

CS489/698

Lecture 20: March 19, 2018

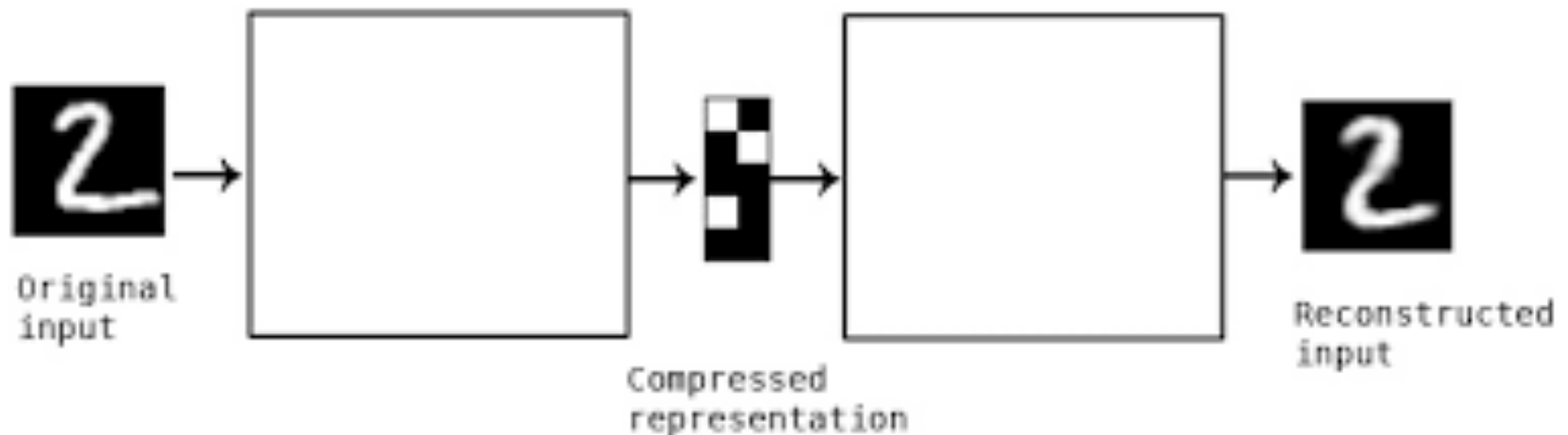
Autoencoders
[GBC] Chap. 14

Autoencoder

- Special type of feed forward network for
 - Compression
 - Denoising
 - Sparse representation
 - Data generation

Autoencoder

- Encoder: $f(\)$
- Decoder: $g(\)$
- Autoencoder: $g(f(x)) = x$



Linear Autoencoder

- f and g are linear
 - Matrix representations: $\mathbf{W}_f$ and $\mathbf{W}_g$
- Picture:

Linear Autoencoder

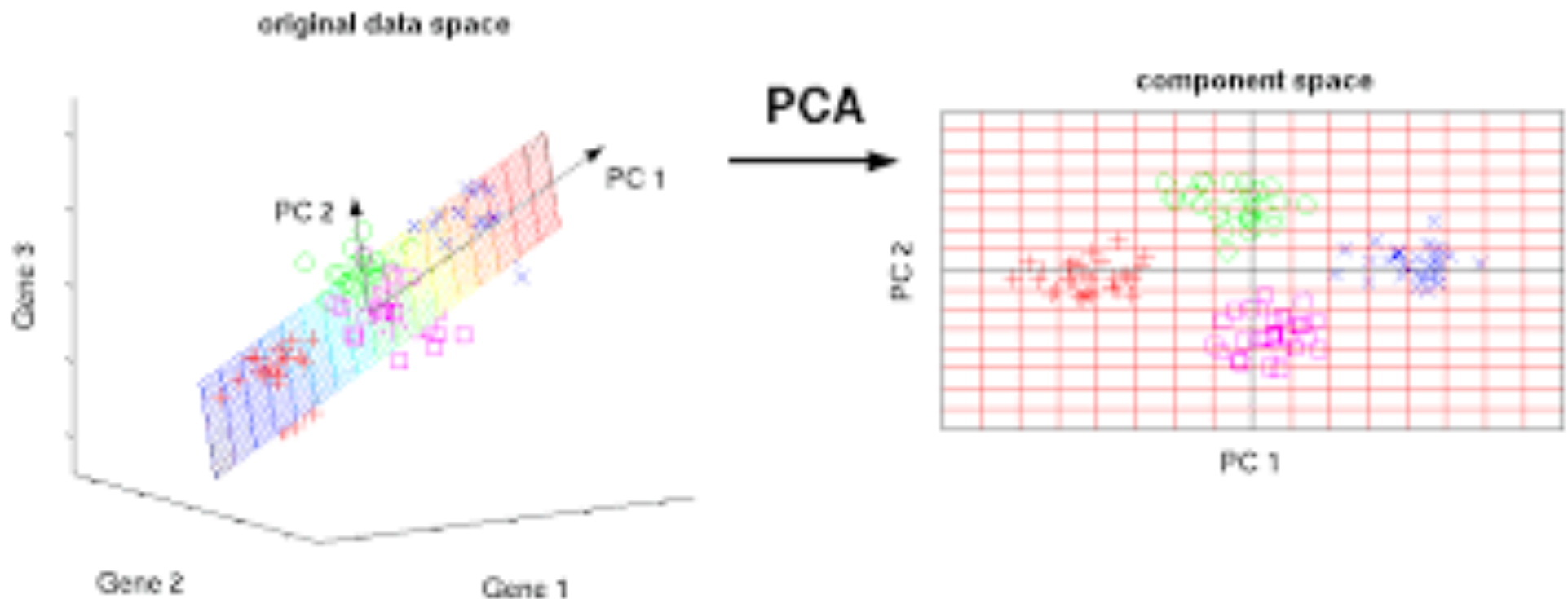
- Objective: find weights $\mathbf{W}_f$ and $\mathbf{W}_g$ that minimize reconstruction error

$$\min_{\mathbf{W}} \frac{1}{2} \sum_n \left\| \mathbf{W}_g \mathbf{W}_f \mathbf{x}_n - \mathbf{x}_n \right\|_2^2$$

- Algorithm: backpropagation
 - Gradient descent
- When using Euclidean norm (i.e., squared loss), solution is the same as principal component analysis (PCA)

Principal Component Analysis

- Hidden nodes: compressed representation

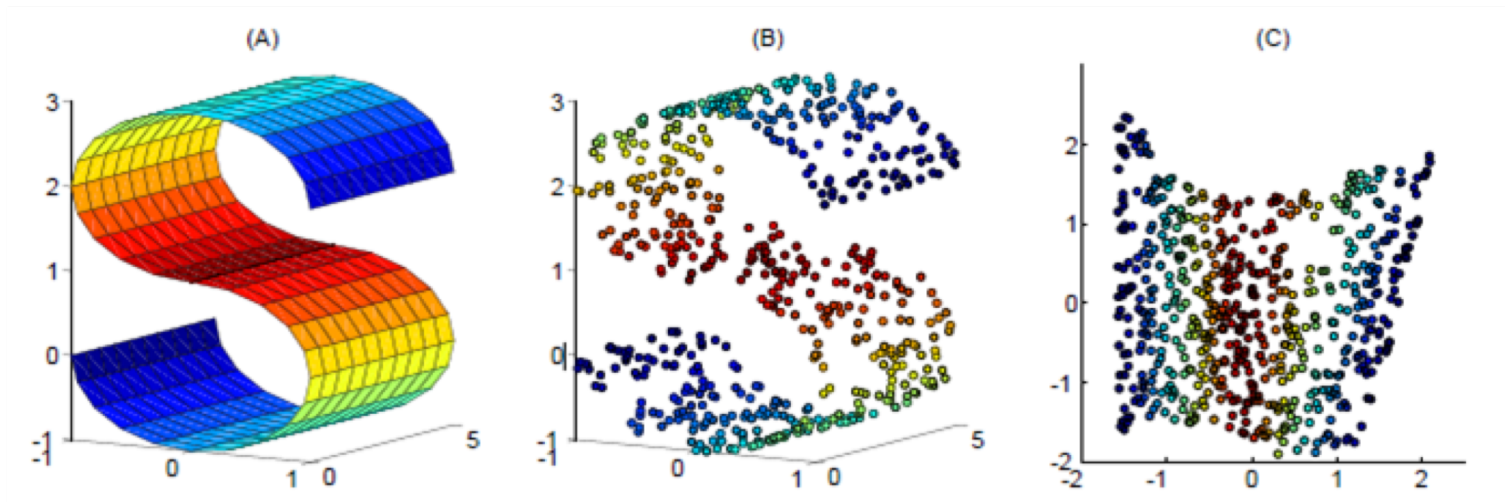


Nonlinear Autoencoder

- f and g are non-linear functions

$$\min_W \frac{1}{2} \sum_n \left\| g(f(\mathbf{x}_n; \mathbf{W}_f); \mathbf{W}_g) - \mathbf{x}_n \right\|_2^2$$

- Hidden nodes: non-linear manifold



Deep Autoencoders

- f and g often consist of multiple layers
- In theory, one hidden layer in f and g is sufficient to represent any possible compression
- Multiple hidden layers in f and g is often better

Sparse Representations

- When more hidden nodes than inputs, use regularization to constrain autoencoder
- Example: force hidden nodes to be sparse

$$\min_{\mathbf{W}} \frac{1}{2} \sum_n \left\| g(f(\mathbf{x}_n; \mathbf{W}_f); \mathbf{W}_g) - \mathbf{x}_n \right\|_2^2 + c \underbrace{nnz\left(f(\mathbf{x}_n; \mathbf{W}_f)\right)}_{\text{Sparse hidden nodes}}$$

where $nnz\left(f(\mathbf{x}_n; \mathbf{W}_f)\right)$ is the number of non-zero entries in the vector produced by f .

- Approximate objective: L1 regularization

$$\min_{\mathbf{W}} \frac{1}{2} \sum_n \left\| g(f(\mathbf{x}_n; \mathbf{W}_f); \mathbf{W}_g) - \mathbf{x}_n \right\|_2^2 + c \left\| f(\mathbf{x}_n; \mathbf{W}_f) \right\|_1$$

Denoising Autoencoder

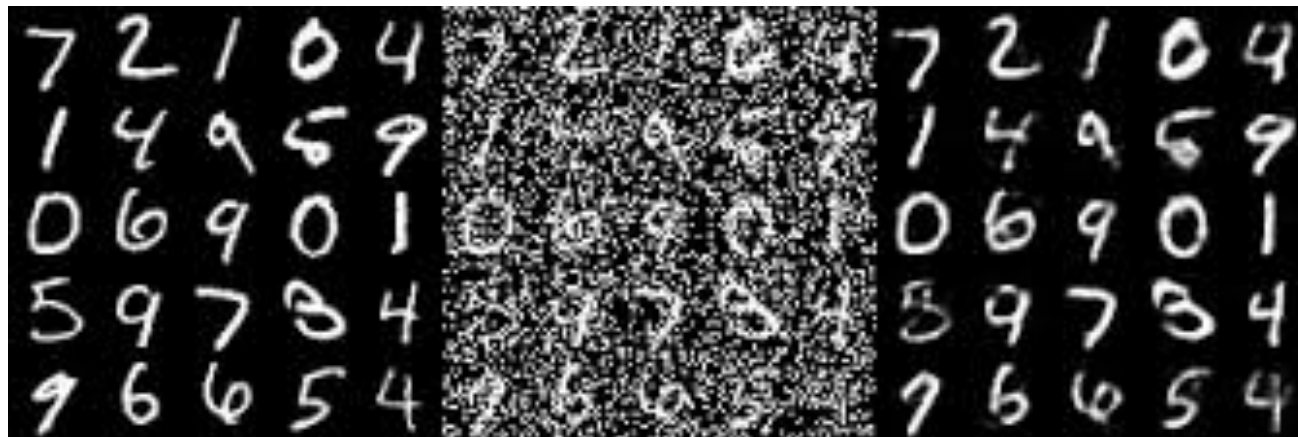
- Consider noisy version $\tilde{x}$ of the input x
- Data denoising

$$\min_W \frac{1}{2} \sum_n \left\| g(f(\tilde{x}_n; \mathbf{W}_f); \mathbf{W}_g) - \mathbf{x}_n \right\|_2^2 + c \left\| f(\tilde{x}_n; \mathbf{W}_f) \right\|_1$$

x

$\tilde{x}$

$g(f(\tilde{x}))$



original

perturbed

reconstructed

Probabilistic Autoencoder

- Let f and g represent conditional distributions

$$f: \Pr(\mathbf{h}|\mathbf{x}; \mathbf{W}_f) \quad \text{and} \quad g: \Pr(\mathbf{x}|\mathbf{h}; \mathbf{W}_g)$$

by using sigmoid, softmax or linear units at the hidden and output layers

- Picture

Generative Model

- Sample $\mathbf{h}$ from some distribution $\Pr(\mathbf{h})$
- Sample $\mathbf{x}$ from decoder $\Pr(\mathbf{x}|\mathbf{h}; \mathbf{W}_g)$

