Computational Intelligence: Methods and Applications

Lecture 8

Projection Pursuit &
Independent Component Analysis

Włodzisław Duch
SCE, NTU, Singapore
Google: Duch

Exploratory Projection Pursuit (PP)

PCA and FDA are linear, PP may be linear or non-linear.
Find interesting "criterion of fit", or "figure of merit" function, that allows for low-dim (usually 2D or 3D) projection.

$$\mathbf{Y}^{(j)\mathrm{T}} = \left(Y_1^{(j)}, Y_2^{(j)}\right) = f\left(\mathbf{X}^{(j)}; \mathbf{W}\right);$$ General transformation with parameters W.

$$I(\mathbf{Y};\mathbf{W}) = I\left(f\left(\mathbf{X};\mathbf{W}\right)\right)$$ Index of "interestingness"

Interesting indices may use *a priori* knowledge about the problem:

1. mean nearest neighbor distance – increase clustering of $Y^{(j)}$
2. maximize mutual information between classes and features
3. find projection that have non-Gaussian distributions.

The last index does not use *a priori* knowledge; it leads to the Independent Component Analysis (ICA), unsupervised method.
ICA features are not only uncorrelated, but also independent.

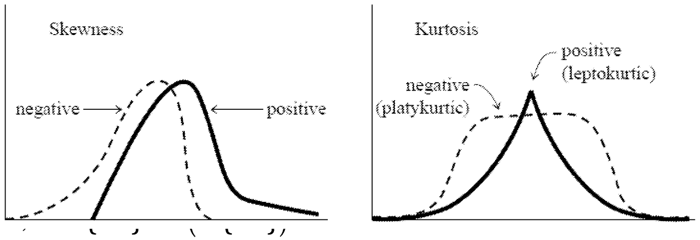
Kurtosis

ICA is a special version of PP, recently very popular.

Gaussian distributions of variable Y are characterized by 2 parameters:

mean value: $\bar{Y} = E\{Y\}$

The … Y).
One…
4-th … s
"cor…
κ_4 (…



Super-Gaussian distribution: long tail, peak at zero, $\kappa_4(Y)>0$, like binary image data.
sub-Gaussian distribution is more flat and has $\kappa_4(Y)<0$, like speech signal data.

Find interesting direction looking for $\max_W |\kappa_4(Y(W))|$

Correlation and independence

Variables Y_i are statistically independent if their joint probability distribution is a product of probabilities for all variables:

$$p\left(Y_1, Y_2 \cdots Y_n\right) = \prod_{i=1}^{n} p_i\left(Y_i\right)$$

Features Y_i, Y_j are uncorrelated if covariance is diagonal, or:

$$E\left\{Y_i Y_j\right\} = E\left\{Y_i\right\} E\left\{Y_j\right\}$$

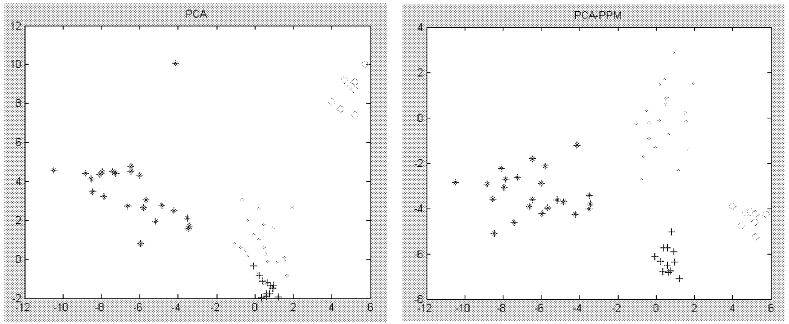
Uncorrelated features are orthogonal, but may have higher-order dependencies, while any functions of independent features Y_i, Y_j

$$E\left\{f_1\left(Y_i\right) f_2\left(Y_j\right)\right\} = E\left\{f_1\left(Y_i\right)\right\} E\left\{f_2\left(Y_j\right)\right\}$$

This is much stronger condition than correlation; in particular the functions may be powers of variables; any non-Gaussian distribution after PCA transformation will still have some feature dependencies.

PP/ICA example

Example: PCA and PP based on maximal kurtosis: note nice separation of the blue class.



Some remarks

- Many algorithms for exploratory PP and ICA methods exist.
- PP is used for visualization, dimensionality reduction & regression.
- Nonlinear projections are frequently considered, but solutions are more numerically intensive.
- PCA may also be viewed as PP, maximizing (for standard. data):

$$\mathbf{W}^{(1)} = \arg\max_{\|\mathbf{W}\|=1} E\left\{\left(\mathbf{W}^{\mathrm{T}}\mathbf{X}\right)^2\right\}$$

Index I(Y;W) is based here on maximum variance.

Other components are found in the space orthogonal to W_1

$$\mathbf{W}^{(k)} = \arg\max_{\|\mathbf{W}\|=1} E\left\{\left(\mathbf{W}^{\mathrm{T}}\left(\left(\mathbf{I}-\sum_{i=1}^{k-1}\mathbf{W}^{(i)}\mathbf{W}^{\mathrm{T}(i)}\right)\mathbf{X}\right)\right)^2\right\}$$

Same index is used, with projection on space orthogonal to k-1 PCs.

PP/ICA description: Chap. 14.6, Friedman, Hastie, Tibshirani.

ICA demos

- ICA has many applications in signal and image analysis.
- Finding independent signal sources allows for separation of signals from different sources, removal of noise or artifacts.

Observations X are a linear mixture W of unknown sources Y

$$\mathbf{X} = \mathbf{W}^{\mathrm{T}}\mathbf{Y}$$

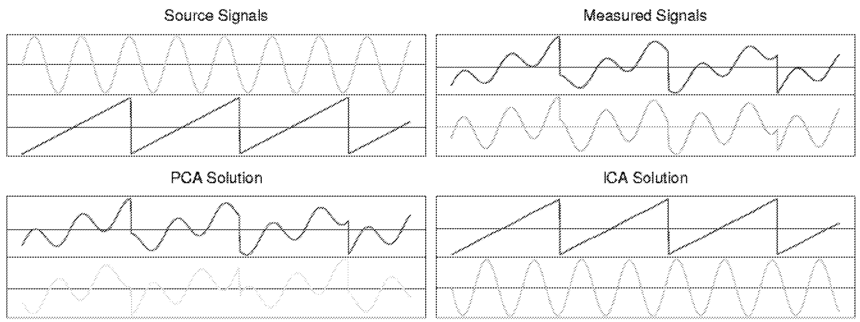
Both W and Y are unknown! This is a blind separation problem. How can they be found?

If Y are Independent Components and W linear mixing the problem is similar to PCA, only the criterion function is different.

Play with ICA-Lab PCA/ICA Matlab software for signal/image analysis: http://www.bsp.brain.riken.go.jp/page7.html

ICA examples

- Mixing simple signals: sinus + chainsaw.
  Vectors X = samples of signals in some time window



From: Chap. 14.6, Friedman, Hastie, Tibshirani: The elements of statistical learning.

ICA demo: images & audio

Play with ICA-Lab PCA/ICA Matlab software for signal/image analysis from Cichocki's lab,
 http://www.bsp.brain.riken.go.jp/page7.html

X space for images:

take intensity of all pixels ⇔ one vector per image, or
take smaller patches (ex: 64x64), increasing # vectors

- 5 images: originals, mixed, convergence of ICA iterations

X space for signals:

sample the signal for some time Δt

- 10 songs: mixed samples and separated samples

Good survey paper on ICA is at:
 http://www.cis.hut.fi/aapo/papers/NCS99web/

Further reading

Many other visualization and dimensionality reduction methods exit.
See the links here:

http://www.ntu.edu.sg/home/aswduch/CI.html#vis
http://www.ntu.edu.sg/home/aswduch/software.html#Visual
Principal curves Web page
http://www.iro.umontreal.ca/~kegl/research/pcurves/
Good page with research papers on manifold learning:
http://www.cse.msu.edu/~lawhiu/manifold/

A. Webb, Chapter 6.3 on projection pursuit, chap. 9.3 on PCA
Duda/Hart/Stork, chap. 3.8 on PCA and FDA
Now we shall turn to non-linear methods inspired by the approach that is used by our brains.

Self-organization

PCA, FDA, ICA, PP are all inspired by statistics, although some neural-inspired methods have been proposed to find interesting solutions, especially for non-linear PP versions.

- Brains learn to discover the structure of signals: visual, tactile, olfactory, auditory (speech and sounds).
- This is a good example of unsupervised learning: spontaneous development of feature detectors, compressing internal information that is needed to model environmental states (inputs).
- Some simple stimuli lead to complex behavioral patterns in animals; brains use specialized microcircuits to derive vital information from signals – for example, amygdala nuclei in rats sensitive to ultrasound signals signifying "cat around".

Models of self-organizaiton

SOM or SOFM (Self-Organized Feature Mapping) – self-organizing feature map, one of the simplest models.

How can such maps develop spontaneously?
Local neural connections: neurons interact strongly with those nearby, but weakly with those that are far (in addition inhibiting some intermediate neurons).

History:
von der Malsburg and Willshaw (1976), competitive learning, Hebb mechanisms, „Mexican hat" interactions, models of visual systems.
Amari (1980) – models of continuous neural tissue.
Kohonen (1981) - simplification, no inhibition; leaving two essential factors: competition and cooperation.