L3 Apprentissage

Michèle Sebag – Benjamin Monmège
LRI – LSV

17 avril 2013

Overview

Introduction

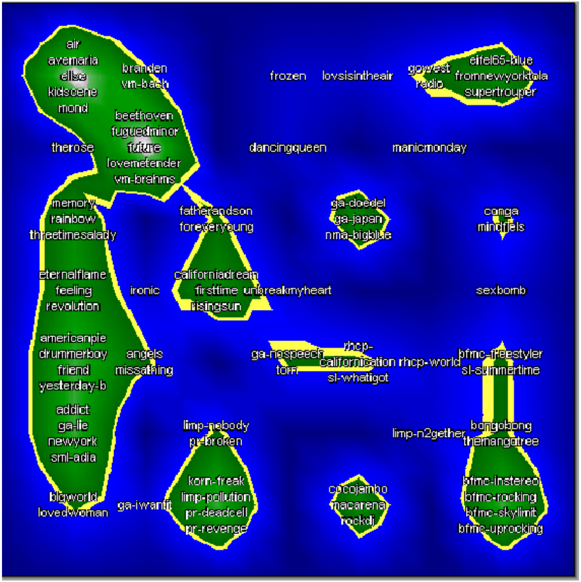
Linear changes of representation

- Principal Component Analysis
- Random projections
- Latent Semantic Analysis

Non linear changes of representation

Clustering

http://www.ofai.at/ elias.pampalk/music/



Unsupervised learning

$$\mathcal{E} = \{\mathbf{x}_1, \ldots, x_n\}$$

Applications

- Documents, text mining
  - indexing, retrieval
- e-commerce, banks, insurance
  - user profiling, recommender systems

Jain, 2010

The representation of the data is closely tied with the purpose of the grouping. The representation must go hand in hand with the end goal of the user. We do not believe there can be a true clustering definable solely in terms of the data – truth is relative to the problem being solved.

Unsupervised learning – WHAT

$$\mathcal{E} = \{\mathbf{x}_1, \ldots, x_n\}$$

Questions

- What is in the data
  divide and conquer
- Abstraction: (lossy) compression
  tradeoff precision/model size
- Regularities/ Anomaly detection
- What is the (generative) model ?
  how to account for prior knowledge

Goals

Unsupervised learning – HOW

Position of the problem

Algorithmic issues

- Stationary data — clustering; density estimation
- Online data — Data streaming

  Tradeoff precision / noise

- Non-stationary data:

  change detection / noise

Real-time ? Limited resources ?

Validation

- exploratory data analysis (subjective)
- density estimation (likelihood)

Abstraction

Artifacts

- Represent a set of instances $\mathbf{x}_i \in X$ by an element $z \in X$
- Examples:
  - Mean state of a system
  - Sociological profile

Prototypes

- Find the most representatives instances among $\mathbf{x}_1, \ldots \mathbf{x}_n$
- How many prototypes ? (avoid overfitting)
- Examples:
  - Faces
  - Molecules

Generative models

Given

$$\mathcal{E} = \{x_i \in X\} \; P(x)$$

Find $\hat{P}(x)$.

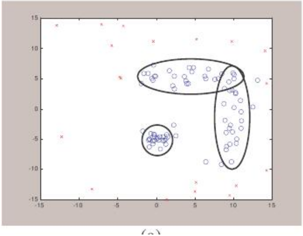
Issues

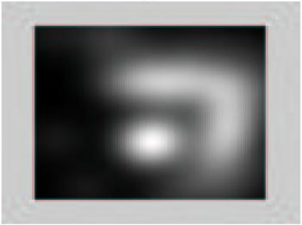
- A distribution (sums to 1)
- Parametric (e.g. mixture of Gaussians) vs non- ?
- Which criterion ? optimize (log) likelihood of data.

One-class SVM

Formulation

$$\begin{cases} \text{Min.} & \frac{1}{2}\,||\mathbf{w}||^2 - \rho \;\; + \mathbf{C}\sum_\mathbf{i}\xi_\mathbf{i} \\ \text{s.t.} & \forall\; i = 1 \dots n \\ & \quad \langle \mathbf{w}, \mathbf{x}_i \rangle \geq \rho\; -\xi_\mathbf{i} \end{cases}$$





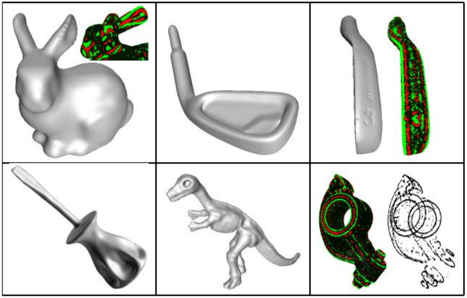
Dual problem

$$\begin{cases} \text{Min.} & \sum_{i,j} \alpha_i \alpha_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle \\ \text{s.t.} & \forall\; i = 1 \dots n \quad 0 \leq \alpha_i \leq C \\ & \sum_i \alpha_i = 0 \end{cases}$$

Implicit surface modelling

Schoelkopf et al, 04 **Goal**: find the surface formed by the data points

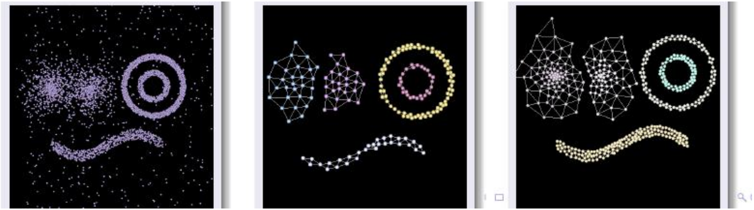
$$\langle \mathbf{w}, \mathbf{x}_i \rangle \geq \rho \text{ becomes } -\varepsilon \leq (\langle \mathbf{w}, \mathbf{x}_i \rangle - \rho) \leq \varepsilon$$



Working assumptions

Clustering assumption

Clusters are separated by a low-density region



Strange clusterings...

http://en.wikipedia.org/wiki/Celestial_Emporium_of_Benevolent_Knowledge%

... a certain Chinese encyclopedia called the Heavenly Emporium of Benevolent Knowledge.
In its distant pages it is written that animals are divided into (a) those that belong to the emperor; (b) embalmed ones; (c) those that are trained; (d) suckling pigs; (e) mermaids; (f) fabulous ones; (g) stray dogs; (h) those that are included in this classification; (i) those that tremble as if they were mad; (j) innumerable ones; (k) those drawn with a very fine camel's-hair brush; (l) etcetera; (m) those that have just broken the flower vase; (n) those that at a distance resemble flies. Borges, 1942

Overview

Introduction

Linear changes of representation

- Principal Component Analysis
- Random projections
- Latent Semantic Analysis

Non linear changes of representation

Dimensionality Reduction – Intuition

Degrees of freedom

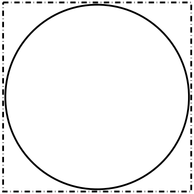
- Image: 4096 pixels; but not independent
- Robotics: (# camera pixels + # infra-red) $\times$ time; but not independent

Goal

Find the (low-dimensional) structure of the data:

- Images
- Robotics
- Genes

Dimensionality Reduction



In high dimension

- Everybody lives in the corners of the space
  Volume of Sphere $V_n = \frac{2\pi r^2}{n} V_{n-2}$
- All points are far from each other

Approaches

- Linear dimensionality reduction
  - Principal Component Analysis
  - Random Projection
- Non-linear dimensionality reduction

Criteria

- Complexity/Size
- Prior knowledge e.g., relevant distance

Linear Dimensionality Reduction

Training set *unsupervised*

$$\mathcal{E} = \{(\mathbf{x}_k), \mathbf{x}_k \in \mathbb{R}^D, k = 1 \ldots N\}$$

Projection from $\mathbb{R}^D$ onto $\mathbb{R}^d$

$$\mathbf{x} \in \mathbb{R}^D \rightarrow \quad h(\mathbf{x}) \in \mathbb{R}^d, \; d << D$$

$$h(\mathbf{x}) = A\mathbf{x}$$

$$s.t. \text{ minimize } \quad \sum_{k=1}^{N} ||\mathbf{x}_k - h(\mathbf{x}_k)||^2$$

Overview

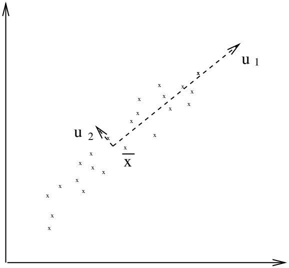
Introduction

Linear changes of representation

- Principal Component Analysis
- Random projections
- Latent Semantic Analysis

Non linear changes of representation

Principal Component Analysis



Covariance matrix S

Mean $\mu_i = \frac{1}{N} \sum_{k=1}^{N} att_i(\mathbf{x}_k)$

$$S_{ij} = \frac{1}{N} \sum_{k=1}^{N} (att_i(\mathbf{x}_k) - \mu_i)(att_j(\mathbf{x}_k) - \mu_j)$$

symmetric $\Rightarrow$ can be diagonalized

$$S = U \Delta U' \qquad \Delta = Diag(\lambda_1, \ldots \lambda_D)$$

Thm: Optimal projection in dimension d

projection on the first d eigenvectors of S

Let u_i the eigenvector associated to eigenvalue λ_i $\qquad \lambda_i > \lambda_{i+1}$

$$h : \mathbb{R}^D \mapsto \mathbb{R}^d, h(\mathbf{x}) = \langle \mathbf{x}, u_1 \rangle u_1 + \ldots + \langle \mathbf{x}, u_d \rangle u_d$$

Sketch of the proof

1. Maximize the variance of $h(\mathbf{x}) = A\mathbf{x}$

$\sum_k ||\mathbf{x}_k - h(\mathbf{x}_k)||^2 = \sum_k ||\mathbf{x}_k||^2 - \sum_k ||h(\mathbf{x}_k)||^2$

$$\text{Minimize } \sum_k ||\mathbf{x}_k - h(\mathbf{x}_k)||^2 \Rightarrow \text{ Maximize } \sum_k ||h(\mathbf{x}_k)||^2$$

$$Var(h(\mathbf{x})) = \frac{1}{N}\left(\sum_k ||h(\mathbf{x}_k)||^2 - ||\sum_k h(\mathbf{x}_k)||^2\right)$$

As

$$||\sum_k h(\mathbf{x}_k)||^2 = ||A\sum_k \mathbf{x}_k||^2 = N^2||A\mu||^2$$

where $\mu = (\mu_1, \ldots .\mu_D)$.
Assuming that $\mathbf{x}_k$ are centered ($\mu_i = 0$) gives the result.

Sketch of the proof, 2

2. Projection on eigenvectors u_i of S

Assume $h(\mathbf{x}) = A\mathbf{x} = \sum_{i=1}^{d} < \mathbf{x}, v_i > v_i$ and show $v_i = u_i$.

$$Var(AX) = (AX)(AX)' = A(XX')A' = ASA' = A(U\Delta U')A'$$

Consider $d = 1,\ v_1 = \sum w_i u_i$

$\sum w_i^2 = 1$
remind $\lambda_i > \lambda_{i+1}$

$$Var(AX) = \sum \lambda_i w_i^2$$

maximized for $w_1 = 1, w_2 = \ldots = w_N = 0$
that is, $v_1 = u_1$.

Principal Component Analysis, Practicalities

Data preparation

- Mean centering the dataset

$$
\begin{aligned}
\mu_i = & \quad \frac{1}{N}\sum_{k=1}^{N} att_i(\mathbf{x}_k) \\
\sigma_i = & \quad \sqrt{\frac{1}{N}\sum_{k=1}^{N} att_i(\mathbf{x}_k)^2 - \mu_i^2} \\
z_k = & \quad \left(\frac{1}{\sigma_i}(att_i(\mathbf{x}_k) - \mu_i)\right)_{i=1}^{D}
\end{aligned}
$$

Matrix operations

- Computing the covariance matrix

$$S_{ij} = \frac{1}{N}\sum_{k=1}^{N} att_i(z_k) att_j(z_k)$$

- Diagonalizing $S = U'\Delta U$ might be not affordable...

Complexity $\mathcal{O}(D^3)$

Overview

Introduction

Linear changes of representation

- Principal Component Analysis
- Random projections
- Latent Semantic Analysis

Non linear changes of representation

Random projection

Random matrix

$$A : \mathbb{R}^D \mapsto \mathbb{R}^d \qquad A[d, D] \qquad A_{i,j} \sim \mathcal{N}(0, 1)$$

define

$$h(\mathbf{x}) = \frac{1}{\sqrt{d}} A\mathbf{x}$$

Property: *h* preserves the norm in expectation

$$E[||h(\mathbf{x})||^2] = ||\mathbf{x}||^2$$

With high probability $1 - 2exp\{-(\varepsilon^2 - \varepsilon^3)\frac{d}{4}\}$

$$(1 - \varepsilon)||\mathbf{x}||^2 \leq ||h(\mathbf{x})||^2 \leq (1 + \varepsilon)||\mathbf{x}||^2$$

Random projection

Proof

$$h(\mathbf{x}) = \frac{1}{\sqrt{d}} A\mathbf{x}$$

$$\begin{aligned}
E(||h(\mathbf{x})||^2) &= \frac{1}{d} E\left[\sum_{i=1}^{d}\left(\sum_{j=1}^{D} A_{i,j}X_j(\mathbf{x})\right)^2\right] \\
&= \frac{1}{d}\sum_{i=1}^{d} E\left[\left(\sum_{j=1}^{D} A_{i,j}X_j(\mathbf{x})\right)^2\right] \\
&= \frac{1}{d}\sum_{i=1}^{d}\sum_{j=1}^{D} E[A_{i,j}^2]E[X_j(\mathbf{x})^2] \\
&= \frac{1}{d}\sum_{i=1}^{d}\sum_{j=1}^{D} \frac{||\mathbf{x}||^2}{D} \\
&= ||\mathbf{x}||^2
\end{aligned}$$

Random projection, 2

Johnson Lindenstrauss Lemma

For $d > \frac{9 \ln N}{\varepsilon^2 - \varepsilon^3}$, with high probability

$$(1-\varepsilon)||\mathbf{x}_i - \mathbf{x}_j||^2 \leq ||h(\mathbf{x}_i) - h(\mathbf{x}_j)||^2 \leq (1+\varepsilon)||\mathbf{x}_i - \mathbf{x}_j||^2$$

More:

`http://www.cs.yale.edu/clique/resources/RandomProjectionMethod.pdf`

Overview

Introduction

Linear changes of representation

Principal Component Analysis

Random projections

Latent Semantic Analysis

Non linear changes of representation

Example

- c1: ***Human*** machine ***interface*** for ABC ***computer*** applications
- c2: A ***survey*** of ***user*** opinion of ***computer*** ***system*** ***response*** ***time***
- c3: The ***EPS*** ***user*** ***interface*** management ***system***
- c4: ***System*** and ***human*** ***system*** engineering testing of ***EPS***
- c5: Relation of ***user*** perceived ***response time*** to error measurement

- m1: The generation of random, binary, ordered ***trees***
- m2: The intersection ***graph*** of paths in ***trees***
- m3: ***Graph*** ***minors*** IV: Widths of ***trees*** and well-quasi-ordering
- m4: ***Graph*** ***minors***: A ***survey***

Example, followed

	c1	c2	c3	c4	c5	m1	m2	m3	m4
human	1	0	0	1	0	0	0	0	0
interface	1	0	1	0	0	0	0	0	0
computer	1	1	0	0	0	0	0	0	0
user	0	1	1	0	1	0	0	0	0
system	0	1	1	2	0	0	0	0	0
response	0	1	0	0	1	0	0	0	0
time	0	1	0	0	1	0	0	0	0
EPS	0	0	1	1	0	0	0	0	0
survey	0	1	0	0	0	0	0	0	1
trees	0	0	0	0	0	1	1	1	0
graph	0	0	0	0	0	0	1	1	1
minors	0	0	0	0	0	0	0	1	1

LSA, 2

Motivations

- Context: bag of words
- Curse of dimensionality $\mathbb{R}^D$
- Synonymy / Polysemy

Goals

- Dimensionality reduction $\mathbb{R}^d$
- Build a decent topology / metric

Remark

- vanilla similarity: cosine
- (why not ?)

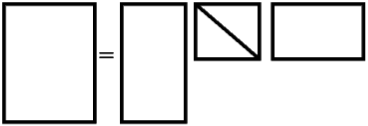
More

`http://lsa.colorado.edu`

LSA, 3

Input

Matrice X = mots $\times$ documents



Principe

1. Changement de base des mots,documents aux concepts
2. Réduction de dimension

Différence Analyse en composantes principales

LSA $\equiv$ Singular Value Decomposition

Input

X matrice mots $\times$ documents $m \times d$

$$X = U' \; S \; V$$

avec

- U : changement de base mots $m \times r$
- V : changement de base des documents $r \times d$
- S : matrice diagonale $r \times r$

Réduction de dimension

- S Ordonner par valeur propre décroissante
- $S' = S$ avec annulation de toutes les vp, sauf les (300) premières.

$$X' = U'S'V$$

Intuition

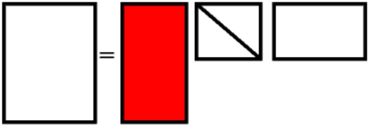
$$X = \begin{pmatrix} & m_1 & m_2 & m_3 & m_4 \\ d_1 & 0 & 1 & 1 & 1 \\ d_2 & 1 & 1 & 1 & 0 \end{pmatrix}$$

m_1 et m_4 ne sont pas “physiquement” ensemble dans les mêmes documents ; mais ils sont avec les mêmes mots ; “donc” ils sont un peu “voisins”...

Après SVD + Réduction,

$$X = \begin{pmatrix} & m_1 & m_2 & m_3 & m_4 \\ d_1 & \epsilon & 1 & 1 & 1 \\ d_2 & 1 & 1 & 1 & \epsilon \end{pmatrix}$$

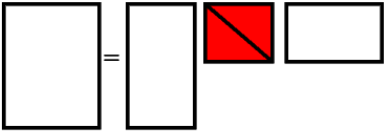
Algorithme



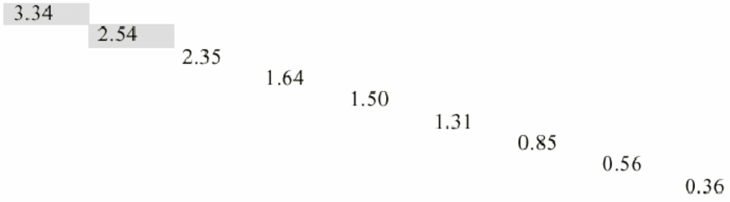
Singular value Decomposition of the words by contexts matrix

0.22	-0.11	0.29	-0.41	-0.11	-0.34	0.52	-0.06	-0.41
0.20	-0.07	0.14	-0.55	0.28	0.50	-0.07	-0.01	-0.11
0.24	0.04	-0.16	-0.59	-0.11	-0.25	-0.30	0.06	0.49
0.40	0.06	-0.34	0.10	0.33	0.38	0.00	0.00	0.01
0.64	-0.17	0.36	0.33	-0.16	-0.21	-0.17	0.03	0.27
0.27	0.11	-0.43	0.07	0.08	-0.17	0.28	-0.02	-0.05
0.27	0.11	-0.43	0.07	0.08	-0.17	0.28	-0.02	-0.05
0.30	-0.14	0.33	0.19	0.11	0.27	0.03	-0.02	-0.17
0.21	0.27	-0.18	-0.03	-0.54	0.08	-0.47	-0.04	-0.58
0.01	0.49	0.23	0.03	0.59	-0.39	-0.29	0.25	-0.23
0.04	0.62	0.22	0.00	-0.07	0.11	0.16	-0.68	0.23
0.03	0.45	0.14	-0.01	-0.30	0.28	0.34	0.68	0.18

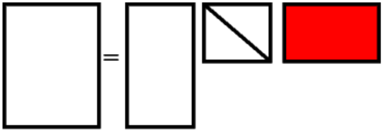
Algorithme, 2



Singular value Decomposition of the words by contexts matrix



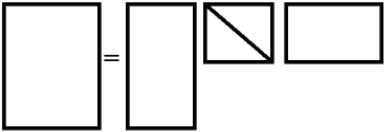
Algorithme. 3



Singular value Decomposition of the words by contexts matrix

0.20	0.61	0.46	0.54	0.28	0.00	0.01	0.02	0.08
-0.06	0.17	-0.13	-0.23	0.11	0.19	0.44	0.62	0.53
0.11	-0.50	0.21	0.57	-0.51	0.10	0.19	0.25	0.08
-0.95	-0.03	0.04	0.27	0.15	0.02	0.02	0.01	-0.03
0.05	-0.21	0.38	-0.21	0.33	0.39	0.35	0.15	-0.60
-0.08	-0.26	0.72	-0.37	0.03	-0.30	-0.21	0.00	0.36
0.18	-0.43	-0.24	0.26	0.67	-0.34	-0.15	0.25	0.04
-0.01	0.05	0.01	-0.02	-0.06	0.45	-0.76	0.45	-0.07
-0.06	0.24	0.02	-0.08	-0.26	-0.62	0.02	0.52	-0.45

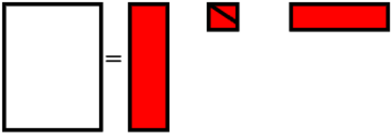
Algorithme, 4



Singular value Decomposition of the words by contexts matrix

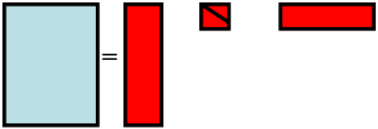
3.34	
	2.54

Algorithme, 5



Singular value Decomposition of the words by contexts matrix

Algorithme, 6



Singular value Decomposition of the words by contexts matrix

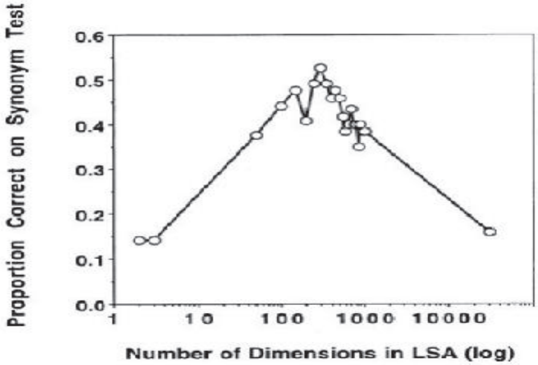
	c1	c2	c3	c4	c5	m1	m2	m3	m4
human	0.16	0.40	0.38	0.47	0.18	-0.05	-0.12	-0.16	-0.09
interface	0.14	0.37	0.33	0.40	0.16	-0.03	-0.07	-0.10	-0.04
computer	0.15	0.51	0.36	0.41	0.24	0.02	0.06	0.09	0.12
user	0.26	0.84	0.61	0.70	0.39	0.03	0.08	0.12	0.19
system	0.45	1.23	1.05	1.27	0.56	-0.07	-0.15	-0.21	-0.05
response	0.16	0.58	0.38	0.42	0.28	0.06	0.13	0.19	0.22
time	0.16	0.58	0.38	0.42	0.28	0.06	0.13	0.19	0.22
EPS	0.22	0.55	0.51	0.63	0.24	-0.07	-0.14	-0.20	-0.11
survey	0.10	0.53	0.23	0.21	0.27	0.14	0.31	0.44	0.42
trees	-0.06	0.23	-0.14	-0.27	0.14	0.24	0.55	0.77	0.66
graph	-0.06	0.34	-0.15	-0.30	0.20	0.31	0.69	0.98	0.85
minors	-0.04	0.25	-0.10	-0.21	0.15	0.22	0.50	0.71	0.62

	c1	c2	c3	c4	c5	m1	m2	m3	m4
human	1	0	0	1	0	0	0	0	0
interface	1	0	1	0	0	0	0	0	0
computer	1	1	0	0	0	0	0	0	0
user	0	1	1	0	1	0	0	0	0
system	0	1	1	2	0	0	0	0	0
response	0	1	0	0	1	0	0	0	0
time	0	1	0	0	1	0	0	0	0
EPS	0	0	1	1	0	0	0	0	0
survey	0	1	0	0	0	0	0	0	1
trees	0	0	0	0	0	1	1	1	0
graph	0	0	0	0	0	0	1	1	1
minors	0	0	0	0	0	0	0	1	1

Discussion

Une application

Test de synonymie — TOEFL



Déterminer le nb de dimensions/vp

Expérimentalement...

Quelques remarques

et la négation ? — aucune importance (!)

battu par: nb de hits sur le Web — P. Turney

Quelques applications

- Educational Text Selection
  Permet de sélectionner automatiquement des textes permettant d'accroître les connaissances de l'utilisateur.
- Essay Scoring
  Permet de noter la qualité d'une rédaction d'étudiant
- Summary Scoring & Revision
  Apprendre à l'utilisateur à faire un résumé
- Cross Language Retrieval
  permet de soumettre un texte dans une langue et d'obtenir un texte équivalent dans une autre langue

LSA – Analyse en composantes principales

Ressemblances

- Prendre une matrice
- La mettre sous forme diagonale
- Annuler toutes les valeurs propres sauf les plus grandes
- Projeter sur l'espace obtenu

Différences

	ACP	LSA
Matrice	covariance attributs	mots $\times$ documents
d	2-3	100-300

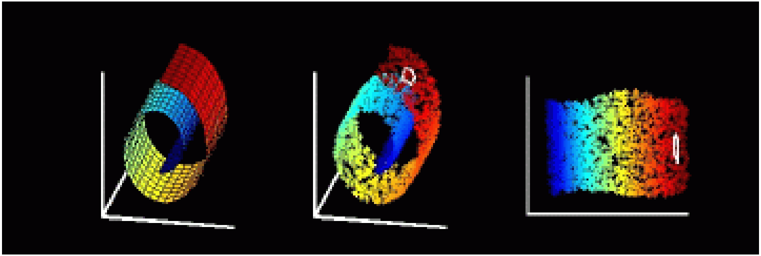
Overview

Introduction

Linear changes of representation
- Principal Component Analysis
- Random projections
- Latent Semantic Analysis

Non linear changes of representation

Non-Linear Dimensionality Reduction



Conjecture

Examples live in a manifold of dimension $d << D$

Goal: consistent projection of the dataset onto $\mathbb{R}^d$

Consistency:

- Preserve the structure of the data
- e.g. preserve the distances between points

Multi-Dimensional Scaling

Position of the problem

- Given $\{\mathbf{x}_1, \ldots, \mathbf{x}_N,\ \mathbf{x}_i \in \mathbb{R}^D\}$
- Given $sim(\mathbf{x}_i, \mathbf{x}_j) \in \mathbb{R}^+$
- Find projection Φ onto $\mathbb{R}^d$

$$\begin{array}{rl} x \in \mathbb{R}^D \rightarrow & \Phi(x) \in \mathbb{R}^d \\ sim(\mathbf{x}_i, \mathbf{x}_j) \sim & sim(\Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j)) \end{array}$$

Optimisation

Define $X,\ X_{i,j} = sim(\mathbf{x}_i, \mathbf{x}_j)$; $X^\Phi,\ X^\Phi_{i,j} = sim(\Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j))$

Find Φ minimizing $||X - X'||$

Rq : Linear Φ = Principal Component Analysis

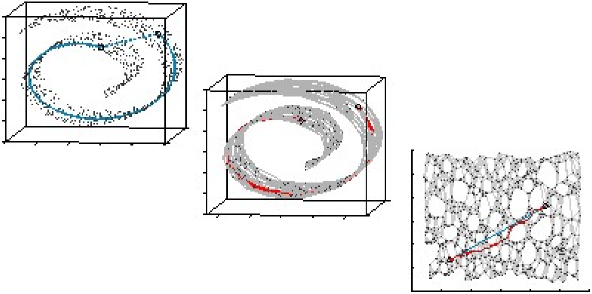
But linear MDS does not work: preserves all distances, while **only local distances are meaningful**

Non-linear projections

Approaches

- Reconstruct global structures from local ones and find global projection **Isomap**
- Only consider local structures **LLE**

Intuition: locally, points live in $\mathbb{R}^d$



Isomap

Tenenbaum, da Silva, Langford 2000
`http://isomap.stanford.edu`

Estimate $d(x_i, x_j)$

- Known if $\mathbf{x}_i$ and $\mathbf{x}_j$ are close
- Otherwise, compute the shortest path between $\mathbf{x}_i$ and $\mathbf{x}_j$
geodesic distance (dynamic programming)

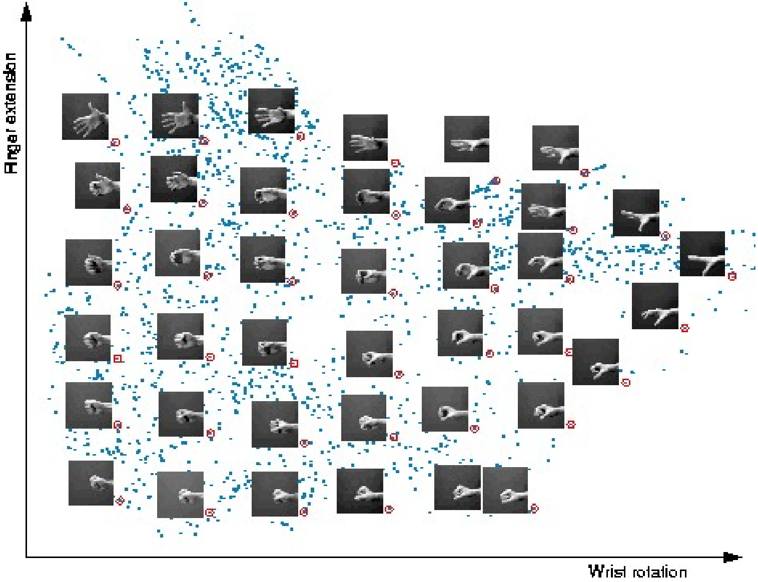
Requisite

If data points sampled in a convex subset of $\mathbb{R}^d$,
then geodesic distance $\sim$ Euclidean distance on $\mathbb{R}^d$.

General case

- Given $d(\mathbf{x}_i, \mathbf{x}_j)$, estimate $< \mathbf{x}_i, \mathbf{x}_j >$
- Project points in $\mathbb{R}^d$

Isomap, 2



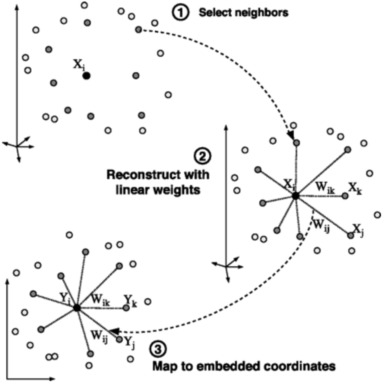
Locally Linear Embedding

Roweiss and Saul, 2000

http://www.cs.toronto.edu/∼roweis/lle/

Principle

- Find local description for each point: depending on its neighbors



Local Linear Embedding, 2

Find neighbors

For each $\mathbf{x}_i$, find its nearest neighbors $\mathcal{N}(i)$

Parameter: number of neighbors

Change of representation

Goal Characterize $\mathbf{x}_i$ wrt its neighbors:

$$\mathbf{x}_i = \sum_{j \in \mathcal{N}(i)} w_{i,j} \mathbf{x}_j \quad \text{with} \quad \sum_{j \in \mathcal{N}(i)} w_{ij} = 1$$

Property: invariance by translation, rotation, homothety
How Compute the local covariance matrix:

$$C_{j,k} = < x_j - x_i, x_k - x_i >$$

Find vector w_i s.t. $Cw_i = 1$

Local Linear Embedding, 3

Algorithm

Local description: Matrix W such that $\sum_j w_{i,j} = 1$

$$W = argmin\{\sum_{i=1}^{N} ||\mathbf{x}_i - \sum_j w_{i,j}\mathbf{x}_j||^2\}$$

Projection: Find $\{z_1, \ldots, z_n\}$ *in* $\mathbb{R}^d$ minimizing

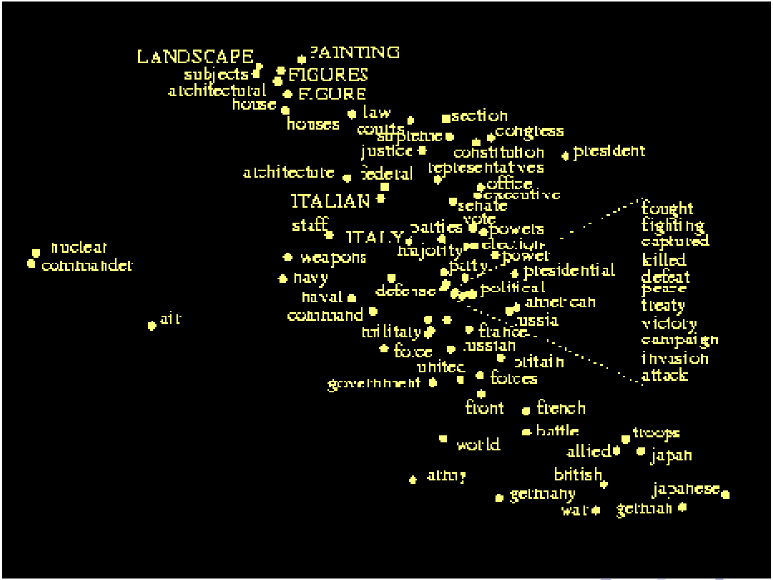
$$\sum_{i=1}^{N} ||z_i - \sum_j w_{i,j} z_j||^2$$

$$Minimize\ ((I - W)Z)'((I - W)Z) = Z'(I - W)'(I - W)Z$$

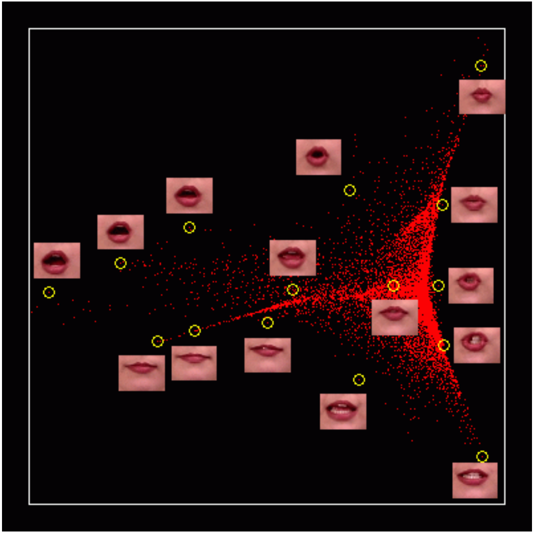
Solutions: vectors z_i are eigenvectors of $(I - W)'(I - W)$

- Keeping the d eigenvectors with lowest eigenvalues > 0

Example, Texts



Example, Images



LLE

Overview

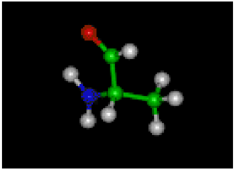
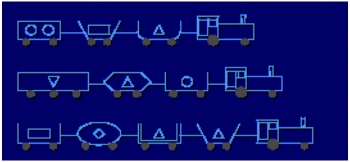
Introduction

Linear changes of representation
- Principal Component Analysis
- Random projections
- Latent Semantic Analysis

Non linear changes of representation

Propositionalization

Relational domains

Relational learning

PROS Inductive Logic Programming

Use domain knowledge

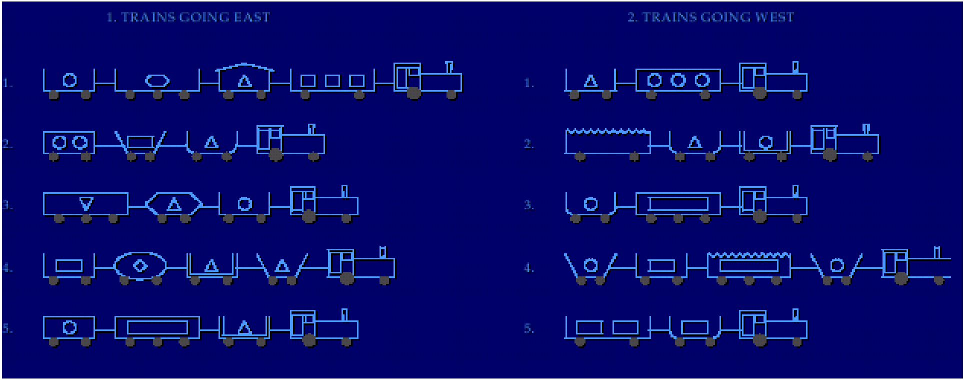
CONS Data Mining

Covering test $\equiv$ subgraph matching exponential complexity

Getting back to propositional representation:
propositionalization

West - East trains

Michalski 1983



Propositionalization

Linus (ancestor)

Lavrac et al, 94

$$West(a) \leftarrow \quad Engine(a, b), first\_wagon(a, c), roof(c), load(c, square, 3)...$$
$$West(a') \leftarrow \quad Engine(a', b'), first\_wagon(a', c'), load(c', circle, 1)...$$

West	Engine(X)	First Wagon(X,Y)	Roof(Y)	$Load_1$ (Y)	$Load_2$ (Y)
a	b	c	yes	square	3
a’	b’	c’	no	circle	1

Each column: a role predicate, where the predicate is determinate linked to former predicates (left columns) with a single instantiation in every example

Propositionalization

Stochastic propositionalization

Kramer, 98 Construct random formulas $\equiv$ boolean features

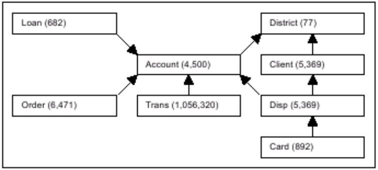
SINUS – RDS

`http://www.cs.bris.ac.uk/home/rawles/sinus`

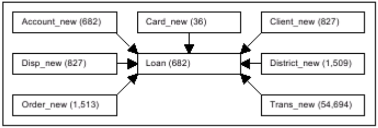
`http://labe.felk.cvut.cz/~zelezny/rsd`

- Use modes (user-declared) `modeb(2,hasCar(+train,-car))`
- Thresholds on number of variables, depth of predicates...
- Pre-processing (feature selection)

Propositionalization



DB Schema



Propositionalization

RELAGGS

Database aggregates

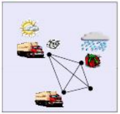
- average, min, max, of numerical attributes
- number of values of categorical attributes

Apprentissage par Renforcement Relationnel

Real Time Strategy Games

- Many objects of various types in complex interactions
- Good players can generalize across situations involving distinct object configurations

The Logistics Domain

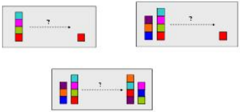


- Move many objects around with many other objects
- Identities and numbers of objects always changing

Robot Soccer

- Reasoning about relationship between objects (players and ball) key to good play

and of course ..... Blocksworld



- Would like a policy that is independent of number of objects/blocks

Propositionalisation

Contexte variable

- Nombre de robots, position des robots
- Nombre de camions, lieu des secours

Besoin: Abstraire et Generaliser

Attributs

- Nombre d'amis/d'ennemis
- Distance du plus proche robot ami
- Distance du plus proche ennemi