BasisDeVAE: Interpretable Simultaneous Dimensionality Reduction and Feature-Level Clustering with Derivative-Based Variational Autoencoders

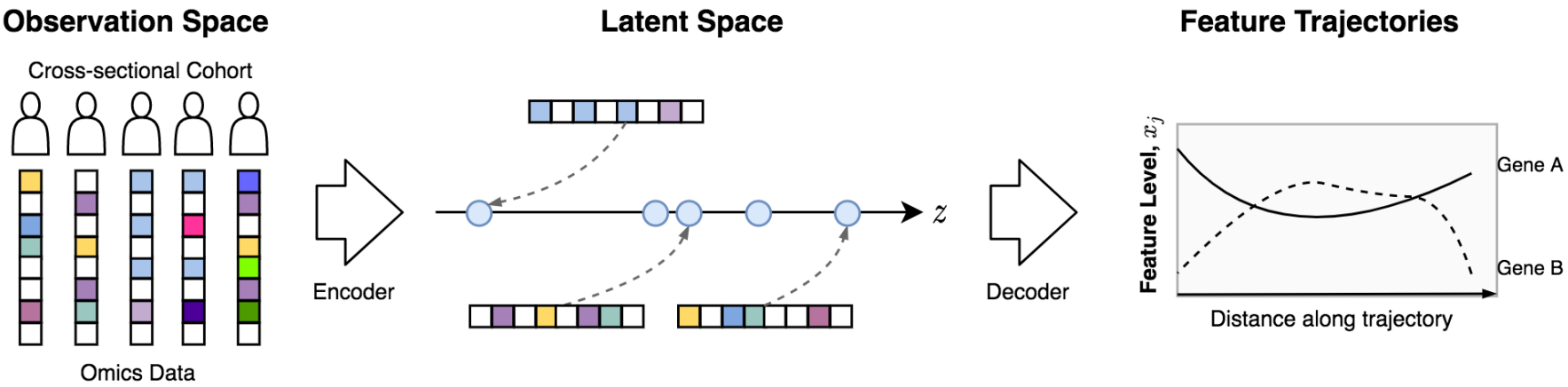
Dominic Danks[1,2] Christopher Yau[2,3,4]

[1]University of Birmingham , [2]The Alan Turing Institute , [3]Health Data Research UK,

[4]University of Manchester

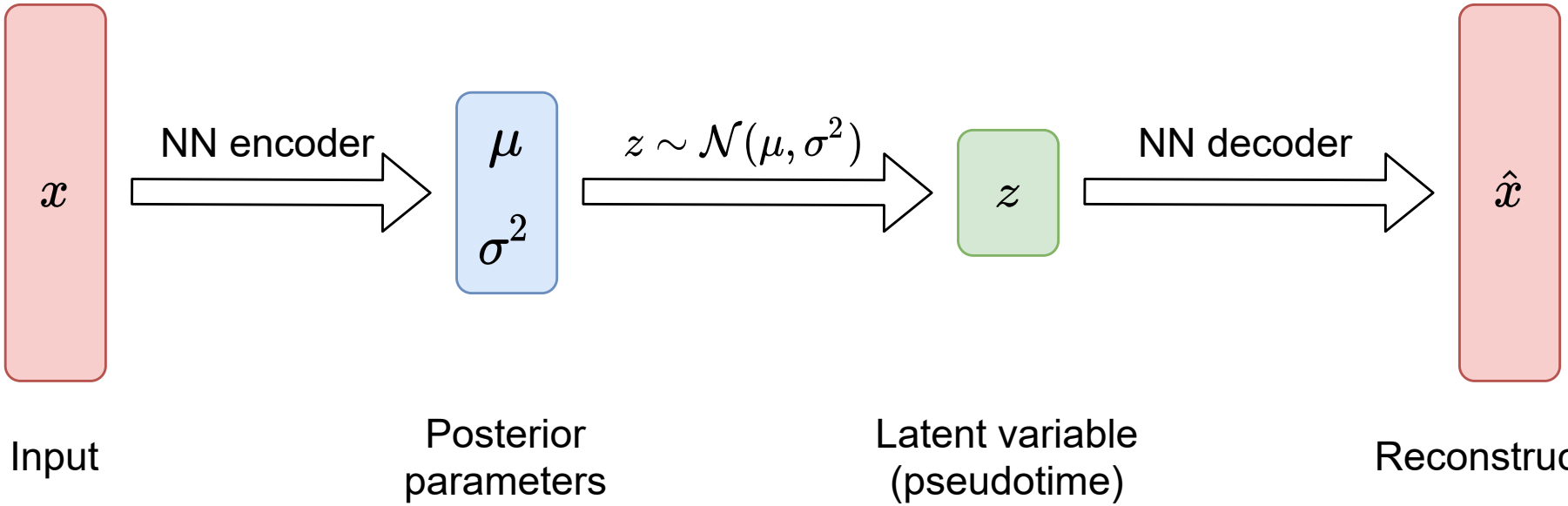
ICML 2021

Motivation



- Primary motivation: biological/disease progression.
- Basic idea: Given cross-sectional data:
  - Extract a one-dimensional latent variable (pseudotime) representing temporal progress.
  - Learn feature behaviours over pseudotime.

Variational Autoencoder



- A Variational Autoencoder (VAE) as shown above can be used to perform the aforementioned task.
- Both the encoder and decoder are neural networks.
- We introduce two VAE-related models (DeVAE and BasisDeVAE) which address shortcomings of the standard VAE in the progression modelling context.

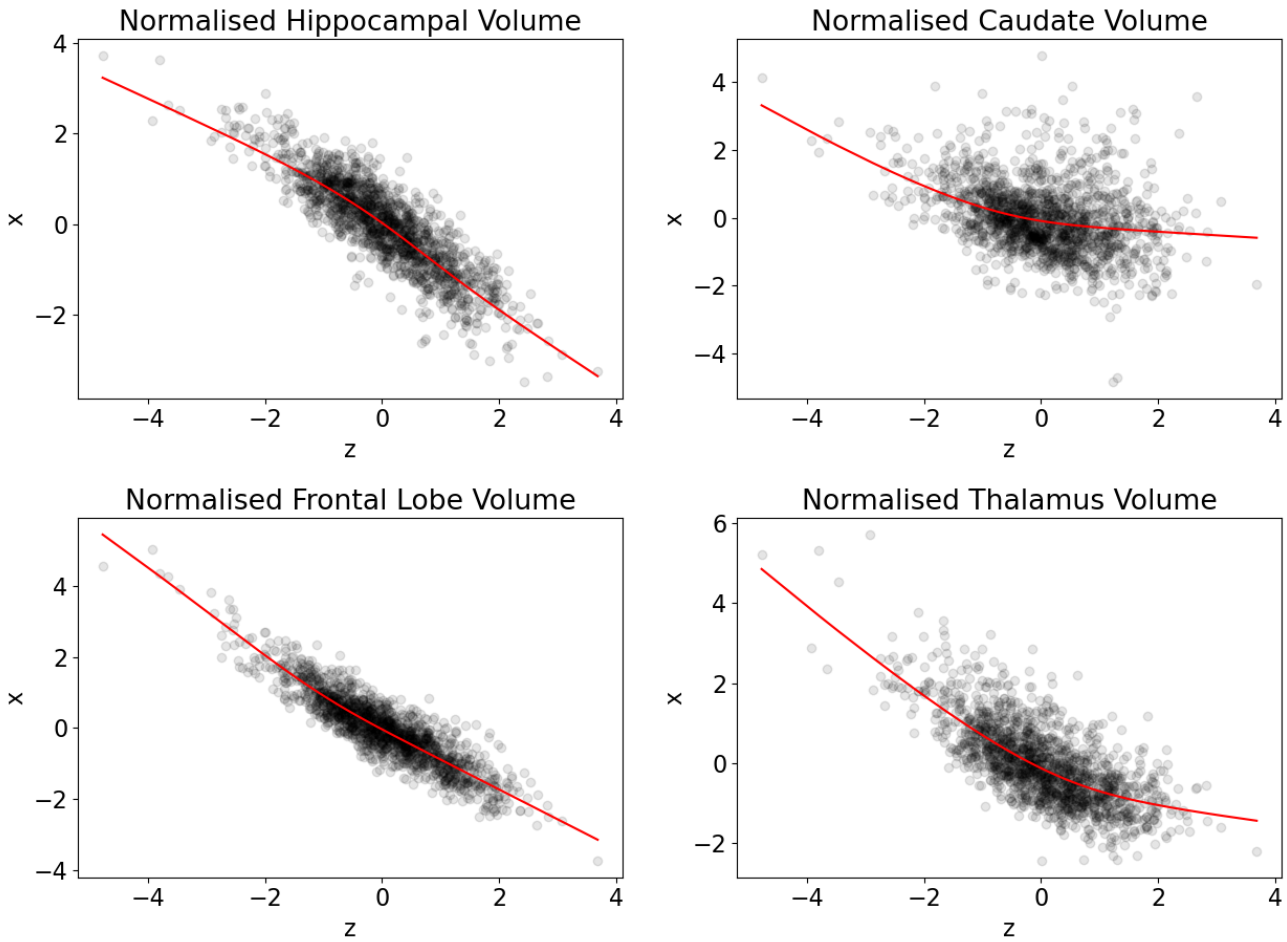
Our work: DeVAE

- DeVAE replaces the deep neural network decoder of a standard VAE with a derivative-based mapping.
- This allows constraints which are particularly common in progression modelling, namely positive and negative monotonicity and Gaussian-like transience (defined in the paper), to be naturally expressed by the decoder.
- For example replacing the NN decoder of the standard VAE, $\hat{x} = f(z)$, with the integral

$$\hat{x}(z) = x_0 - \int_0^z f(t)\,\mathrm{d}t$$

with $f(t)$ a NN with positive range results in $\hat{x}(z)$ monotonicallly decreasing with z.

Our work: DeVAE

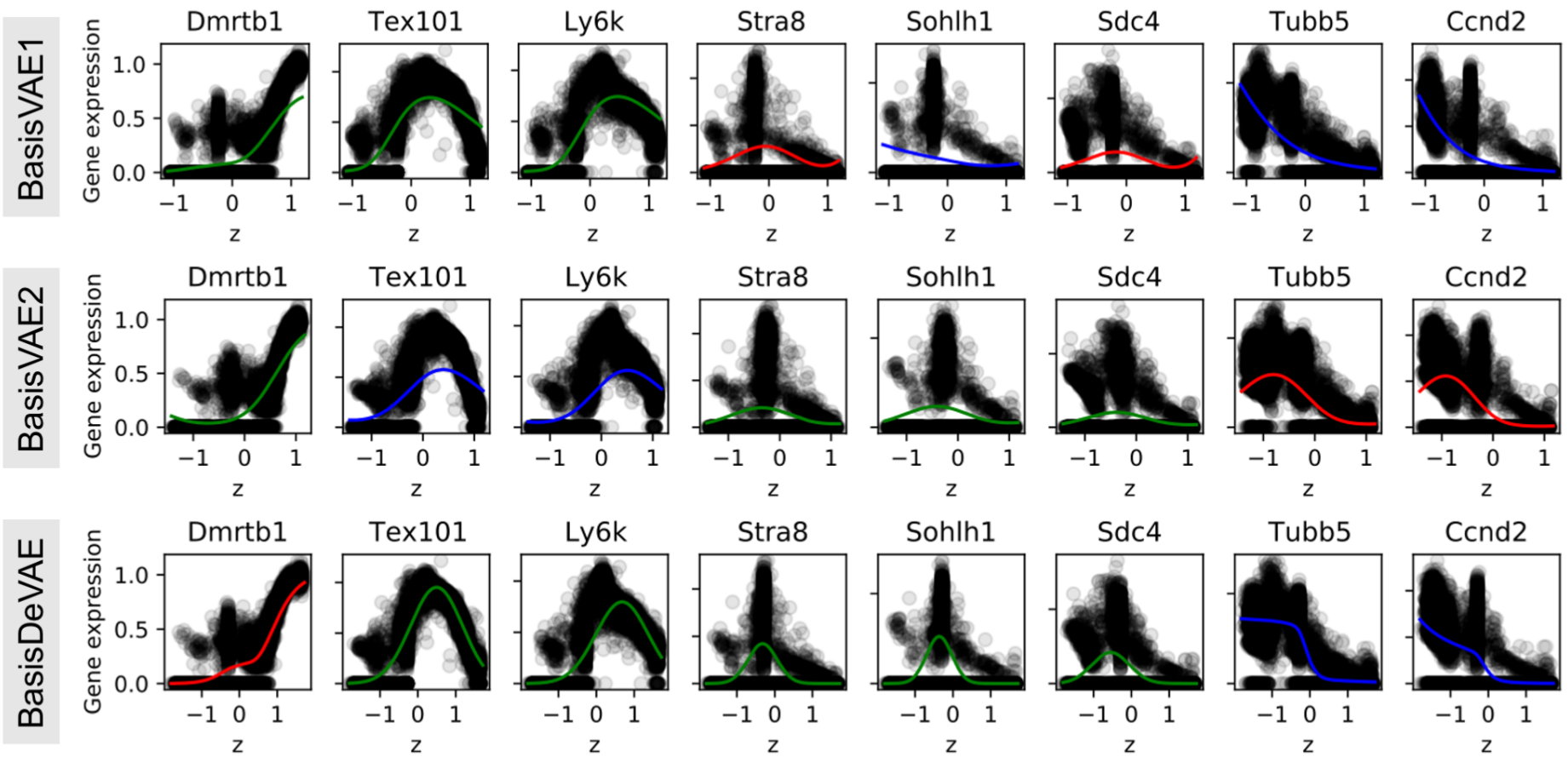


OASIS regional brain volumes. Inferred pseudotemporal profiles of 4 regional brain volumes. DeVAE extracts the monotonically decreasing pseudotemporal trajectories associated with cognitive decline.

Our work: BasisDeVAE

- DeVAE is particularly appropriate when the behaviour class of each feature is known *a priori* (e.g. brain volumes in Alzheimer's models).
- However, it is more common to know that each feature can be modelled as having one of a few known behaviours, for example positive monotonicity, negative monotonicity or Gaussian-like transience.
- BasisDeVAE addresses this setting. It uses the methodology of BasisVAE (Märtens & Yau, 2020) to assign each feature to one of the defined classes and learns its specific behaviour. It therefore performs *simultaneous dimensionality reduction and feature-level clustering*.
- As the cluster behaviours are defined *a priori*, the meaning of the cluster assignments can be immediately interpreted.

Our work: BasisDeVAE



Single-cell spermatogenesis data. Inferred clustering and pseudotemporal gene expression trajectories of eight genes involved in spermatogenesis. BasisDeVAE captures the pseudotemporal variation and naturally clusters the genes according to pseudotemporal dynamics.

Concluding remarks

See the full paper for:

- The derivative-based definition of Gaussian-like transient behaviour.
- Synthetic data experiments demonstrating behavioural differences between BasisDeVAE and BasisVAE.
- Additional results associated with real-world data experiments.
- Methodological details including mathematical model definitions, loss functions and implementation considerations.
- Expanded motivations.
- Source code.