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

- Superfluous Dimensionality
  - Multiple measurements 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:

\[ Y = TP^T + E \]


\[ P^T P = I \] Solve eigen vectors

\[ T = YP \] In feature space

\[ Y' = TP^T \] Back to original space.
PCA VS NLPCA

• NLPCA is analogous, but is more complex:
  o Likewise:

\[
T = G(Y) \quad \text{In feature space}
\]

\[
Y' = H(T) \quad \text{In original space}
\]

• G and H use sigmoidal functions
Architecture

Networks implementing mapping and demapping 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 demapping nodes (○), reconstruction with no mapping nodes (+), reconstruction using PCA (△).
Example 1 - simple test

<table>
<thead>
<tr>
<th>Technique</th>
<th>Adjust. Param.</th>
<th>Error $E$</th>
<th>FPE</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA</td>
<td>2</td>
<td>27.8</td>
<td>0.0708</td>
<td>-2.65</td>
</tr>
<tr>
<td>ANN, no mapping layers</td>
<td>7</td>
<td>26.4</td>
<td>0.0708</td>
<td>-2.65</td>
</tr>
<tr>
<td>NLPCA, no. mapping nodes</td>
<td>2</td>
<td>19</td>
<td>10.5</td>
<td>-3.45</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>27</td>
<td>1.35</td>
<td>-5.42</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>35</td>
<td>0.348</td>
<td>-6.70</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>51</td>
<td>0.336</td>
<td>-6.57</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>67</td>
<td>0.307</td>
<td>-6.50</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>83</td>
<td>0.302</td>
<td>-6.36</td>
</tr>
</tbody>
</table>
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