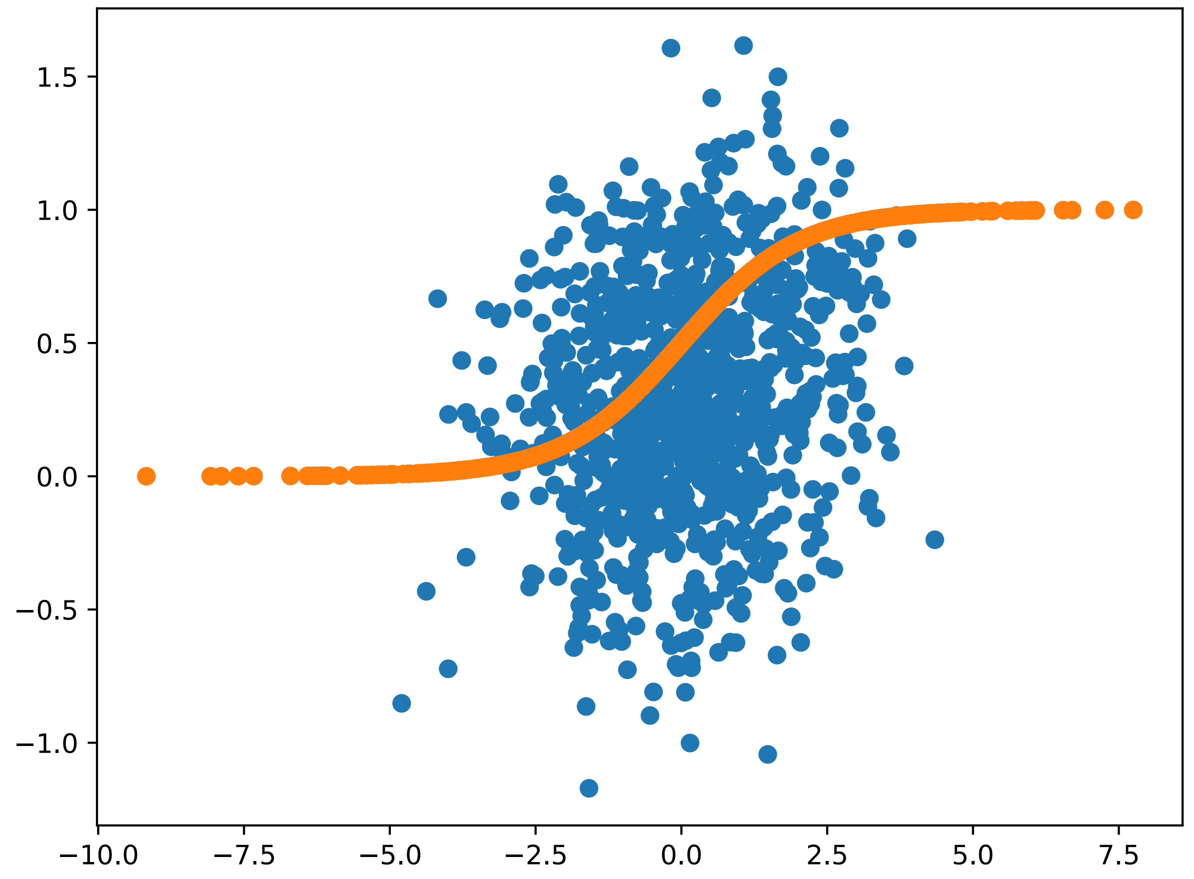
Neither the manifold ✗

nor the density ✗

is recovered.

Informal Theorem:

The learnt decoder strictly contains the ground-truth manifold.



x-axis: $\langle a^*, x_{:r^*}\rangle$;

y-axis: x_{r^*+1}

ground-truth data x in orange

generated data $\tilde{x}$ in blue

Thank You
Please Visit Poster No. 6885