Analysis of nonnegative data objects

Lingsong Zhang

Nov. 13, 2012

Outline

1. Background and Introduction
2. PCA, SVD and NMF
3. Nested Cone Analysis Method

Object-oriented data analysis

- *Unit of interests*
  - Introductional Statistics: Number
  - Multivariate Statistics: Vector
  - Functional Data Analysis: Function, Curve
  - Object-Oriented Data Analysis: data objects
- Important reference on FDA and OODA
  - FDA: Ramsay and Silverman (2002, 2005), Ferraty and Vieu (2006), $\cdots$.
  - OODA: Wang and Marron (2007), Aydin et al (2010), Shen et al (2012+).
- Central tool: Principal Component Analysis

Some Challenges

- Sometimes, the notion of mean and variability is hard to define
- Collected data may have some intrinsic structure
- Usually the number of parameters are much larger than the sample size
- Data may not lie in Euclidean Space

In this talk, we will focus on a special data: nonnegative data objects

Nonnegativity : Examples

- Image analysis: e.g. gray scale images: values are between 0 and 255.
- Branch length representation of tree objects: lengths of the branches are nonnegative.
- Proteomics and chemometrics: spectrum are nonnegative values.
- Monotonic functional data: first derivative are nonnegative values.

Related work

- Data transformation: $\log(X)$, $\log(1+X)$
- Nonnegative least squares $Y = \mathbf{X}\beta + \varepsilon$, where $\beta \geq 0$. (Lawson and Hanson, 1995)
- Nonnegative matrix factorization (Paatero and Tapper, 1994, Lee and Seung, 1999)
- Nonnegative independent component analysis (Plumbley, 2003)
- Nonnegative Garrotte method (Breiman, 1995)
- Nonnegative time series (Hutton, 1990, Bougerol and Picard, 1992)

Constrained Statistical Analysis

- Shape constraints : smoothness, monotonicity, convexity, log-concavity
- Model complexity constraints: sparsity and false discovery rate
- Group constrains: group variable selection, gene set analysis

$\vdots$

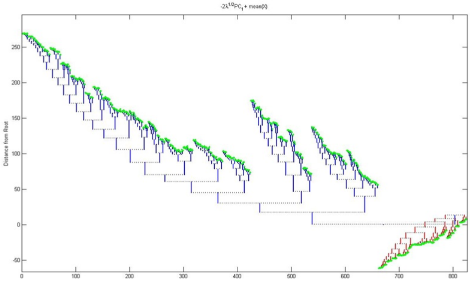
Motivating example - analysis of tree objects

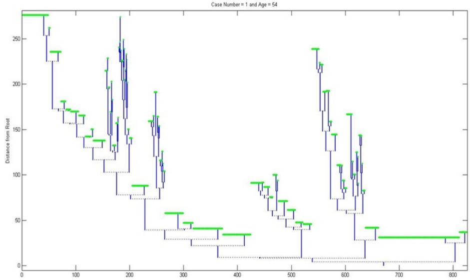
Early tutorial from J. S. Marron and Dan Shen

- Dyck Path and Branch Length Presentation provide a nice functional view (or multivariate view) of the trees.
- Every tree corresponds to a data point in the first quadrant.
- PCA method provides some interesting projections for exploration

Motivating example

- However, examples show that PCA projections may leave the first quadrant. This makes the projections less interpretable.
- PCA projections is not sparse (no flat parts in the tree)
- Nonnegative matrix factorization may solve these issues.





Dyck Path vs. Branch Length

- For either method, a tree corresponds to a data point in the first quadrant.
- Dyck Path Presentation: a data point in the first quadrant may not correspond to a tree
    - The curve of the Dyck Path is more structured.
- Branch Length Presentation: $1-1$ mapping between trees and data points in the first quadrant.
- Our analysis uses the Branch Length Presentation.

PCA quick review

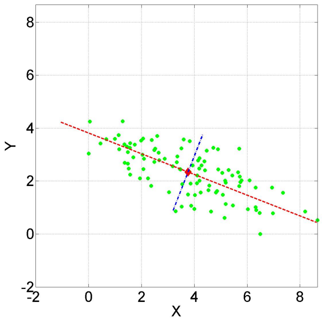
- Let X_1, X_2, $\cdots$, X_d be d random variables, PCA is to find a linear combination of X's that has the largest variability.

$$\text{maximize} \quad \text{Var}(\alpha_1 X_1 + \cdots + \alpha_d X_d),$$

  where $\alpha_1^2 + \cdots + \alpha_d^2 = 1$.
- After identify the first set α's, we will continue to find the linear combination (β's) of X's that has the largest variability. In addition, α and β are orthogonal to each other.
- It turns out this corresponds to the eigen-decomposition of the covariance of X's.

PCA quick review

- Note that α, β are the eigenvectors of Σ, the covariance
- For a data matrix $\mathbf{X}$, estimate $\widehat{\Sigma}$ and its eigen-decomposition (λ_i, ξ_i) $(i = 1, \cdots, k)$, where $\lambda_1 \geq \lambda_2 \geq \cdots$. ξ_i's are the estimators for α, β's.
- This can be done by an SVD of $\mathbf{X}$.



SVD quick review

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{pmatrix} \qquad (\mathrm{rank}(X) = r)$$

SVD quick review

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{pmatrix} = USV^T \qquad (\mathrm{rank}(X) = r)$$

SVD quick review

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{pmatrix} = USV^T \qquad (\text{rank}(X) = r)$$

$$= \begin{pmatrix} u_{11} & u_{12} & \cdots & u_{1r} \\ u_{21} & u_{22} & \cdots & u_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ u_{m1} & u_{m2} & \cdots & u_{mr} \end{pmatrix} \begin{pmatrix} s_1 & 0 & \cdots & 0 \\ 0 & s_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s_r \end{pmatrix} \begin{pmatrix} v_{11} & v_{12} & \cdots & v_{1r} \\ v_{21} & v_{22} & \cdots & v_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ v_{n1} & v_{n2} & \cdots & v_{nr} \end{pmatrix}$$

SVD quick review

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{pmatrix} = USV^T \qquad (\mathrm{rank}(X) = r)$$

$$= \begin{pmatrix} u_{11} & u_{12} & \cdots & u_{1r} \\ u_{21} & u_{22} & \cdots & u_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ u_{m1} & u_{m2} & \cdots & u_{mr} \end{pmatrix} \begin{pmatrix} s_1 & 0 & \cdots & 0 \\ 0 & s_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s_r \end{pmatrix} \begin{pmatrix} v_{11} & v_{12} & \cdots & v_{1r} \\ v_{21} & v_{22} & \cdots & v_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ v_{n1} & v_{n2} & \cdots & v_{nr} \end{pmatrix}$$

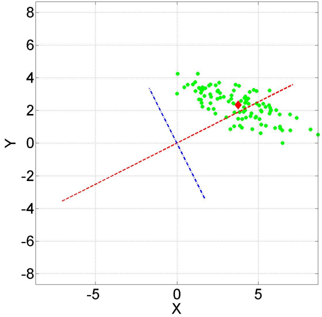
$$= s_1 \mathbf{u}_1 \mathbf{v}_1^T + s_2 \mathbf{u}_2 \mathbf{v}_2{}^T + \cdots + s_r \mathbf{u}_r \mathbf{v}_r^T$$

s's - *singular vector*; $\mathbf{u}$'s - *singular column*; $\mathbf{v}$'s - *singular row*;

$s_i \mathbf{u}_i \mathbf{v}_i^T$ - *SVD component*

SVD quick review

- SVD is a mathematical tool. It can be used to estimate PCA
- SVD can be viewed as a non-central PCA
- Minimize a noncentral second moments.
- Sometimes provides better interpretation
- Especially suitable for two-way data sets



Nonnegative Matrix Factorization

- The Nonnegative Matrix Factorization (NMF) of a data matrix $X = (x_{ij})_{p\times n}$ (where $x_{ij} \geq 0$) is defined as

$$X \approx W_{p\times k} H_{k\times n},$$

  where entries in W and H are nonnegative. Usually rows of H have norm 1.
  See Lee and Seung (1999, 2001), Berry et al (2007).

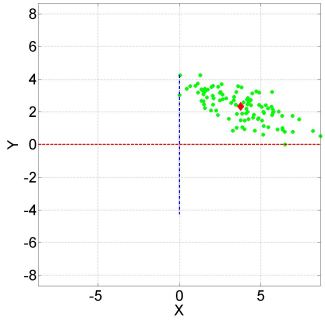
- Very similar to PCA or SVD, except:
  - entries of the resulting matrices are nonnegative
  - columns of W, rows of H are not orthogonal to each other
  - W and H usually are sparse

Interpretation of W and H

- Columns of W forms scaled (non-orthogonal) basis for the trees
    - Similar to the principal component (directions)
- Rows of H are scaled projections of the subjects
    - Similar to the projections to the principal component directions.

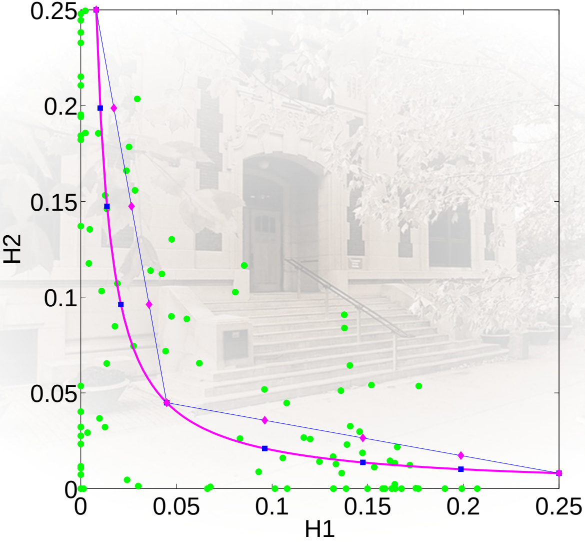
NMF quick review

- Very similar to SVD, but the directions and coefficients are nonnegative
- Typically find the extreme rays of a cone
- Better interpretation than SVD.

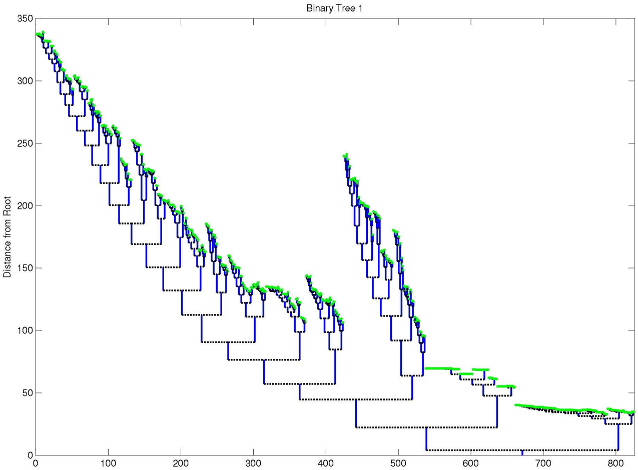


First exploration: Scatter plot of $H1$ vs. $H2$

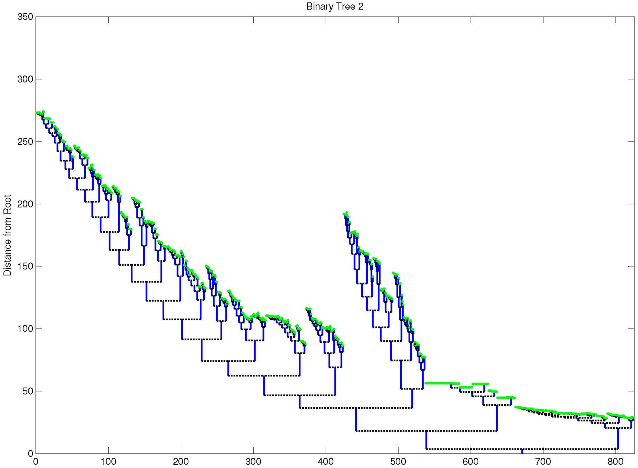
Projections on the first component vs. the second component



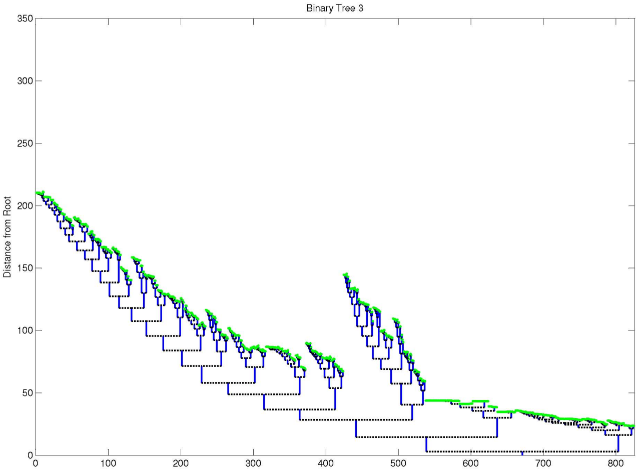
March along the “principal hyperbola” - Page 1



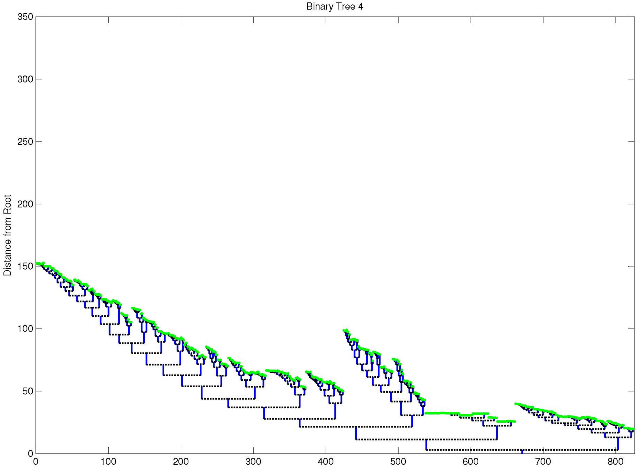
March along the “principal hyperbola” - Page 2



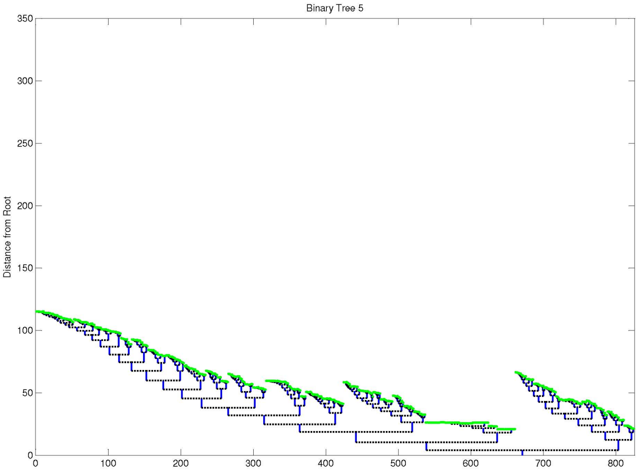
March along the “principal hyperbola” - Page 3



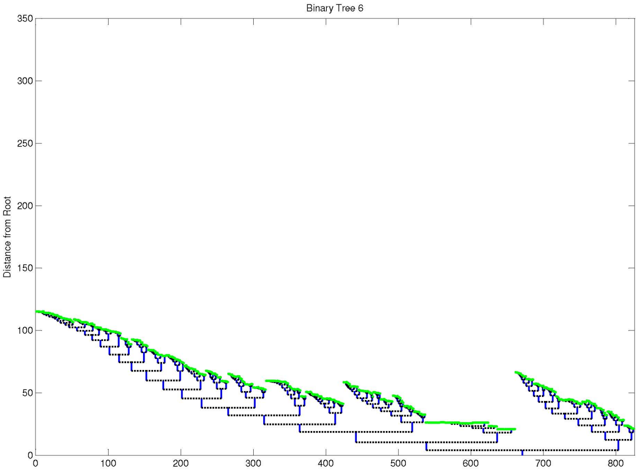
March along the “principal hyperbola” - Page 4



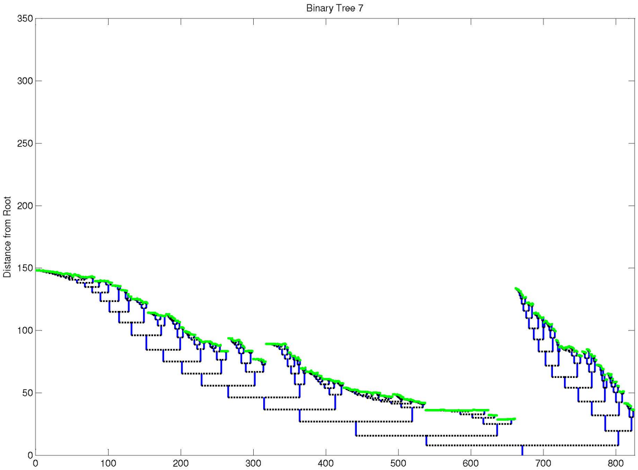
March along the “principal hyperbola” - Page 5



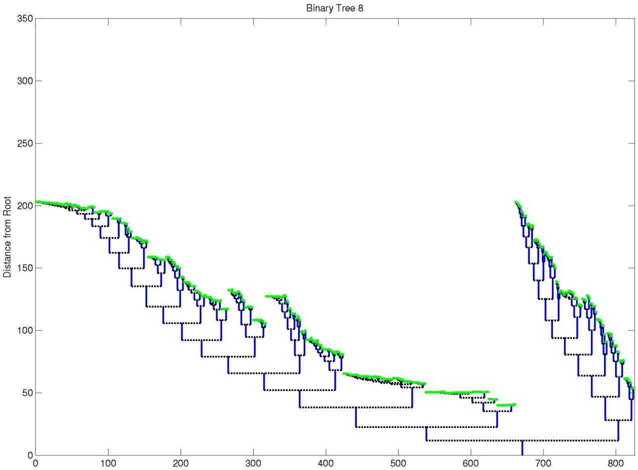
March along the “principal hyperbola” - Page 6



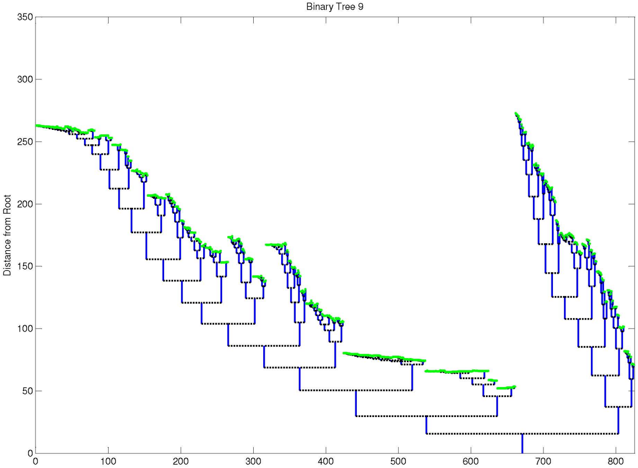
March along the “principal hyperbola” - Page 7



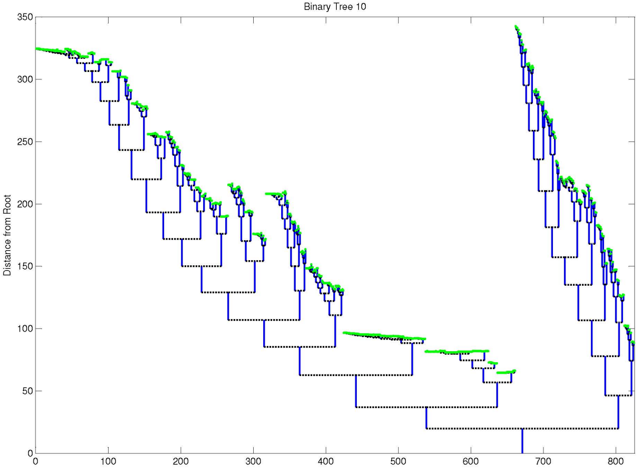
March along the “principal hyperbola” - Page 8



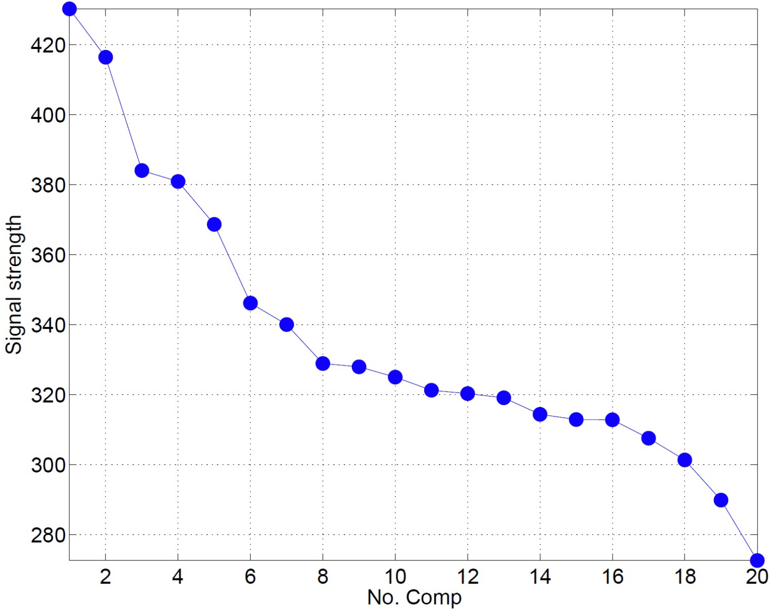
March along the “principal hyperbola” - Page 9



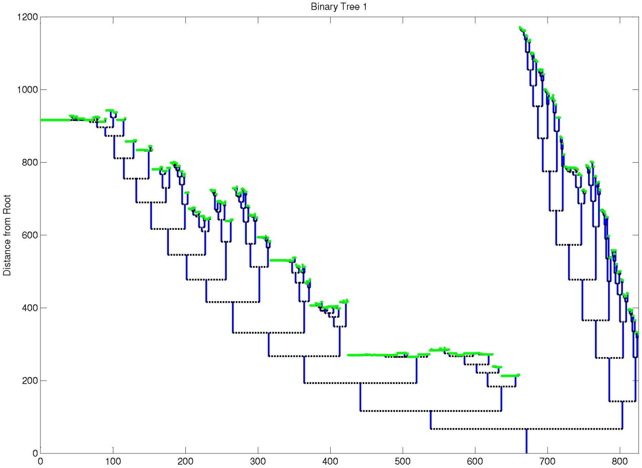
March along the "principal hyperbola" - Page 10



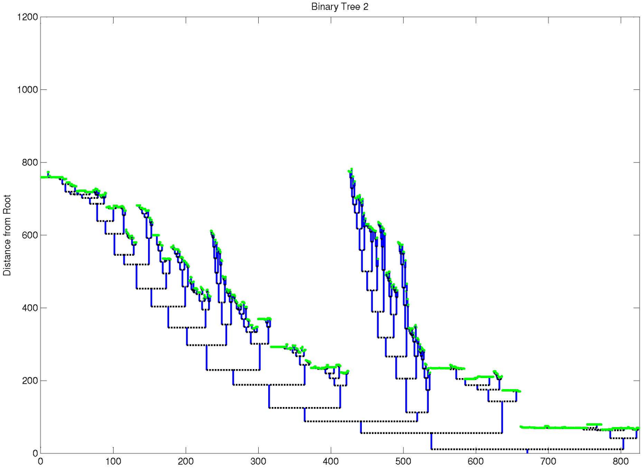
Scree plot for the NMF



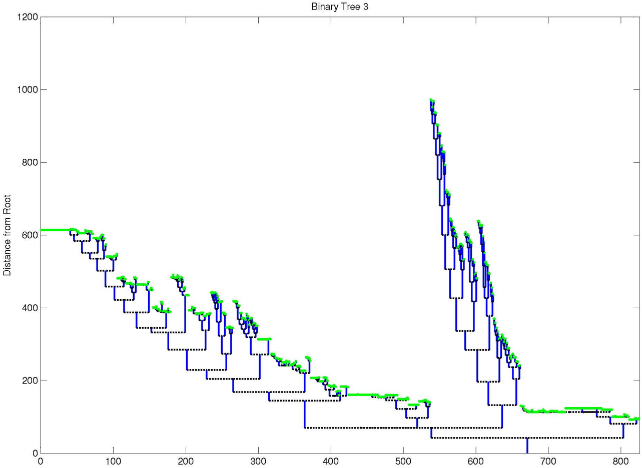
The first 7 columns of W



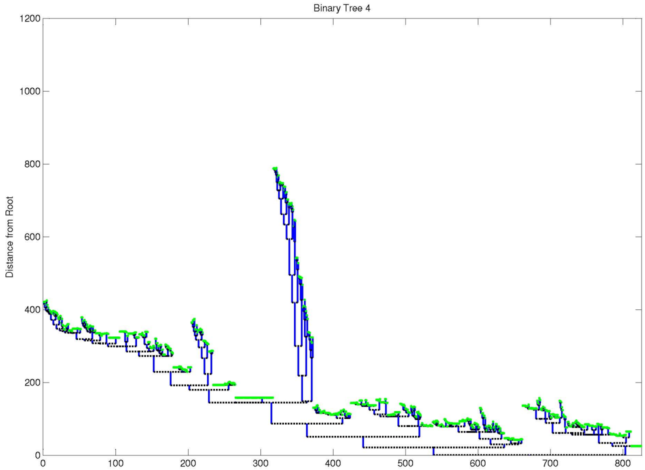
The first 7 columns of W



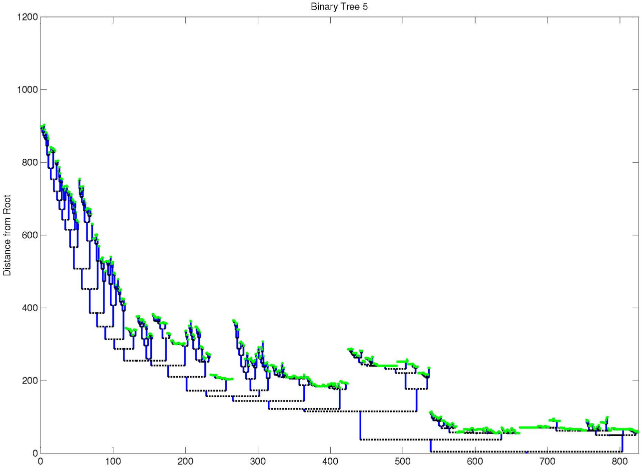
The first 7 columns of W



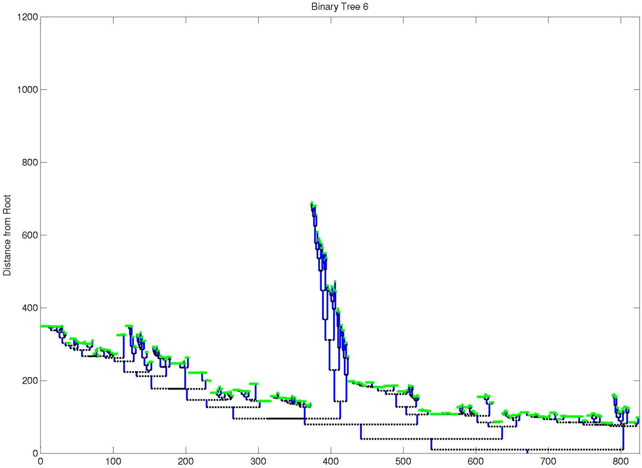
The first 7 columns of W



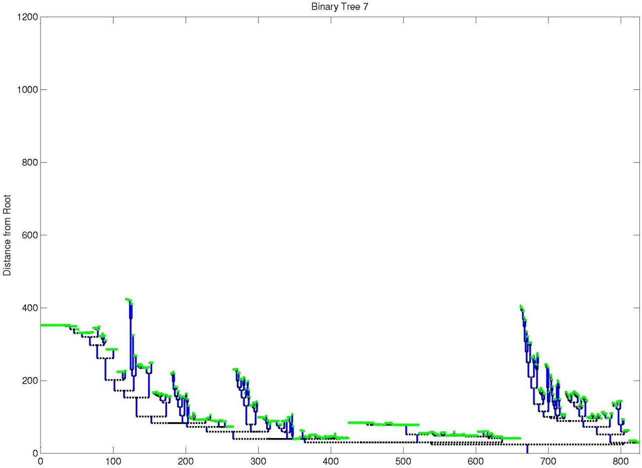
The first 7 columns of W



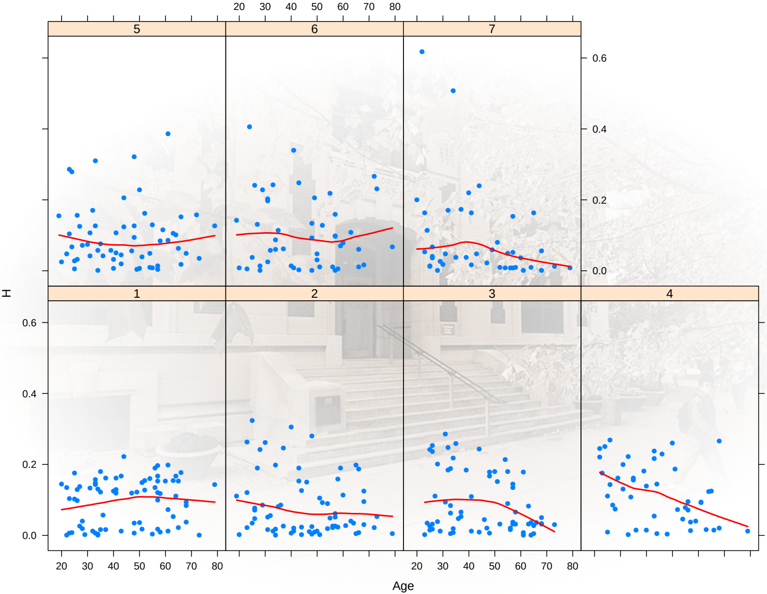
The first 7 columns of W



The first 7 columns of W



Age effects



Results

- Different components (i.e., columns in W) explain different parts in the support tree structure
- Components are more interpretable than the PCs
- Components do have some flat regions

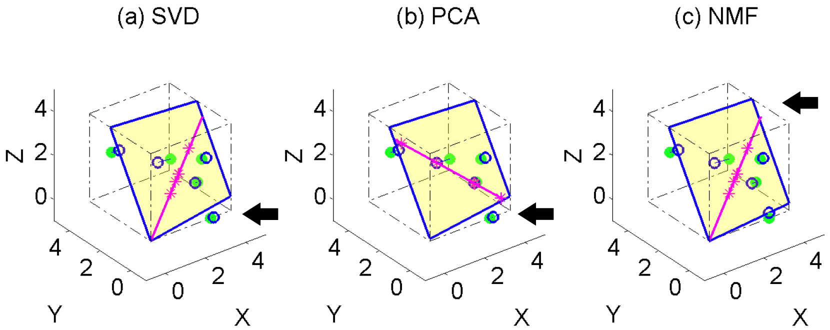
Drawbacks on analysis of nonnegative data objects

- PCA/SVD - good properties
  - PCA projections usually have the "best" interpretation
  - SVD is suitable for two-way data
  - Calculation are easy
  - Different ranks are nested with each other
- PCA/SVD - bad properties
  - SVD is less intuitive in interpretation
  - Both PCA/SVD may leave the first orthant

Drawbacks on analysis of nonnegative data objects

- NMF - good properties
  - Directly target on nonnegative data
  - Usually sparse coefficient, and sparse direction (Learning from parts)
- NMF - bad properties
  - Different ranks are not nested.
  - Decomposition may not be unique
  - Lack of probability/mathematical statistics support

Illustration



Backwards PCA method and Nested Cone Analysis Method

- The above drawbacks lead to a new research
- PCA provides a series of approximation subspaces, and nested
- NMF provides a series of approximation cones, but may not be nested
- Could we propose a novel nonnegative matrix analysis method, that is a series of approximation cones, and they are nested?

Backwards PCA method and Nested Cone Analysis Method

- The above drawbacks lead to a new research
- PCA provides a series of approximation subspaces, and nested
- NMF provides a series of approximation cones, but may not be nested
- Could we propose a novel nonnegative matrix analysis method, that is a series of approximation cones, and they are nested? **My current research, see in 2013 JSM and other conferences**

Summary and discussion

- OODA - active research area
- Nonnegative - challenges and opportunities
- Other constraint statistical analysis
- Knowledge needed : mathematical statistics, visualization, optimization, algorithms, $\cdots$.

Summary and discussion

- OODA - active research area
- Nonnegative - challenges and opportunities
- Other constraint statistical analysis
- Knowledge needed : mathematical statistics, visualization, optimization, algorithms, $\cdots$.

Thank you!