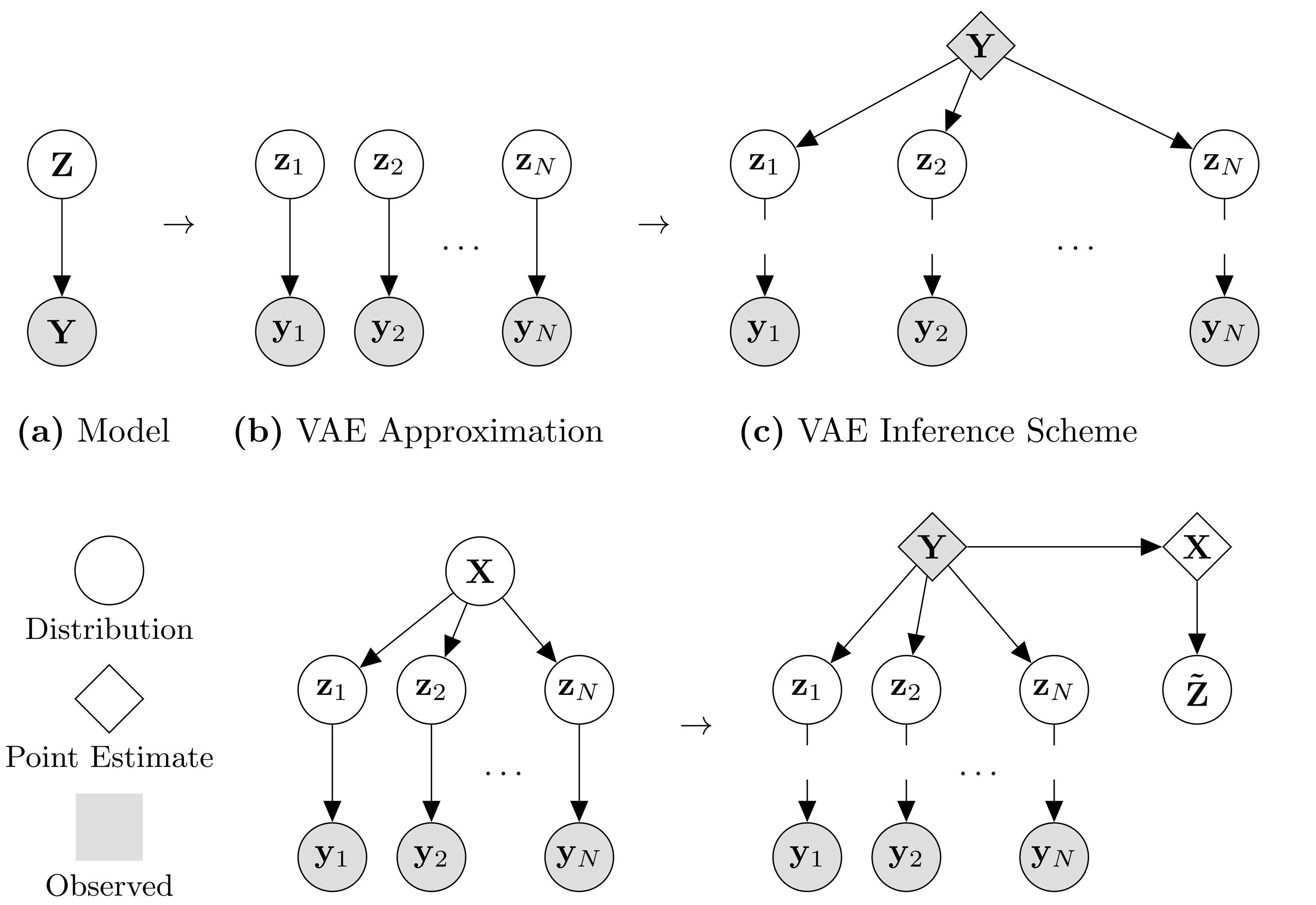
Nonparametric Inference for Auto-Encoding Variational Bayes

Erik Bodin [*], Iman Malik [*], Carl Henrik Ek [*] and Neill D.F. Campbell [†]

[*] University of Bristol [†] University of Bath

Abstract

We would like to learn latent representations that are low-dimensional and highly interpretable. A model that has these characteristics is the Gaussian Process Latent Variable Model (GP-LVM). The benefits and negative of the GP-LVM are complementary to the VAE, the former provides useful low-dimensional latent representations while the latter is able to handle large amounts of data and can use non-Gaussian likelihoods. Our inspiration for this paper is to marry these two approaches and reap the benefits of both.



(a) Model **(b)** VAE Approximation **(c)** VAE Inference Scheme

(d) Our Nonparametric VAE **(e)** Our Inference Scheme

Approximation

- $\mathbf{Z}$ no longer independent but conditionally independent given $\mathbf{X}$.
- Dependence modelled using a Gaussian Process.
- Gaussian predictive posterior parameterised directly as a convex combination.
- Match first order moment of predicted $\tilde{\mathbf{z}}$ of the GP with $\mathbf{z}$.

$$\mu(\mathbf{z}_i) = k(\mathbf{x}_i, \mathbf{X}_{\neg i})k(\mathbf{X}_{\neg i}, \mathbf{X}_{\neg i})^{-1}\mathbf{z}_{\neg i} \approx \mathbf{W}_i \mathbf{z}_{\neg i}. \tag{1}$$

$$\tilde{\mathcal{L}}_s = \mathcal{L}_g - \sum_i (\mathbb{E}[\mathbf{z}_i] - \mathbb{E}[\tilde{\mathbf{z}}_i])^2, \tag{2}$$

Assumption

- **Low-dimensional manifold in latent high-dimensional Z space**
- Tested by: Embedding high-dimensional Z space positions recovered from the VAE in a low-dimensional space X via a GP-LVM (**see initial results**).

Variational Autoencoder

$$q(\mathbf{Z} \mid \mathbf{Y}) = \prod_{i=1}^{N} q(\mathbf{z}_i \mid \mathbf{y}_i).$$

$$\tilde{\mathcal{L}}(\mathbf{y}_i) = \frac{1}{L}\sum_{l=1}^{L} \log p(\mathbf{y}_i \mid \mathbf{z}_{i,l}) - \mathrm{KL}(q(\mathbf{z} \mid \mathbf{y}_i) || p(\mathbf{z}))$$

Standard VAE formulation [Kingma and Welling(2014)]:

- unit Gaussian prior $p(\mathbf{z}) = \mathcal{N}(\mathbf{0}, \mathbf{1})$.
- trade-off between the embedded data residing at the same location in the latent space and the ability to reconstruct the data in the observed space.

Gaussian Process Latent Variable Model

A Gaussian Process can:

- be used to model functions nonparametrically
- be fully defined by a covariance function

$$p(\mathbf{Y} \mid \mathbf{X}) = \int p(\mathbf{Y} \mid \mathbf{F})p(\mathbf{F} \mid \mathbf{X})\mathrm{d}\mathbf{F} = \mathcal{N}(\mathbf{0}, k(\mathbf{X}, \mathbf{X}))$$

$$p(\mathbf{y}_* \mid \mathbf{x}_*, \mathbf{Y}, \mathbf{X}) = \int p(\mathbf{y}_* \mid \mathbf{x}_*, \mathbf{F})p(\mathbf{F} \mid \mathbf{Y}, \mathbf{X})\mathrm{d}\mathbf{F} = \mathcal{N}(\mu(\mathbf{x}_*), \Sigma(\mathbf{x}_*))$$

$$\mu(\mathbf{x}_*) = k(\mathbf{x}_*, \mathbf{X})k(\mathbf{X}, \mathbf{X})^{-1}\mathbf{Y}$$

$$\Sigma(\mathbf{x}_*) = k(\mathbf{x}_*, \mathbf{x}_*) - k(\mathbf{x}_*, \mathbf{X})k(\mathbf{X}, \mathbf{X})^{-1}k(\mathbf{X}, \mathbf{x}_*)$$

Gaussian Process Latent Variable Model [Lawrence(2005)]:

$$p(\mathbf{Y}) = \int p(\mathbf{Y} \mid \mathbf{F})p(\mathbf{F} \mid \mathbf{X})p(\mathbf{X})\mathrm{d}\mathbf{F}\mathrm{d}\mathbf{X}$$

Motivation np-VAE

VAE:

- **non-Gaussian likelihoods**

GP-LVM:

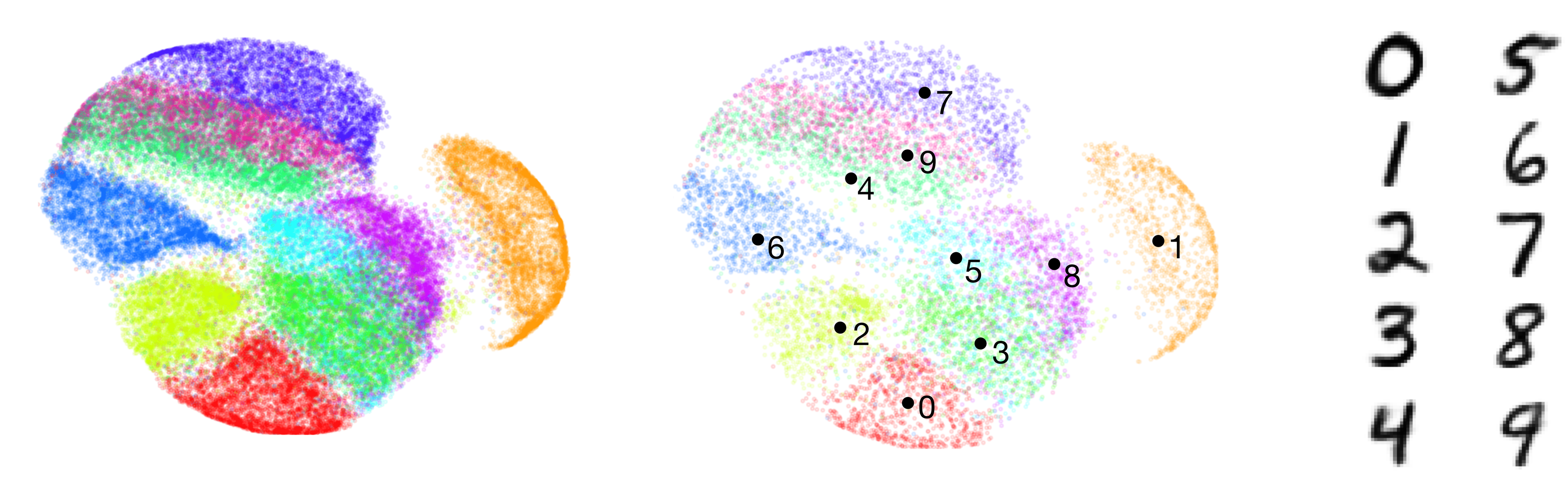
- high **interpretabilty**.
- **explicit prior over structure** via the choice of covariance function.
- **uncertainty estimation**.
- **model complexity growing with the size of the data set**.

Our inspiration for this paper is to marry these two approaches and reap the benefits of both.

Model

$$p(\mathbf{Y}) = \iint \left(\prod_{i=1} p(\mathbf{y}_i \mid \mathbf{z}_i)\right) p(\mathbf{Z} \mid \mathbf{X})p(\mathbf{X})\mathrm{d}\mathbf{Z}\mathrm{d}\mathbf{X}$$

Results



(a) Training X (b) Test X (c) Samples

Figure: Learned X space embeddings from the nonparametric VAE. Inferred X locations for (a) the training data and (b) the test data with colors encoding the MNIST digit classes. (c) Generated samples from the corresponding locations in (b) using a Z space with 500 dimensions.

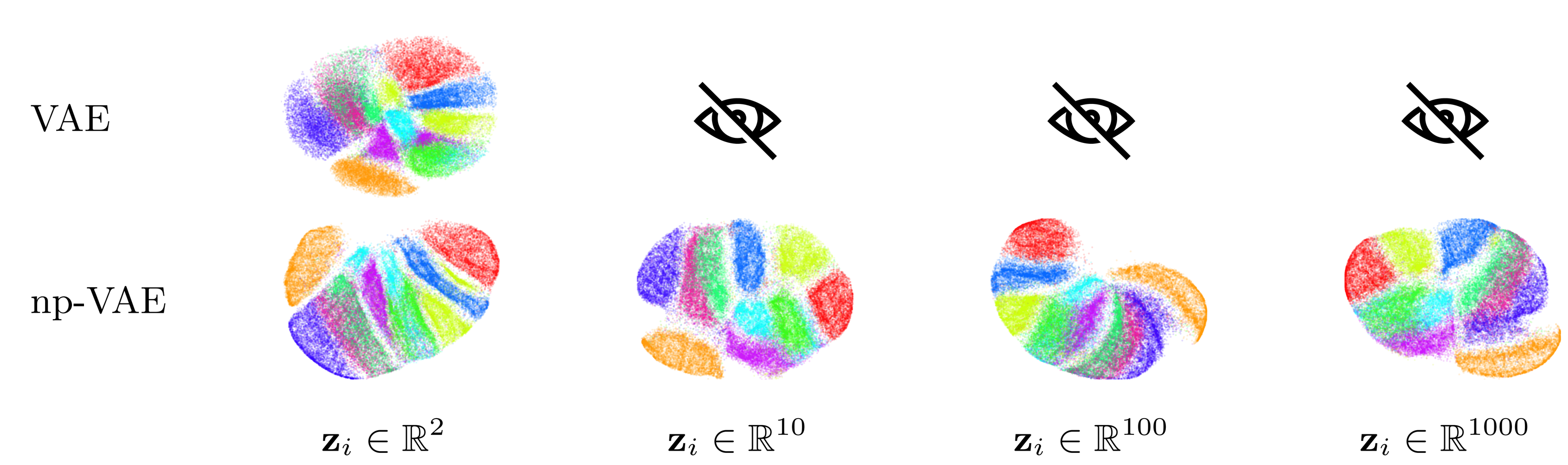


Figure: Latent space visualisation. Upper row: the Z space embedding is visualised for the standard VAE where possible (we are unable to do this for high-dimensional Z). Bottom row: the same Z space dimensionalities are used but the nonparametric VAE allows the X space to be visualised and sampled (set to be 2-dimensional). Z spaces of higher dimension become impractical to visualise and interpret whereas the X space provides an embedding for easy display and interpretation.

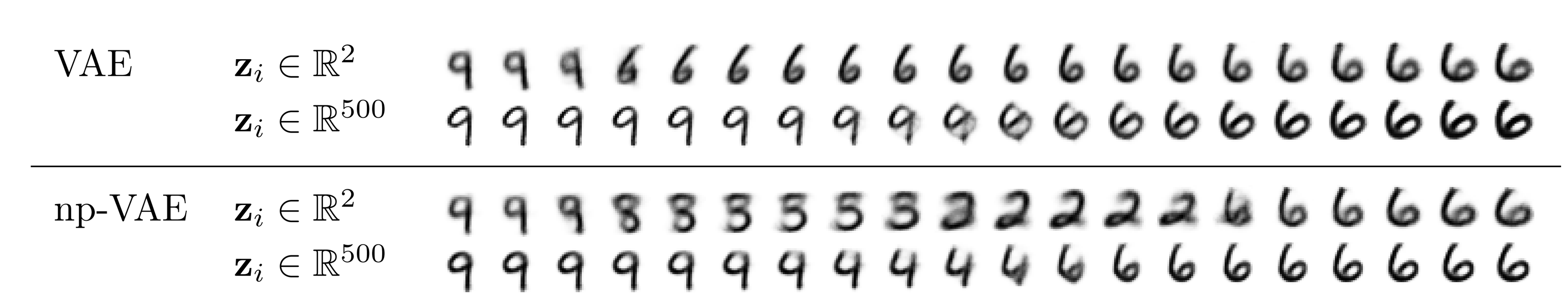


Figure: Latent space interpolation. The upper two rows show interpolants between two MNIST training examples for a standard VAE with a Z latent dimensionality of 2 and 500. The bottom two rows show interpolants between the same training examples for our nonparametric VAE with the same respective dimensionalities for Z but where the interpolation is performed in the inferred latent space X of dimension 2. We observe that a similar reconstruction quality is obtained by corresponding Z-dimensionalities, however, the interpolants from the X space of the nonparametric VAE are more meaningful with credible intermediate states between digits. Thus we can obtain a low dimensional latent space that provides interpretability without sacrificing reconstruction quality.

References

[Kingma and Welling(2014)] Diederik P Kingma and Max Welling.
Auto-Encoding Variational Bayes.
In *International Conference on Learning Representations (ICLR)*, 2014.

[Lawrence(2005)] Neil D Lawrence.
Probabilistic non-linear principal component analysis with Gaussian process latent variable models.
Journal of Machine Learning Research, 6:1783–1816, 2005.

Contact Information

Email: erik.bodin@bristol.ac.uk

EPSRC Engineering and Physical Sciences Research Council
IVA Kungl. Ingenjörsvetenskapsakademien