Nonlinear Principle Component Analysis Using Autoassociative Neural Networks

By Mark A. Kramer

Presented by Rhema Linder

Intro

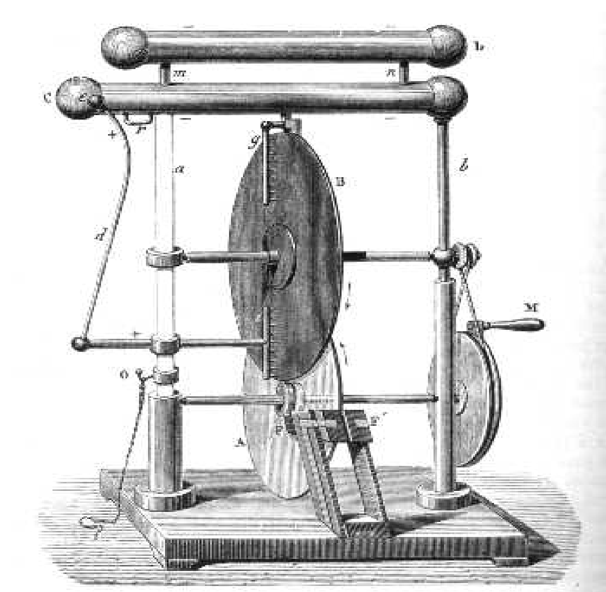
- Principle Component Analysis (PCA)
  - Reduces Dimensionality
  - Maps Features Into A Distribution
  - Visual Aid
  - Analysis
- Nonlinear Principle Component Analysis (NLPCA)
  - Solves nonlinear problems
  - Uses hidden layers NN architecture
  - Can fit to any non-random data
  - Many uses

Outline

- Stochastic Viewpoint
- PCA vs NLPCA
- Architecture
- Hidden Layers and Bottleneck
- Training
- Examples
    - Circle
    - Batch Reaction (Error Comparison)
- Summary

Stochastic Viewpoint

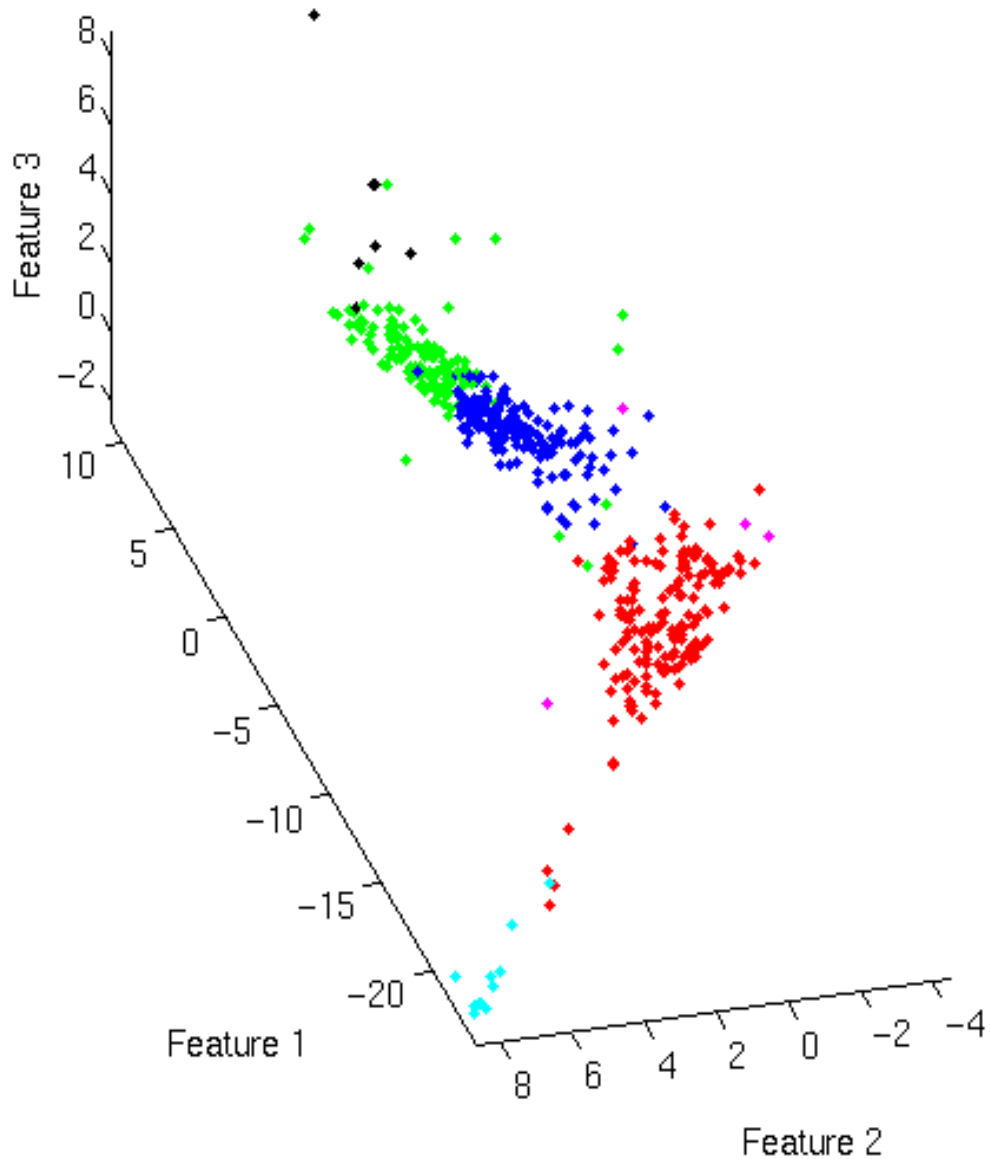
- 1991 Attitude of NNs
- No claims of representing biology
- Mathematically forces compact representations of data



Dimensionality

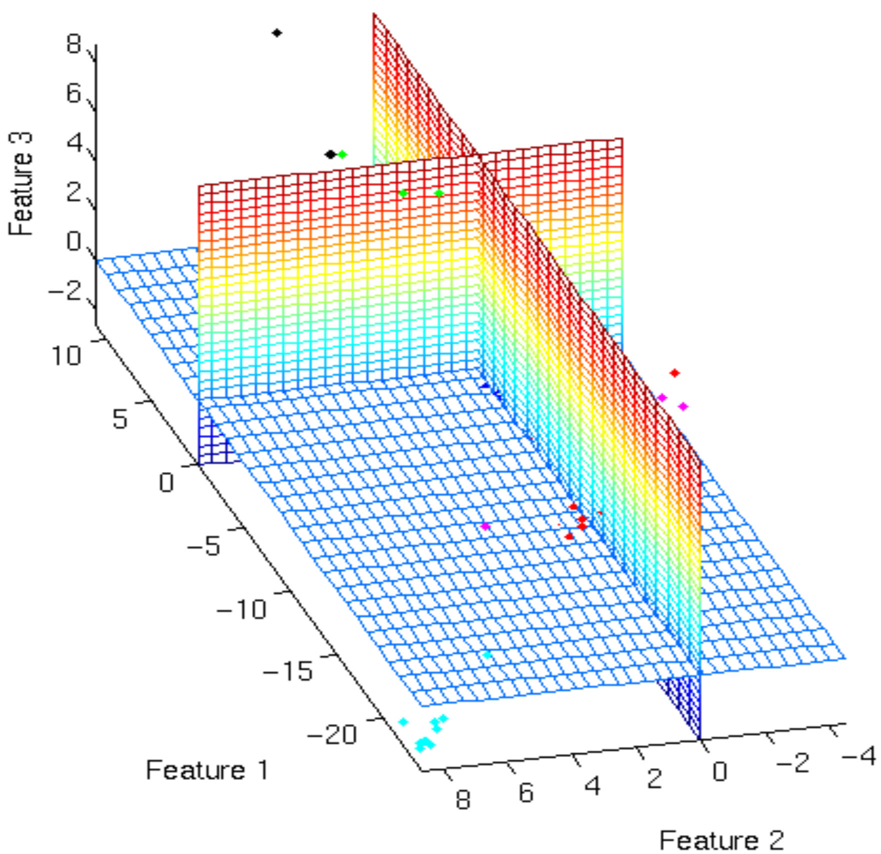
- Superflous Dimensionality
  - Mutilple measurments of the same thing
- Intrinsic Dimensionality
  - The number of independent variables underlying observation

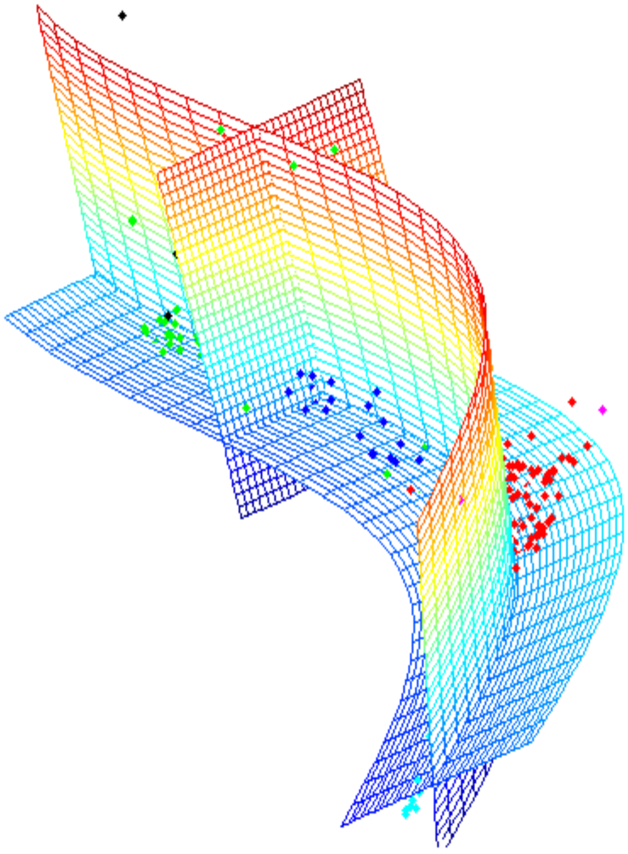
PCA vs NLPCA



PCA vs NLPCA

NLPCA involves nonlinear mappings between the original and reduced dimension spaces.





PCA VS NLPCA

- PCA can be solved simply:

Table of n observations by m variables.

$$Y = TP^T + E$$

Scores matrix.

Loadings matrix.

Residual.

$$P^T P = I$$

Solve eigan vectors

$$T_{row} = Y_{row} P$$

In feature space

$$Y'_{row} = T_{row} P^T$$

Back to original space.

PCA VS NLPCA

- NLPCA is analogous, but is more complex:
  - Likewise:

$$T_{row} = G_i(Y_{row})$$ — In feature space

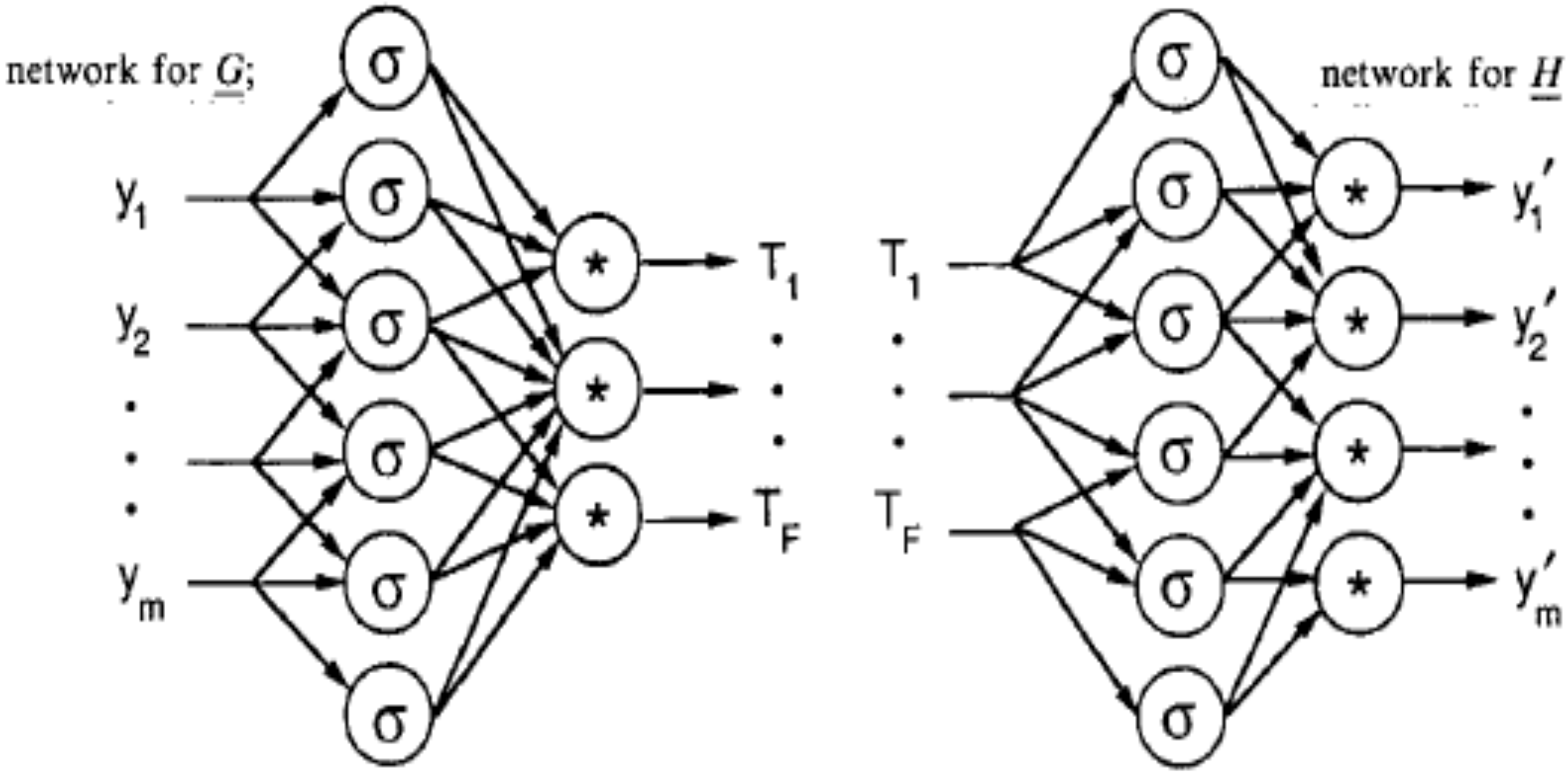
F individual nonlinear functions.

$$Y'_{row} = H_j(T_{row})$$ — In original space.

M individual nonlinear functions.

- G and H use sigmoidal functions

Architecture

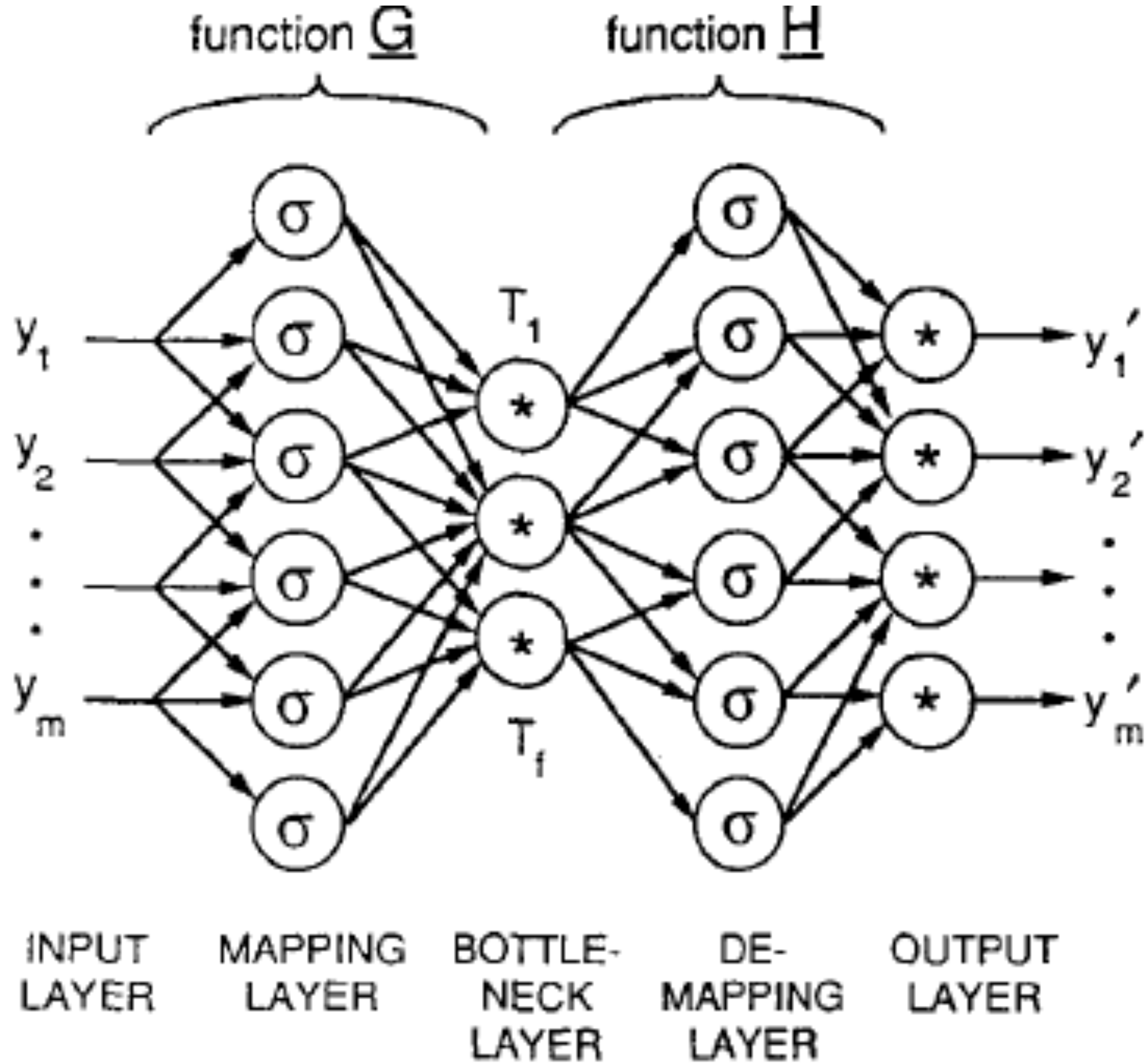


Networks implementing mapping and demaping functions.

Hidden Layers and Bottleneck

- Hidden layers necessary to represent non linear data.
- Supervised learning not tractable for these networks
- Because Y is the input and Y' is the output, we can combine the learning of these networks
    - Self-supervised backpropagation == autoassociation
- The bottleneck limits the dimensionality of the data and the layer does not need to be nonlinear
- The combined network cannot be converted into a two layer network.

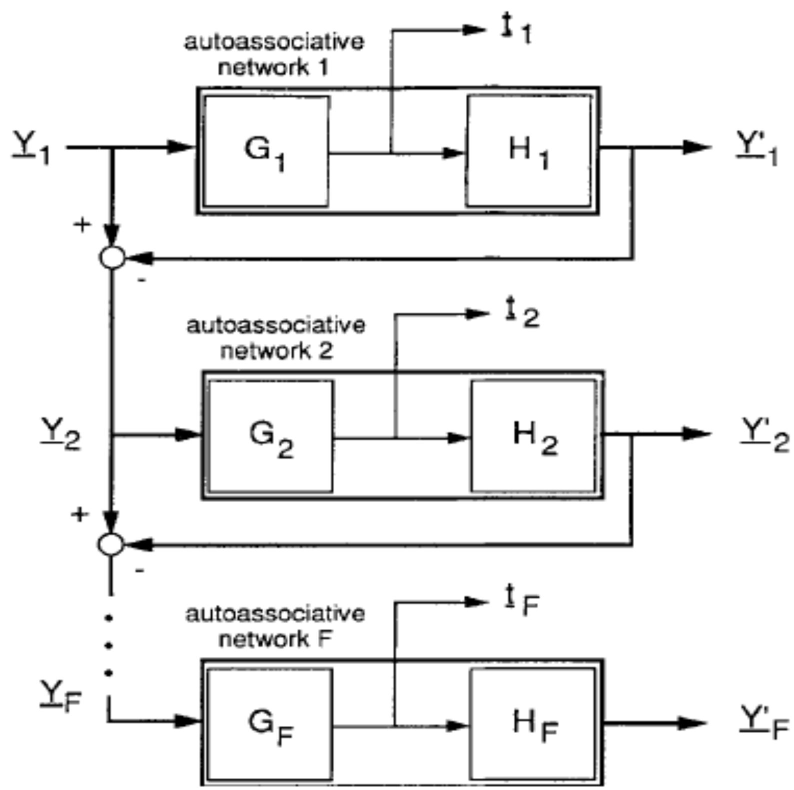
Training



Training is finished when sum of squared errors is minimized.

Sequential NLPCA

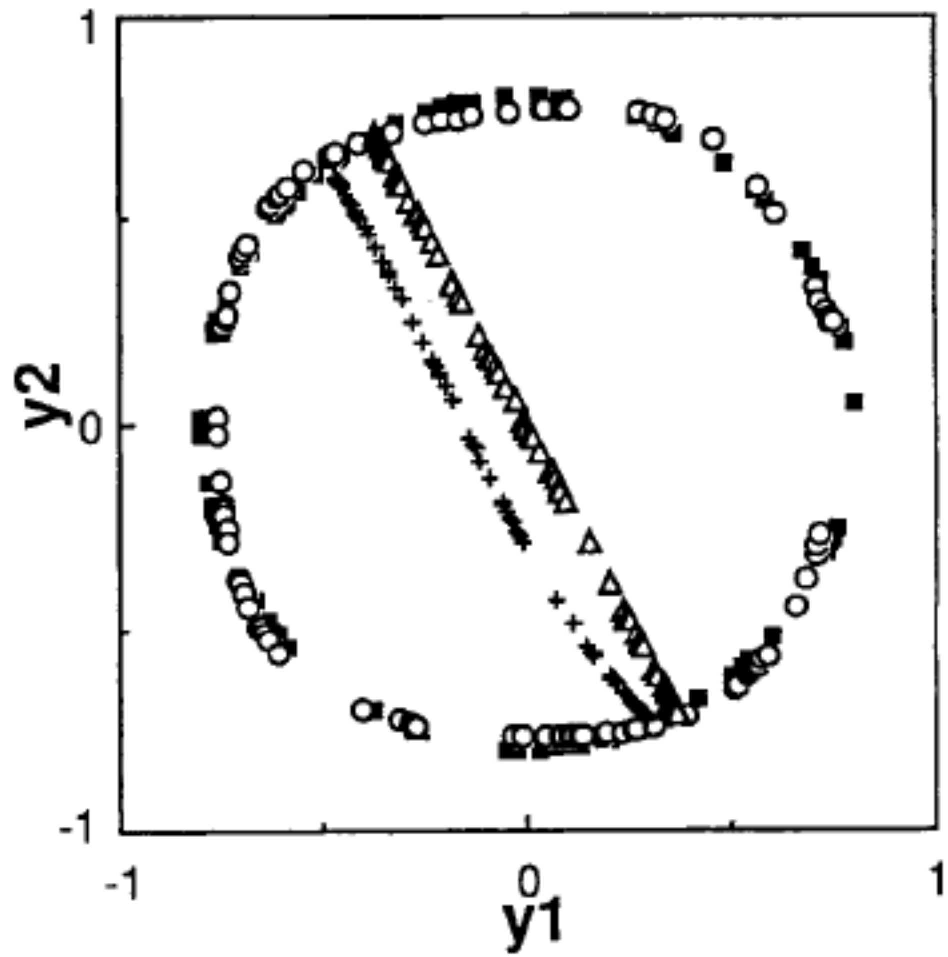
- Can rescale between steps
- Better at including more than just the primary factor



Sequential determination of nonlinear factors by training *F* networks with one bottleneck node each.

Example 1 - simple test

- NLPCA Outperforms PCA
- Both reduce to one factor but only one can be reconstructed



Reconstructed data from one factor

Original data (■), reconstruction using four mapping and de-mapping nodes (o), reconstruction with no mapping nodes (+), reconstruction using PCA (△).

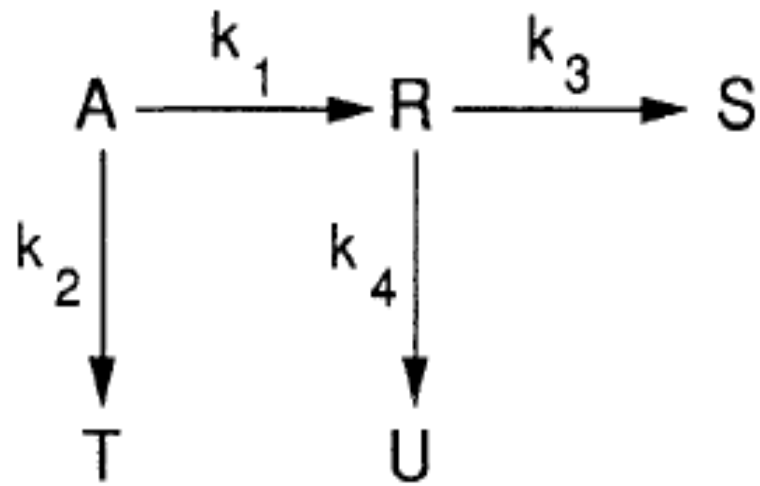
Example 1 - simple test

Table 1. Results of One-Factor Representations for Example 1

Technique	Adjust. Param.	Error E	FPE	AIC
PCA	2	27.8	0.0708	−2.65
ANN, no mapping layers	7	26.4	0.0708	−2.65
NLPCA, no. mapping nodes				
2	19	10.5	0.0318	−3.45
3	27	1.35	0.00444	−5.42
4	35	0.348	0.00124	−6.70
6	51	0.336	0.00142	−6.57
8	67	0.307	0.00154	−6.50
10	83	0.302	0.00183	−6.36

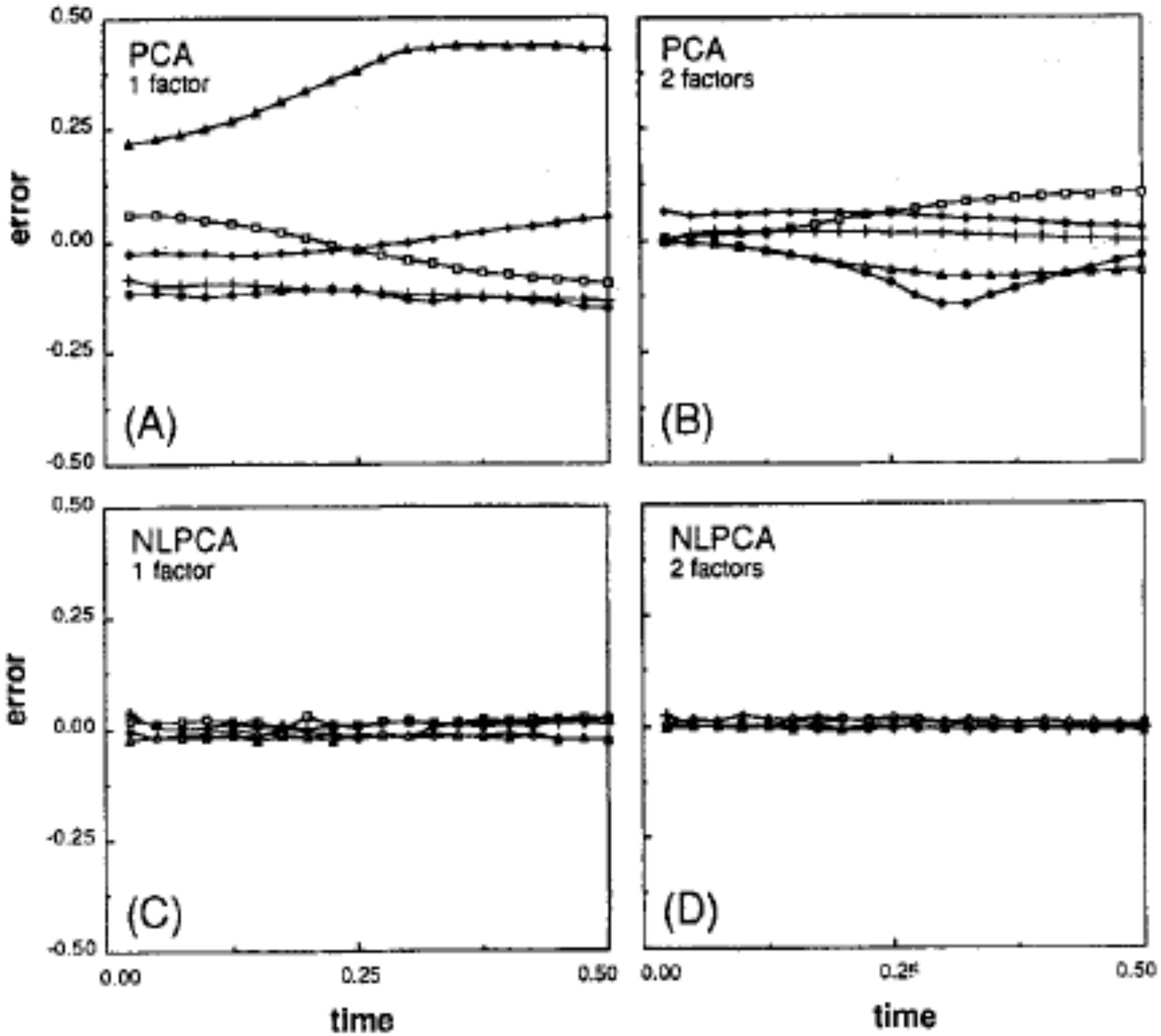
Example 2 - Batch Reactor

- High dimensional, (100 measurements) data with 25 batches.



Example 2 - Batch Reactor

.



Summary

- The NLPCA is can remove superfluous dimensionality.
- NLPCA uses a 3 layer NN.
- NLPCA can be applied to the same problems as PCA
  - Data reduction and visualization
  - Quality control
  - Principle component regression
  - et cetera
- NLPCA is generally better than PCA