Wednesday, November 26, 2014 3:21 PM

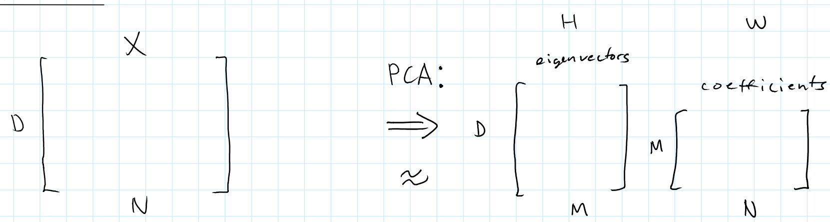
Last Time

factor analysis

- similar to probabilistic PCA

PCA noise: Gaussian, isotropic)
FA noise: different entries on diag (not isotropic)

On PCA



- finding these matrices that approximates the full data set
- difference ideally as close as possible

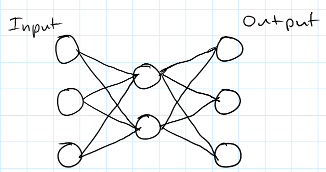
PCA: columns of H orthonormal

NMF: H, W no negative elements   Non-Negative Matrix Factorization   - requires iterative optimizer
why? - represent as additive sum of 'templates

Sparse Coding: coefficents have sparsity

Auto encoders

- neural network for unsupervised learning



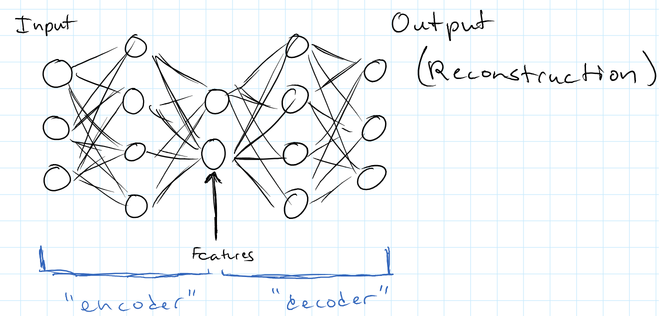
- same dimensionality output

"Information Bottleneck"

- 3d worth of info, 2d squeeze, recover as close as possible in output.

- multiple layers allows for a more rich representation
- Linear version (one hidden layer) performs PCA

Alternative Picture:



Specific Models:

- sparse auto encoder (sparse feature vectors)
  - penalty on hidden layer (ie. l_1 penalty)

$$\min_{\{w_j\}} \frac{1}{N} \sum_{n=1}^{N} \|x_n - \tilde{x}_n\|_2^2 + L_1 \text{ makes it sparse}$$

- De-noising auto encoder
  - corrupt input with noise
  - try to reconstruct original
  - features shouldn't be sensitive to small variations
  - robustness

Information Theory angle ("information bottleneck")

- imagine discrete: started with '3' bits, brought it to '2' bits, and back to '3'
- "Fit this thing into some size"
  - JPEG encoding
- "Lossy Compression"

Analogy to brain: only firing sparse subset of neurons for some given input.