Factorization of High Dimensional Data using an Auto-Encoder Framework

Ross Goroshin

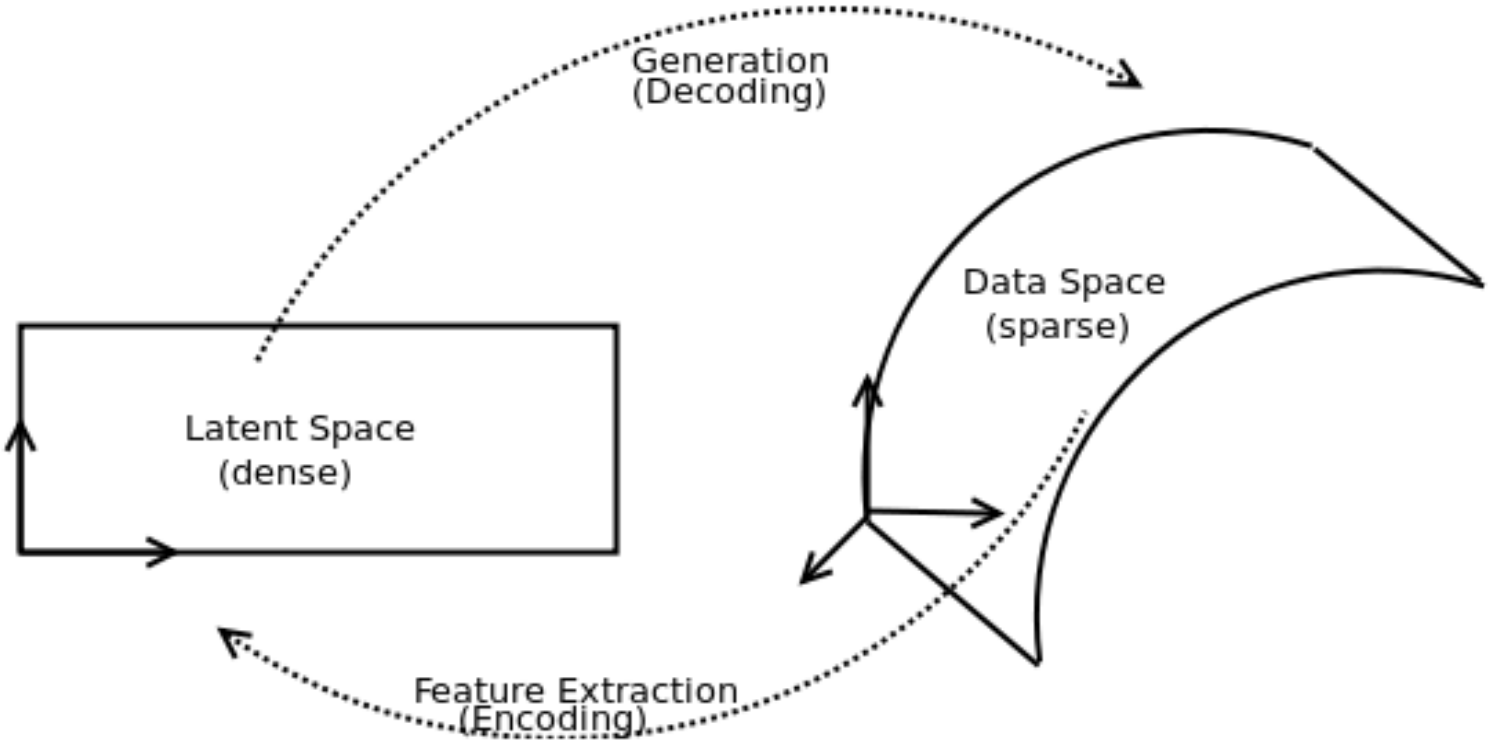
December 3, 2013

NEW YORK UNIVERSITY

Dimensionality of Data & Statistical Dependence

- Suppose we have a 42 second video played at 24 frames/second, with a resolution of 1000 by 1000 pixels
- In theory each pixel can vary independently from frame to frame, which implies that there are $\approx 10^9$ degrees of freedom
- If all of these pixels were to vary independently of one another, the picture would not be very interesting

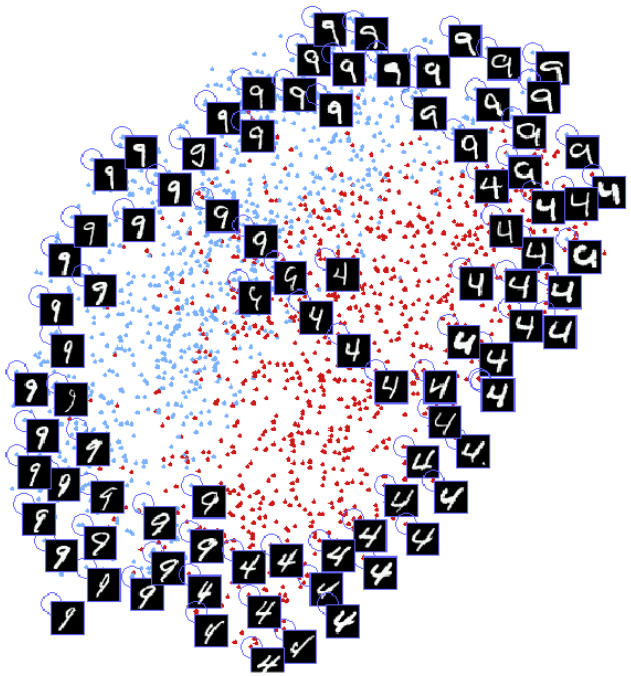
Dimensionality of Data & Statistical Dependence



- This illustration is representative of many processes
- However, dependence can be introduced without increasing the dimensionality
- Latent representation is NOT unique for generative processes of interest

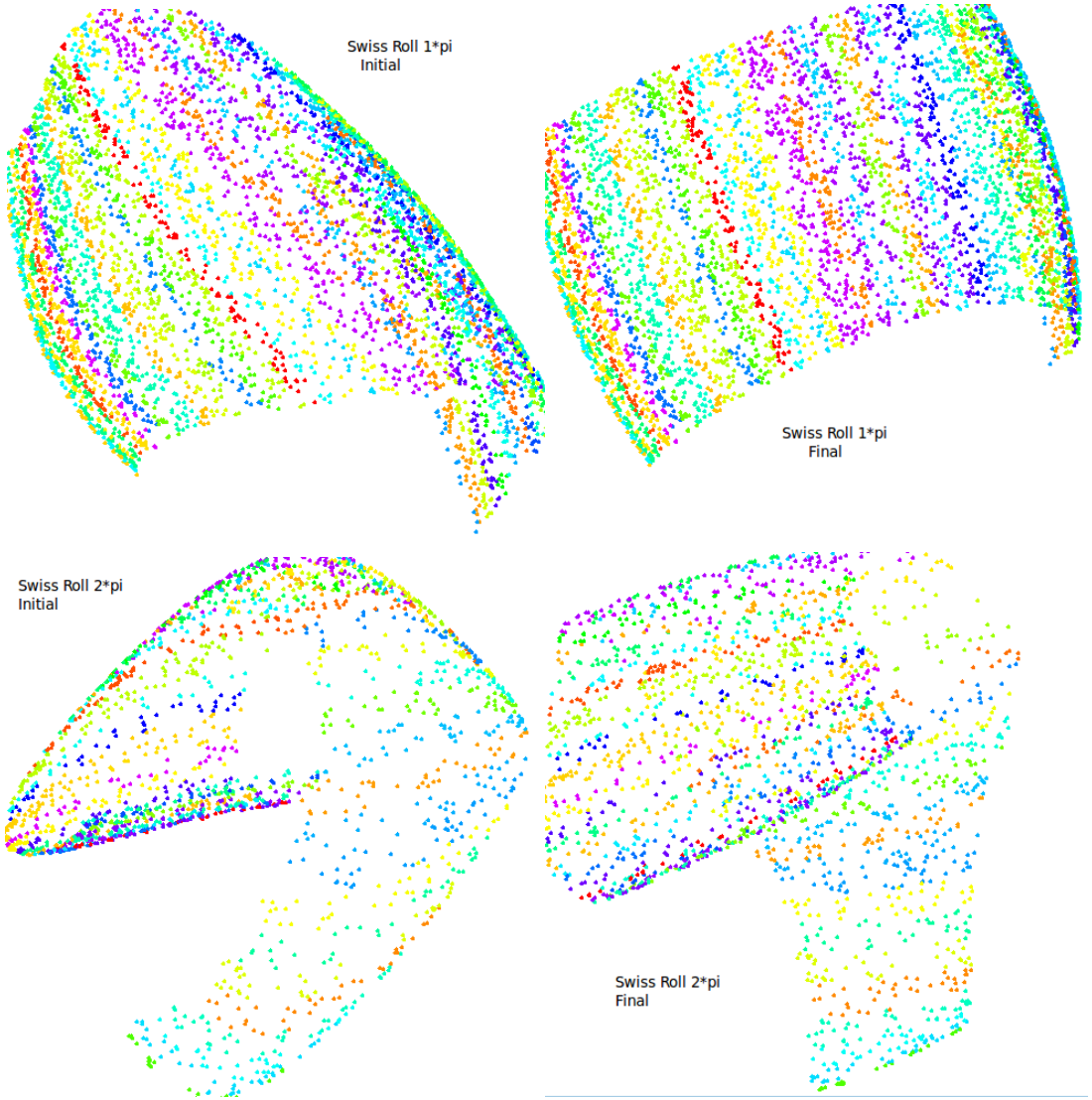
DrLIM

- We wish to find a mapping $G_W(X_i) : \mathbb{R}^D \rightarrow \mathbb{R}^d$, where $D > d$ which translates labeled similarity relationships in the input space to Euclidean distances in the output space
- If (X_1, X_2) are similar then $Y = 0$, otherwise $Y = 1$
- Let $D_W(X_1, X_2) = \|G_W(X_1), G_W(X_2)\|_2$
- $L(W, Y, X_1, X_2) = (1 - Y)\frac{1}{2}D_W^2 + Y\frac{1}{2}\{max(0, m - D_W)\}^2$



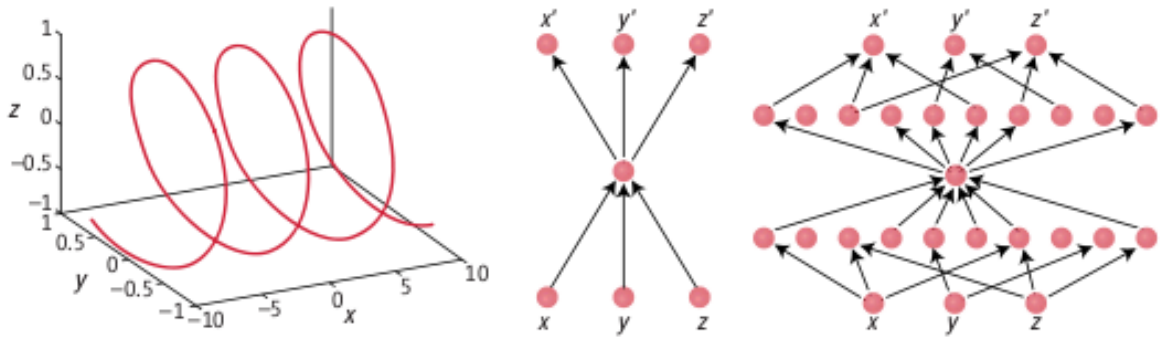
DrLIM

We can do classical metric learning using DrLIM by using L_2 as a measure of similarity in the ambient space (n-nearest neighbors)



Auto-Encoder Framework

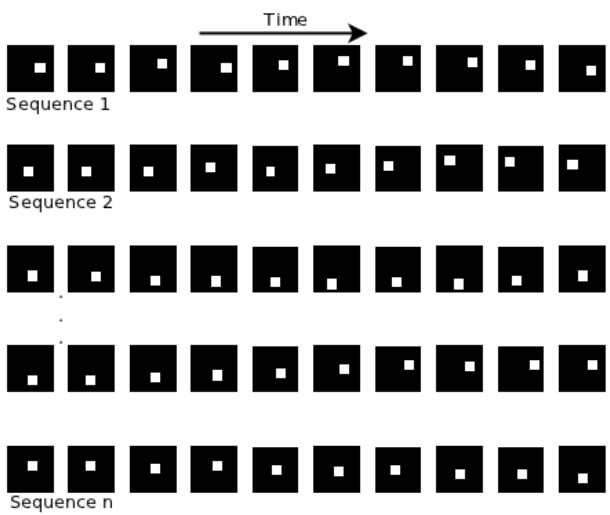
- An Auto-Encoder(AE) is composed of an "encoder" and "decoder"
- The encoder and decoder may correspond to completely different procedures (bases)
- The encoder transforms the data to latent space, and the decoder reconstructs the data from the latent representation
- AEs unify many data representation concepts and algorithms

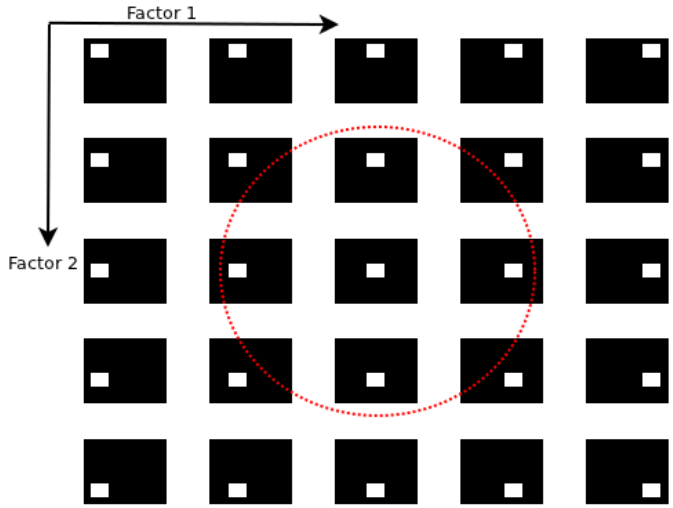


Searching for structure. (**Left**) Three-dimensional data that are inherently one-dimensional. (**Middle**) A simple "autoencoder" network that is designed to compress three dimensions to one, through the narrow hidden layer of one unit. The inputs are labeled x, y, z, with outputs x', y', and z'. (**Right**) A more complex autoencoder network that can represent highly nonlinear mappings from three dimensions to one, and from one dimension back out to three dimensions.

The Role of Time

- With no overlap, all samples are equidistant from each other
- Meaningful neighborhood relationships can only be deduced from temporally coherent sequences of images (i.e. movie clips)





Thank You

THE END