Dimensionality reduction

December 7, 2016

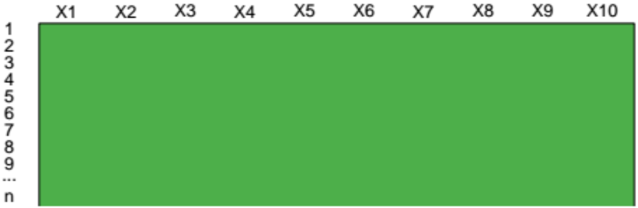
Outline

Principal Component Analysis

- What is it?
- Extensions
  - Sparse PCA
  - Simultaneous Component Analysis (SCA)
  - Independent Component Analysis (ICA)

Unsupervised learning: Dimension reduction

- Goal: Estimate a probability distribution $P(X)$
- X mostely multivaiate, possibly **huge** $\boldsymbol{p}$
- Different characteristics of interest
- Trying to find a density function $\hat{P}(X)$ that is close to $P(X)$



PCA: What is it?

Goal: Reduce matrix X of dimension p to alternative matrix T of dimension r

- $p =$ number of variables
- $r =$ number of components
- with $r < p$

PCA: What is it?

Dimension reduction by taking **orthogonal linear combinations** of the original variables such that the new dimensions contain as much variance as possible

$$T = XP$$

- T = standardised component scores (n x r)
- X = original standardised data matrix (n x p)
- P = component loadings (p x r)

PCA: What is it?

Dimension reduction by taking linear combinations of the original variables

$$T = XP$$

1. Constrain variance to 1: $\sum P^2 = 1$
2. Each linear combination is orthogonal with the others: $T_j^T T_g = 0$
3. Each linear combination explains as much variance as possible: $var(T_i) > var(T_{i+1})$

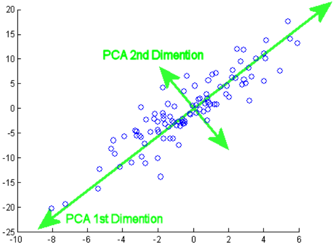
PCA: What is it?

Dimension reduction by taking **orthogonal linear combinations** of the original variables such that the new dimensions contain as much variance as possible

$$T = XP$$

- $T =$ standardised component scores (n x r)
- $X =$ original standardised data matrix (n x p)
- $P =$ component loadings (p x r)
  - combination of eigenvectors and eigenvalues

PCA: What is it?



- Eigenvectors: direction of maximal variance
- Eigenvalues: scale of maximal variance

PCA: What is it?

Singular value decomposition (SVD)

$$X = USV^T$$

- X is the original n x p data matrix,
- columns of n x n matrix U contains the left-singular vectors,
- columns of p x p matrix V contain the right-singular vectors,
- S is a diagonal n x p matrix that contains the singular values in descending order.

PCA: What is it?

Singular value decomposition (SVD)

$$X = USV^T$$

$$T = XP$$

$$T = \sqrt{n-1}U \; (= standardised \; scores)$$

$$P = \frac{SV^T}{\sqrt{n-1}} \; (= loadings)$$

PCA: What is it?

Singular value decomposition (SVD)

$$X = USV^T$$

$$US = \textit{principal scores}$$

$$V^T = \textit{principal directions } (= \textit{eigenvectors})$$

PCA: What is it?

Least squares minimisation problem

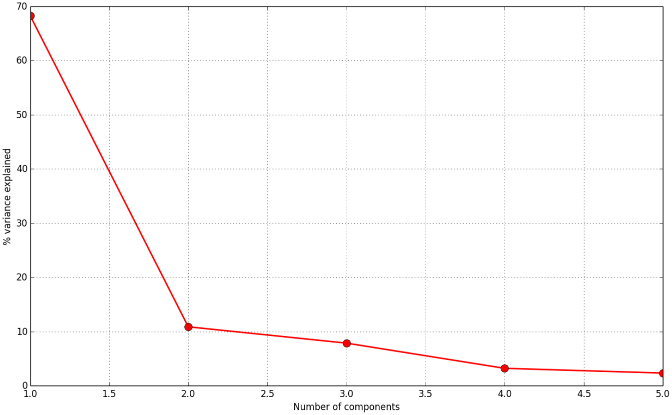
$$(\hat{T}, \hat{P}) = \underset{T,P}{\text{argmin}} ||X - TP^T||_2^2$$

PCA: What is it?

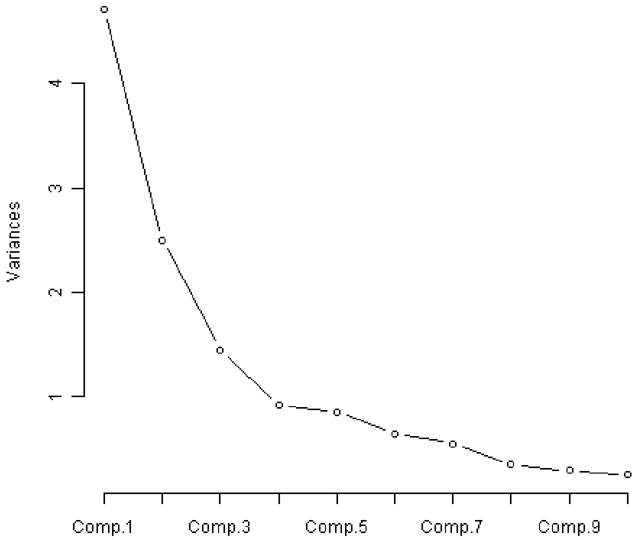
Based on some cut-off, take the first r-components

- Proportion variance explained: Choose all components until they cumulative explain certain amount of variance
- Eigenvalue criterion: Choose all components with eigenvalues higher than 1
- Scree plot: Look at the graph of the components and their eigenvalues. Choose all components before the ‘elbow’

PCA: What is it?



PCA: What is it?



Extensions: Sparse PCA

- Manually (e.g. set all loadings <0.3 to 0)
- Penalty, such as lasso

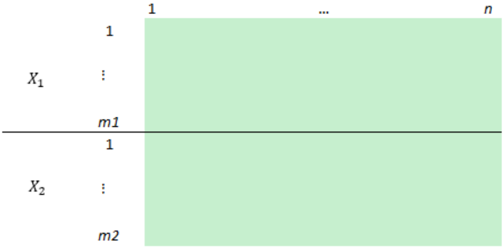
$$(\hat{T}, \hat{P}) = \underset{T,P}{\text{argmin}} ||X - TP^T||_2^2 + \lambda_l ||P||_1$$

Extension: Simultaneous Component Analysis (SCA)

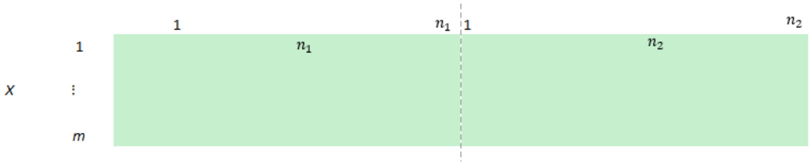
Integrate over multiple data blocks K with either

- common subjects (T) or
- common variables (P)

Extension: Simultaneous Component Analysis (SCA)



Extension: Simultaneous Component Analysis (SCA)



Extension: Simultaneous Component Analysis (SCA)

Integrate over multiple data blocks K with common subjects (T)

$$(\hat{T}, \hat{P}_k) = \underset{T, P_k}{\operatorname{argmin}} ||X_k - TP_k^T||_2^2$$

Extension: Sparse SCA

$$(\hat{T}, \hat{P}_k) = \underset{T, P_k}{\operatorname{argmin}} ||X_k - TP_k^T||_2^2 +$$

$$\sum(\lambda_g \sqrt{J_k}||P_k||_2 \ + \ \lambda_e||P_k||_{1,2}$$

- λ_g group lasso penatly: Selecting groups
- λ_e elitist lasso penalty: Selecting variables within groups

Extension: Independent Component Analysis (ICA)

In PCA, you maximise a second-order moment (variance).
In ICA, you maximise higher order moment.

Extension: Independent Component Analysis (ICA)

Standardised data X is a linear mixture of **independent, non-Gaussian** source signals

$$X = AS^T$$

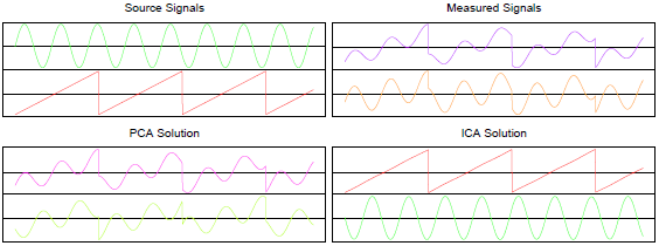
- X = Data
- A = Mixing weights
- S = Independent components (= sources)

Extension: Independent Component Analysis (ICA)

Maximise **independence** of components

- Minimise mutual information (maximum entropy)
- Maximise non-Guassianity (kurtosis)

Extension: Independent Component Analysis (ICA)



Extension: Independent Component Analysis (ICA)

Use ICA when data are not

- Guassian
- stationary
- linear

Extension: Independent Component Analysis (ICA)

ICA **cannot**

- identify the number of source signals
- uniquely order the source signals
- properly scale source signals

Often PCA as preprocessing step

Rotation: What is it about?

- Goal: Make PCA results more interpretable
- How: Rotate T and P as to make P as sparse as possible.

Rotated loadings do not respond to orthogonal eigenvectors