INTERPOLATION IN AUTO ENCODERS WITH BRIDGE PROCESSES

Carl Ringqvist, Judith Butepage, Hedvig Kjellström and Henrik Hult

The Royal Institute of Technology, Stockholm, Sweden

Introduction

Auto encoding models have been extensively studied in recent years. They provide an efficient framework for sample generation, as well as for analysing feature learning. Furthermore, they are efficient in performing interpolations between data-points in semantically meaningful ways. In this paper, we introduce a method for generating sequence samples from auto encoders trained on flattened sequences (e.g video sample from auto encoders trained on single frames); as well as a canonical, dimension independent method for generating stochastic interpolations. The distribution of interpolation paths is represented as the distribution of a bridge process constructed from an artificial random data generating process in the latent space, having the prior distribution as its invariant distribution. The suggested method can be used in many latent variable model frameworks. We concentrate on the *Variational Auto Encoder* (VAE) model for image reconstruction.

Model and Interpolation

The VAE model [2] consists of:

- An encoder $q_\phi(z|x)$, transforming images x to a representation (of lower dimension) z, through a neural network function.
- A prior $p(z)$, enforcing a structure on the latent distribution of data
- A decoder, $p_\theta(x|z)$, transforming a z-representation to an image representation x.

Most common interpolation method is *linear interpolation*:

1. Encode $x^{(i)}, x^{(j)}$ through sampling $z^{(i)} \sim q_\phi(z|x^{(i)}), z^{(j)} \sim q_\phi(z|x^{(j)})$
2. Pick points of suitable distance along the line between decoded data points: $[z^{(i)}, z, ..., z^{(j)}]$
3. Decode the latent path by sampling

$$[x^{(i)} \sim p_\theta(x|z^{(i)}), x \sim p_\theta(x|z), ..., x^{(j)} \sim p_\theta(x|z^{(j)})]$$

Fig. 1: Classical linear interpolation over the Celeb A dataset

Problem and idea

- Good samples are generally produced close to data latent representation.
- Many prior structures, especially in higher dimension, and especially the commonly used normal prior, enforces latent data representations with "holes".
- Lines between data points hence often traverse through "empty" areas of the latent space, creating images of low fidelty.
- Many methods have been developed to remedy this problem, notably *spherical interpolation* for higher dimensional normal priors.
- We suggest a novel *stochastic* interpolation scheme, that also address the above problem. We argue for that stochasticity is preferable and interesting in its own right for some applications.

Method

Our approach starts from the following observation. Given a probability p of the form $p(\mathbf{x}) \propto e^{-E(\mathbf{x})}$ a Langevin diffusion of the form

$$d\mathbf{X} = -\nabla E(\mathbf{X})dt + \sqrt{2}d\mathbf{W} \tag{1}$$

has p as its stationary distribution (under suitable conditions) In the context of VAE with $p(x)$ as the prior, the process will reside "close" to latent data representations.
In order to create an interpolation scheme from the stochastic process, we note that the corresponding bridge process, $\mathbf{X}^{x_0 x_T}$ from x_0 to x_T is given by [1]

$$\begin{aligned} d\mathbf{X}^{x_0 x_T} = &\Big(-\nabla E(\mathbf{X}^{x_0 x_T}) + \sigma\sigma^T \nabla \log p(\mathbf{x}_T, T|\mathbf{X}^{x_0 x_T}, t)\Big)dt \\ &+ \sigma d\mathbf{W}, \end{aligned} \tag{2}$$

This gives a stochastic interpolation scheme for a completely general prior. However, the term p is hard to calculate for most priors. This can be solved with numerical methods. For some priors, notably the normal distribution, p can be solved for explicitly. If the prior $p(\mathbf{z})$ is an n-dimensional standard normal distribution (e.g the prior in the VAE setting),

$$p(\mathbf{z}) = (2\pi)^{-n/2} e^{-2^{-1}\mathbf{z}\mathbf{z}^T}, \tag{3}$$

it follows that the bridge for the corresponding diffusion process reads

$$\begin{aligned} dZ = &\Big[-Z + \frac{2e^{-(T-t)}}{1-e^{-2(T-t)}}(z_T - Ze^{-(T-t)})\Big]dt \\ &+ \sqrt{2}dW, \end{aligned} \tag{4}$$

We use this bridge process for interpolation between two latent data representations, when the VAE prior p is a normal distribution. The approach outlined has the advantage of being very general. However, for normal priors, Gaussian processes can be deployed as well. The kernel parameterization of Gaussian processes allows for greater control over the properties of the bridge. For our examples, we use two kernels

$$k(h) = \exp\Big\{-\beta|h|^\alpha\Big\} \tag{5}$$

$$k(h) = \exp\Big\{-\frac{2}{\ell^2}\sin^2\Big(\pi\frac{|h|^2}{p}\Big)\Big\} \tag{6}$$

Kernel (5) is suitable for strong control over smoothness in image transitions. Kernel (6) allows for a periodic behavior of the interpolation path. In order to construct an interpolation path with a Gaussian Processes of kernel k, we consider the joint Gaussian distribution of $(Z(0), Z(t_1), \ldots, Z(t_m), Z(T))$, conditioning on $(Z(0), Z(T))$. Using the properties of conditional Gaussian distributions we obtain the mean $\hat{\mu}_{z_0,z_T}(t)$ and covariance $\hat{k}(t,s)$ for the bridge process

$$\hat{\mu}_{z_0,z_T}(t) = \frac{z_0[k(t) - k(T-t)k(T)] + z_T[k(T-t) - k(t)k(T)]}{1-k(T)^2} \tag{7}$$

and similar holds for the bridge covariance.

Relation to linear and spherical interpolation

For normal distributions p of high dimension n, it is well-known that samples are located around a sphere with radius $\sqrt{n}$ [3]. In this setting, linear interpolation with prior p can create middle images of low fidelity, since lines between points on a sphere passes through the interior. In order to remedy this phenomenon, spherical interpolation was introduced [4]. Here, interpolation is performed along geodesics on the sphere, thus assuring that the path stays within the data manifold. However, normal priors of low dimension is concentrated around the origin, rendering linear interpolation more suitable. Our method has the benefit of being somewhat a generalisation that encompasses both spherical and linear interpolation as special cases. For large T and high dimension, the interpolation path stays on the sphere of radius $\sqrt{n}$ (shown in article). For small T, randomness is eliminated, and the interpolation path is essentially linear. Figure 2 and 3 illustrate this phenomenon for the MNIST data set.

MNIST example

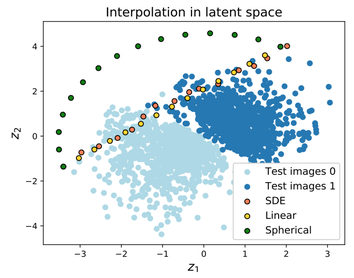

Fig. 2: Spherical interpolation does not pass through data for normal priors of low dimension

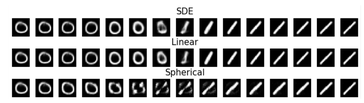

Fig. 3: This is also seen in the quality of images for the respective methods

Human poses

Fig. 4: Samples of human pose images

The following samples where generated with a Gaussian process bridge between two human pose pictures. Here, the latent space is of high dimension, and so data resides close to the sphere. We apply the periodic kernel, and use large T, to promote stochasticity and proximity to the sphere. Note that for every sample, the start and end point is the same, and that the method hence produces an interesting and plausible variability.

Fig. 5: Random walks from human poses. Scan with phone to see video

References

[1] U. Cetin and A. Danilova. "Markov Bridges: SDE Representation". In: *Stochastic Processes and their Applications, 126(3):651-679, 2016* (2016).

[2] D. Kingma and M. Welling. "Auto-encoding variational Bayes." In: *ICLR* (2014).

[3] J.S. Marron P. Hall and A. Neeman. "Geometric representation of high dimension, low sample size data." In: *Journal of the Royal Statistical Society: Series B (Stastistical Methodology), 67(3):427-444* (2005).

[4] T. White. "Sampling generative network". In: *Preprint, arXiv:1609.04468* (2016).