references.txt - Notepad

File Edit Format View Help

MAIN:

Doersch, C. (2016). Tutorial on variational autoencoders. arXiv preprint arXiv:1606.05908.

VAEs

VARIATIONAL INFERENCE:

Murphy's book, Chapter 21

Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. (2017). Variational inference: A review for statisticians. Journal of the American statistical Association, 112(518), 859-877.

PCA AND AUTOENCODERS:

Plaut, E. (2018). From principal subspaces to principal components with linear autoencoders. arXiv preprint arXiv:1804.10253.

EXAMPLES:

Lopes, R. G., Ha, D., Eck, D., & Shlens, J. (2019). A Learned Representation for Scalable Vector Graphics. In Proceedings of the IEEE International Conference on Computer Vision (pp. 7930-7939).

Roberts, A., Engel, J., Raffel, C., Hawthorne, C., & Eck, D. (2018). A hierarchical latent vector model for learning long-term structure in music. arXiv preprint arXiv:1803.05428. (https://magenta.tensorflow.org/music-vae)

Ha, D., & Eck, D. (2017). A neural representation of sketch drawings. arXiv preprint arXiv:1704.03477. (https://ai.googleblog.com/2017/04/teaching-machines-to-draw.html)

A Learned Representation for Scalable Vector Graphics

Raphael Gontijo Lopes*, David Ha, Douglas Eck, Jonathon Shlens
Google Brain
{iraphael, hadavid, deck, shlens}@google.com

Abstract

Dramatic advances in generative models have resulted in near photographic quality for artificially rendered faces, animals and other objects in the natural world. In spite of such advances, a higher level understanding of vision and imagery does not arise from exhaustively modeling an object, but instead identifying higher-level attributes that best summarize the aspects of an object. In this work we attempt to model the drawing process of fonts by building sequential generative models of vector graphics. This model has the benefit of providing a scale-invariant representation for imagery whose latent representation may be systematically manipulated and exploited to perform style propagation. We demonstrate these results on a large dataset of fonts and highlight how such a model captures the statistical dependencies and richness of this dataset. We envision that our model can find use as a tool for graphic designers to facilitate font design.

1. Introduction

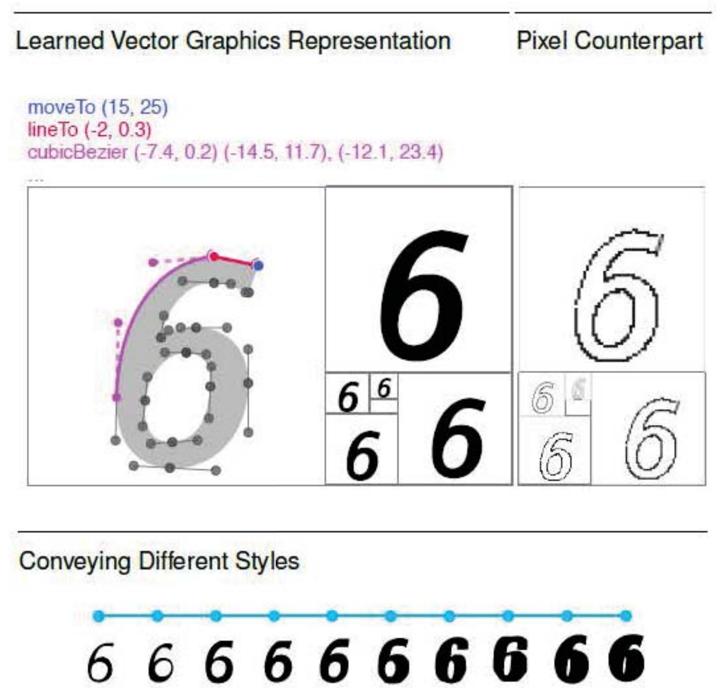


Figure 1: **Learning fonts in a native command space**. Unlike pixels, scalable vector graphics (SVG) [11] are scale-invariant representations whose parameterizations may be

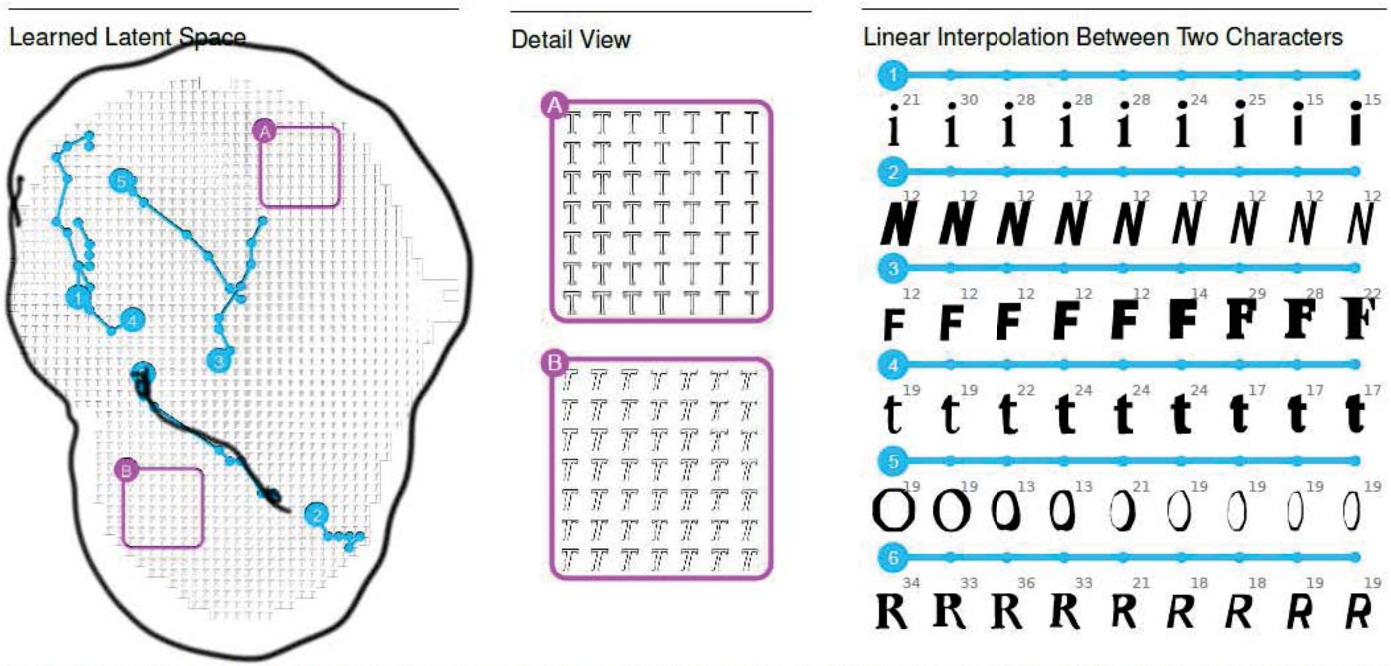


Figure 4: **Learning a smooth, latent representation of font style**. UMAP visualization [39] of the learned latent space z across 1 M examples (left). Purple boxes (A, B) provide a detail view of select regions. Blue lines (1-9) indicate *linear* interpolations in the full latent space z between two characters of the dataset. Points along these linear interpolations are rendered as SVG images. Number in upper-right corner indicates number of strokes in SVG rendering. Best viewed in digital color.

0 1 2 3 4 5 6 7 8 9
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
a b c d e f g h i j k l m n o p q r s t u v w x y z
0 1 2 3 4 5 6 7 8 9
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
a b c d e f g h i j k l m n o p q r s t u v w x y z
0 1 2 3 4 5 6 7 8 9
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
a b c d e f g h i j k l m n o p q r s t u v w x y z
0 1 2 3 4 5 6 7 8 9
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
a b c d e f g h i j k l m n o p q r s t u v w x y z
0123456789
ABCDEFGHIJKLMNOPQRSTUVWXYZ
abcdefghijklmnopqrstuvwxyz
0 1 2 3 4 5 6 7 8 9
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
a b c d e f g h i j k l m n o p q r s t u v w x y z
0 1 2 3 4 5 6 7 8 9
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
a b c d e f g h i j k l m n o p q r s t u v w x y z
0 1 2 3 4 5 6 7 8 9
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
a b c d e f g h i j k l m n o p q r s t u v w x y z
0 1 2 3 4 5 6 7 8 9
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
a b c d e f g h i j k l m n o p q r s t u v w x y z

A Neural Representation of Sketch Drawings

David Ha
Google Brain
hadavid@google.com

Douglas Eck
Google Brain
deck@google.com

Abstract

We present sketch-rnn, a recurrent neural network (RNN) able to construct stroke-based drawings of common objects. The model is trained on a dataset of human-drawn images representing many different classes. We outline a framework for conditional and unconditional sketch generation, and describe new robust training methods for generating coherent sketch drawings in a vector format.

A Hierarchical Latent Vector Model for Learning Long-Term Structure in Music

Adam Roberts [1] Jesse Engel [1] Colin Raffel [1] Curtis Hawthorne [1] Douglas Eck [1]

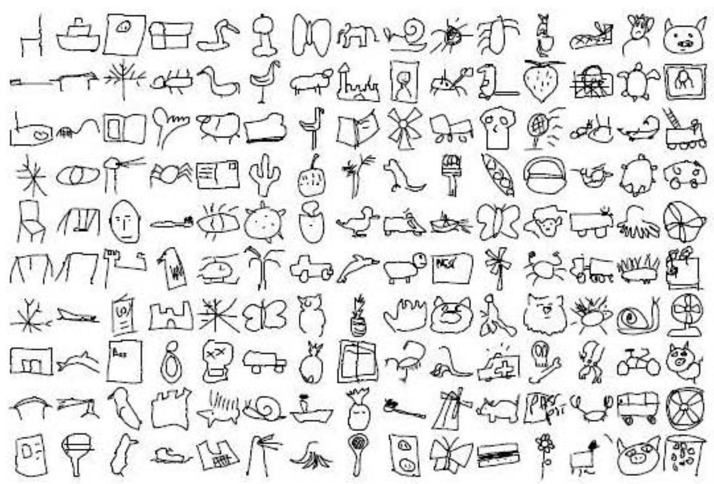
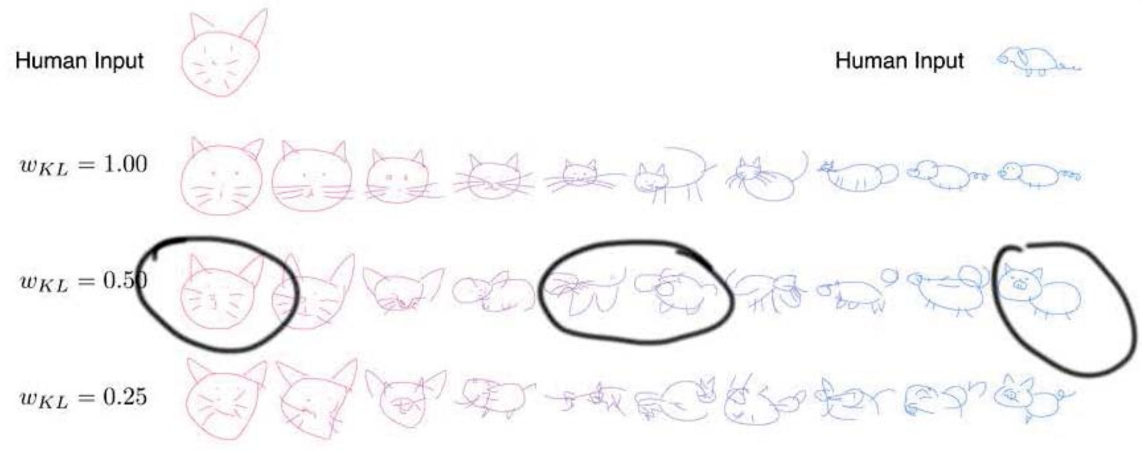


Figure 1: Example sketch drawings from QuickDraw dataset.



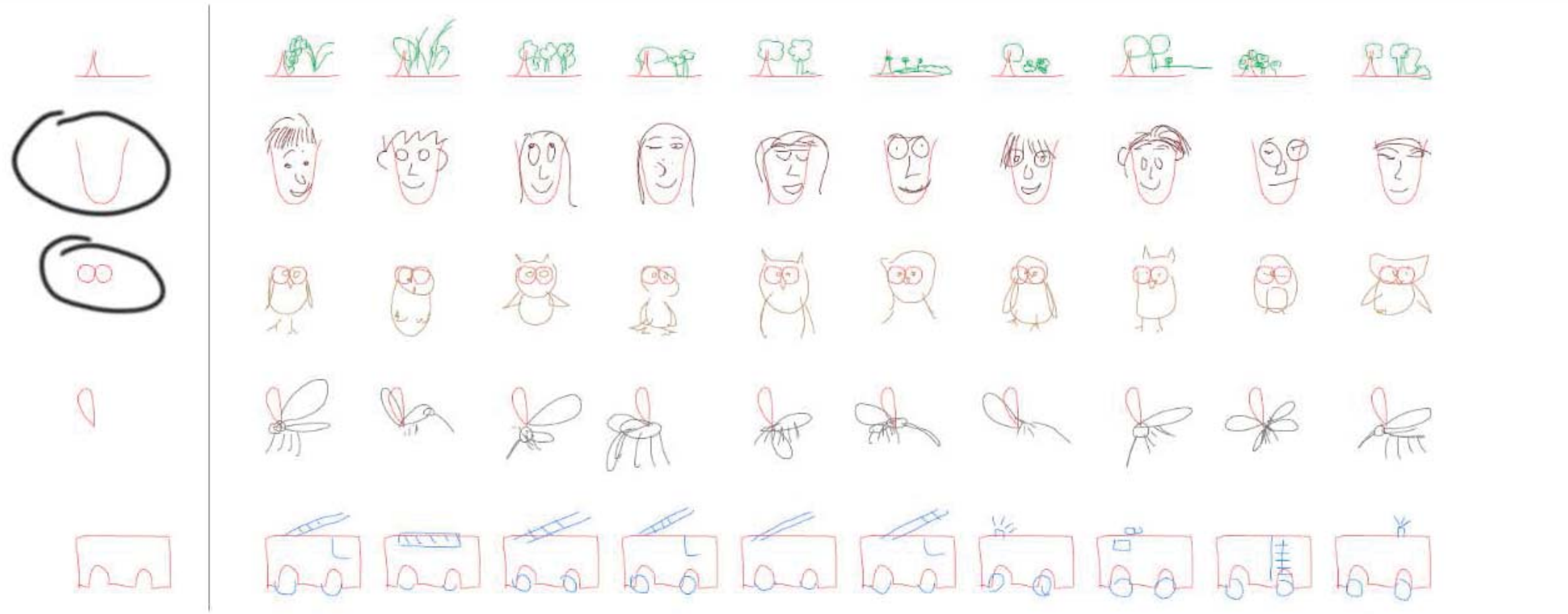
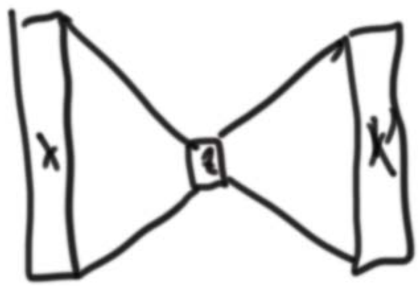


Figure 7: sketch-rnn predicting possible endings of various incomplete sketches (the red lines).

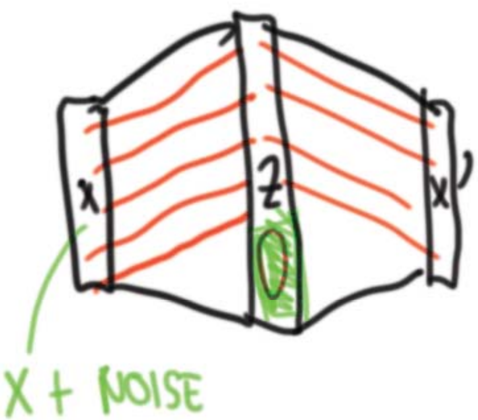
TYPICAL TASKS

DIM. RED. ($Q << p$, UNDERCOMPLETE AUTOENCODER)



- FASTER
- EASIER TO INTERPRET

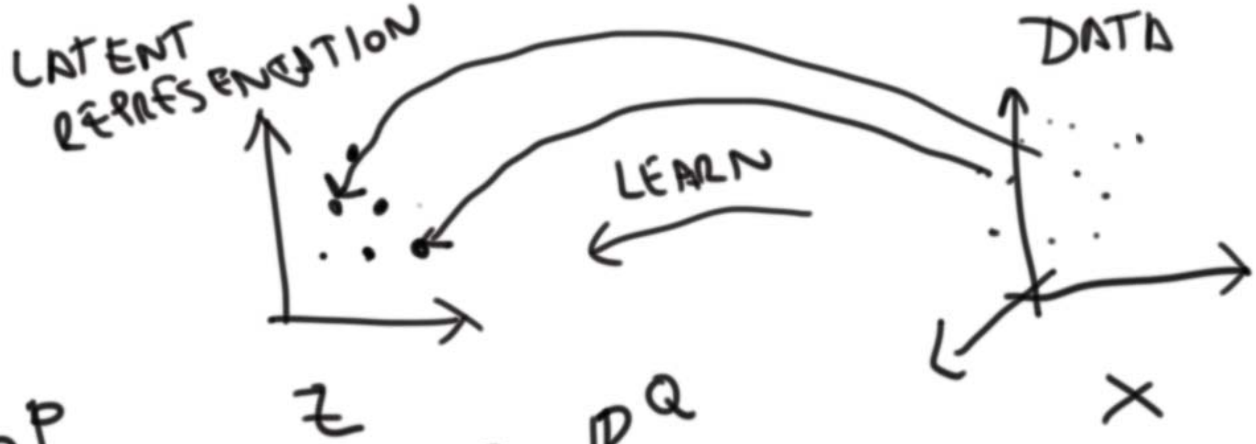
DENOISING ($Q > p$, OVERCOMPLETE AUTOENC.)



$l(x, x')$

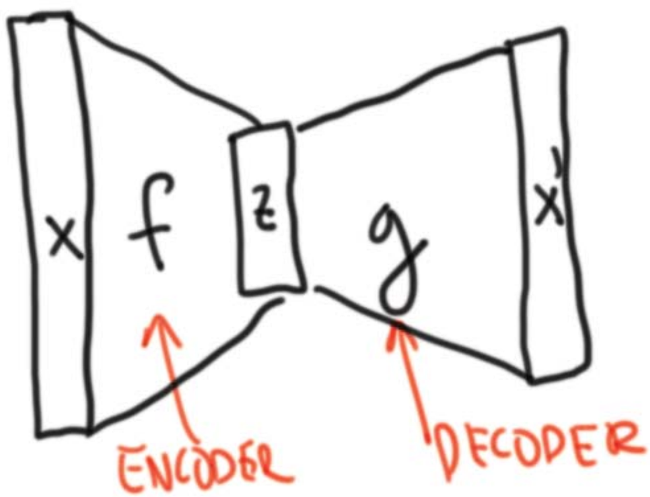
AUTOENCODER

UNSUPERVISED LEARNING TECHNIQUE



$x \in \mathbb{R}^P$

$z \in \mathbb{R}^Q$



MINIMIZE LOSS

$\ell(x, x')$

(TYPICALLY NEURAL NETS)

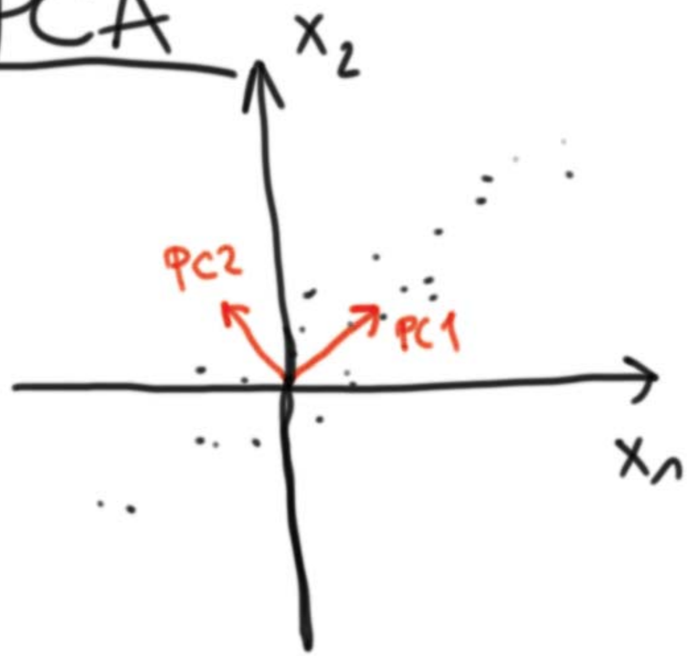
FIND f, g THAT MINIMIZE $\ell(x, x')$

$f: \mathbb{R}^P \to \mathbb{R}^Q$, $g: \mathbb{R}^Q \to \mathbb{R}^P$

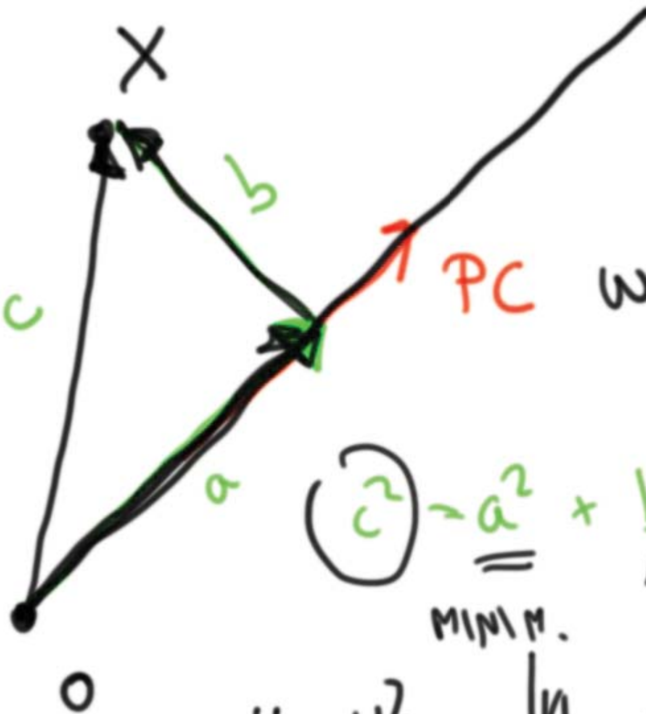
VAE (VARIATIONAL AUTOENCODER)

- AUTOENCODER
- PROBABILISTIC (BAYESIAN)
- VARIATIONAL INFERENCE

PCA



MAX. VAR. ≡ MIN. MSE



$$c^2 = a^2 + b^2$$

MINIM. MAX.

$$\|x\|_2^2 = \|w^T x\, w\|_2^2 + \|x_i - w^T x\, w\|^2$$

A LINEAR AUTOENCODER

$f(x) = Ax$ — A IS A $q \times p$ MATRIX

$g(x) = Bx$ — B IS A $p \times q$ MATRIX

LOSS = MSE

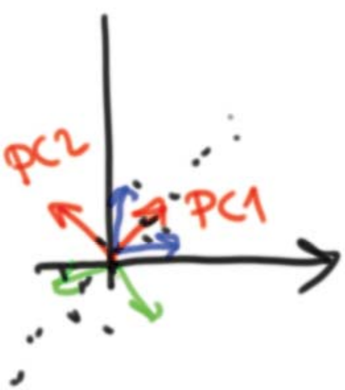
$$\sum_{i=1}^{m} (x_i - BAx_i)^2$$

PCA:

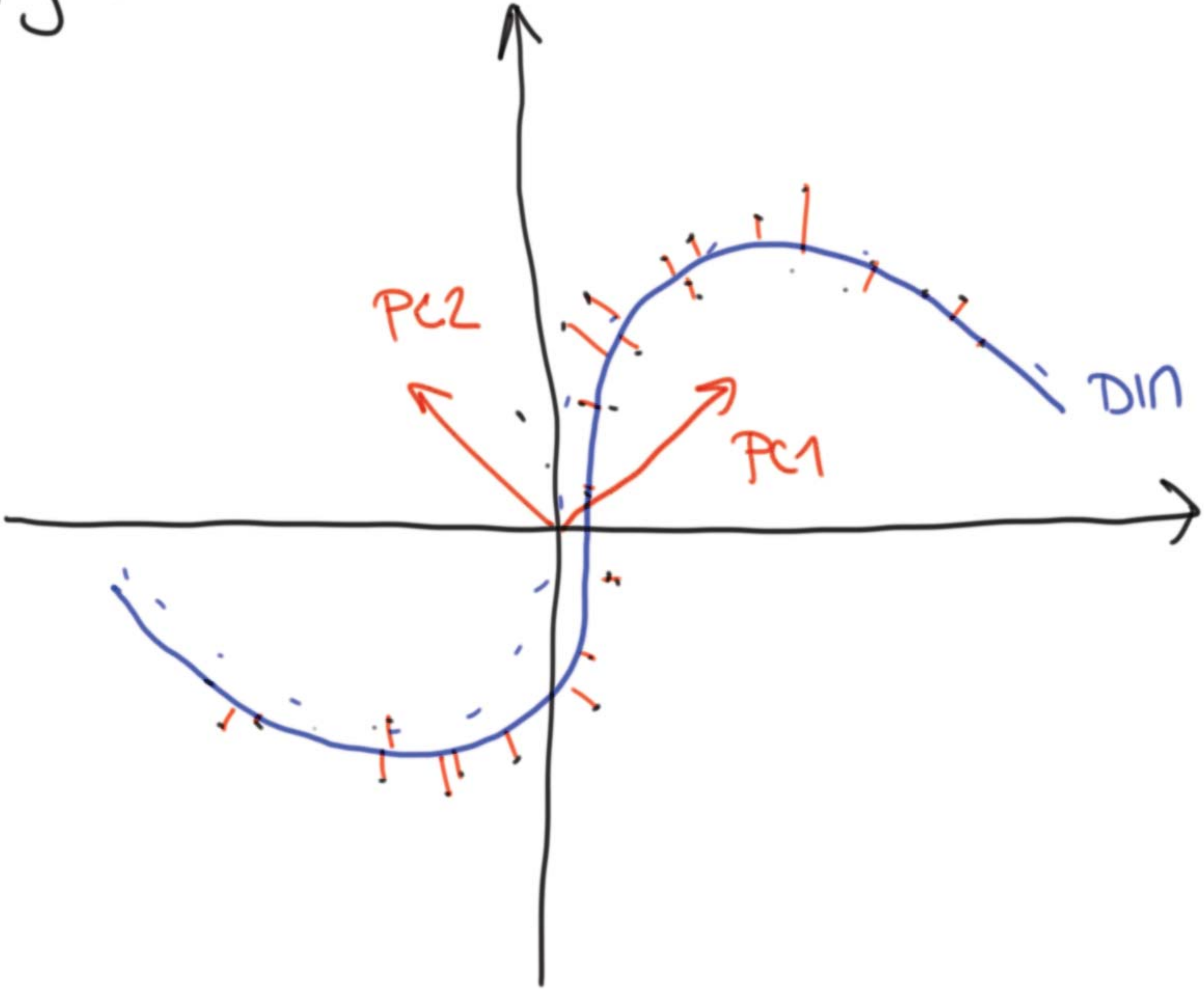
$$A = W^T$$
$$B = W$$

$$\sum_{i=1}^{m} (x_i - WW^T x_i)^2$$

PC NEED TO BE **ORTHOGONAL**



f, g ARE NONLINEAR ⇒ NONLIN. DIM. REDUCTION



VAE

- PROBABILISTIC AUTOENCODER ("BAYESIAN") (GENERATIVE)

$p(z)$ z $\quad z_i \sim p(z) \xrightarrow{g} p(x_i|z), \quad x_i \sim p(x|z_i)$

TYPICAL: $z \sim N(0, I)$ TYPICAL NN

$p_\theta(x|z) \qquad x|z \sim N(\mu_\theta(z), \Sigma_\theta(z))$

HOW DO WE "LEARN"?

$$p_\theta(x) = \underbrace{\int p_\theta(x|z)\, p(z)\, dz}_{\text{TYPICALLY NOT TRACTABLE}}$$

$p_\theta(x|z)$: N; $p(z)$: N

μ_x IS A LINEAR FUNCT. OF z

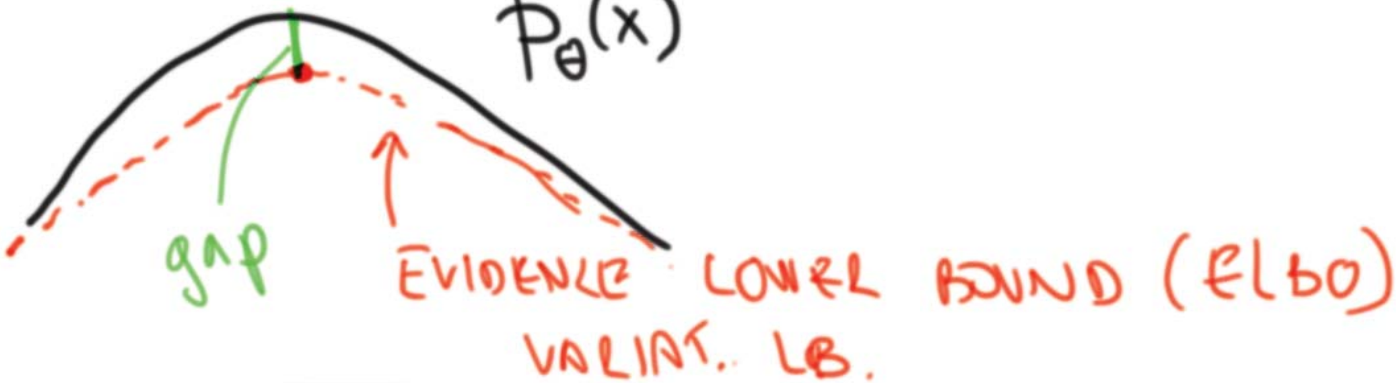
$$\theta^* = \arg\max_\theta p_\theta(x)$$

MONTE CARLO

$$p_\theta(x) \approx \frac{1}{m} \sum_{i=1}^{m} p_\theta(x|z^{(i)}) \qquad z^{(i)} \sim p(z)$$

IS ALSO INFEASIBLE

KEY IDEA: INSTEAD OF MAX. $p_\theta(x)$,
MAXIMIZE A LOWER BOUND OF $p_\theta(x)$



$$p_\theta(x) = \frac{p_\theta(x|z)\,p(z)}{p_\theta(z|x)}$$

BAYES TH.

$$p_\theta(z|x) = \frac{p(x|z)\,p(z)}{\int p(x|z)\,p(z)\,dz}$$

$Q_\phi(z|x)$ typically $= N(\mu_\phi(z), \Sigma_\phi(z))$

SOME MATH...

BAYES TH.

$$\log p_\theta(x) = E_{z\sim Q(z|x)}\left[\log p_\theta(x)\right] = E_Q\left[\log \frac{p_\theta(x|z)p(z)}{p_\theta(z|x)}\right]$$

$E_z[c] = c\,E[1]$

$KL(p\|Q) = \int p(x) \log\frac{p(x)}{Q(x)}dx$

$$= E_Q\left[\log \frac{p_\theta(x|z)\,p(z)}{p_\theta(z|x)} \cdot \frac{Q_\phi(z|x)}{Q_\phi(z|x)}\right] =$$

$= E_Q[\log p_\theta(x|z)] - KL(p_{z|x}\|p)$

$$= E_Q[\log p_\theta(x|z)] - E_Q\left[\log\frac{Q_\phi(z|x)}{p(z)}\right] + E\left[\log\frac{Q_\phi(z|x)}{p_\theta(z|x)}\right]$$

$KL(Q\|p_z)$ $\qquad$ $KL(Q|p_{z|x})$

RELATIVELY EASY TO COMPUTE

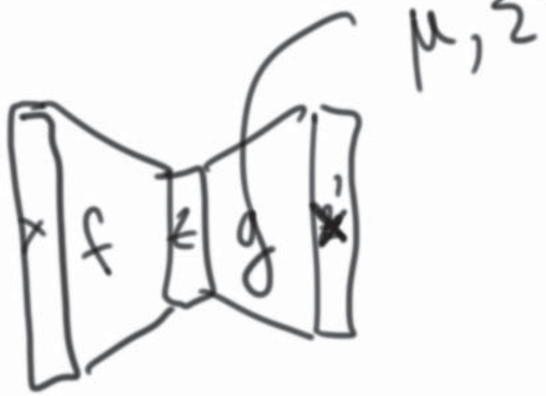
$$\geq E_Q[\log p_\theta(x|z)] - KL(Q_\phi\|p_z)$$

(ELBO)

OUR LOSS FUNCTION

EASY TO COMPUTE

$\theta, \phi \leftarrow$ MAXIMIZE OVER THESE



- AUTOENCODERS ARE USEFUL
- VAE ARE "BETTER"
- VI (VARIATIONAL INF.)
  (ALT./COMPL. TO MCMC)

GEN. NEW OBS.

$z_i \sim p(z)$

$\mu(z_i), \Sigma(z_i)$

$x_i \sim N(\mu(z_i), \Sigma(z_i))$

~~COMPUTE x~~

COMPUTE LATENT REP. OF x_i

$p(z|x)$

$Q(z|x)$

PROBABILISTIC ENCODER

PROB. DECODER

$p_\theta(x|z)$

10:30

GAUSSIAN PROCESSES