NYU | COURANT INSTITUTE OF MATHEMATICAL SCIENCES

NYU DATA SCIENCE

Low-rank Models for Data Analysis

Carlos Fernandez-Granda

www.cims.nyu.edu/~cfgranda

2/27/2018

Background

Low-rank models

Matrix completion

Structured low-rank models

Data-driven Analysis of Infant Sleep Patterns

Rank

For any matrix A

$$\dim(\text{col}(A)) = \dim(\text{row}(A))$$

This is the rank of A

Singular value decomposition

Every rank r real matrix $A \in R^{m \times n}$, has a singular-value decomposition (SVD) of the form

$$
\begin{aligned}
A &= \begin{bmatrix} \vec{u}_1 & \vec{u}_2 & \cdots & \vec{u}_r \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ & & \ddots & \\ 0 & 0 & \cdots & \sigma_r \end{bmatrix} \begin{bmatrix} \vec{v}_1^{\,T} \\ \vec{v}_2^{\,T} \\ \vdots \\ \vec{v}_r^{\,T} \end{bmatrix} \\
&= USV^T
\end{aligned}
$$

Singular value decomposition

- The singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$ are positive real numbers
- The left singular vectors $\vec{u}_1, \vec{u}_2, \ldots \vec{u}_r$ form an orthonormal set
- The right singular vectors $\vec{v}_1, \vec{v}_2, \ldots \vec{v}_r$ also form an orthonormal set
- The SVD is unique if all the singular values are different
- If $\sigma_i = \sigma_{i+1} = \ldots = \sigma_{i+k}$, then $\vec{u}_i, \ldots, \vec{u}_{i+k}$ can be replaced by any orthonormal basis of their span (the same holds for $\vec{v}_i, \ldots, \vec{v}_{i+k}$)
- The SVD of an $m \times n$ matrix with $m \geq n$ can be computed in $\mathcal{O}\left(mn^2\right)$

Column and row space

- The left singular vectors $\vec{u}_1$, $\vec{u}_2$, $\ldots \vec{u}_r$ are a basis for the column space
- The right singular vectors $\vec{v}_1$, $\vec{v}_2$, $\ldots \vec{v}_r$ are a basis for the row space

Best rank-k approximation

Let USV^T be the SVD of a matrix $A \in \mathbb{R}^{m \times n}$

The truncated SVD $U_{:,1:k} S_{1:k,1:k} V_{:,1:k}^T$ is the best rank-k approximation

$$U_{:,1:k} S_{1:k,1:k} V_{:,1:k}^T = \arg\min_{\{\widetilde{A} \mid \text{rank}(\tilde{A})=k\}} \left|\left| A - \widetilde{A} \right|\right|_{\text{F}}$$

Motivation

Quantity $y[i, j]$ depends on indices i and j

We observe examples and want to predict new instances

In collaborative filtering, $y[i, j]$ is rating given to a movie i by a user j

Collaborative filtering

$$Y := \begin{array}{c} \\ \end{array} \begin{array}{cccc} \text{Bob} & \text{Molly} & \text{Mary} & \text{Larry} \end{array}$$

$$Y := \begin{pmatrix} 1 & 1 & 5 & 4 \\ 2 & 1 & 4 & 5 \\ 4 & 5 & 2 & 1 \\ 5 & 4 & 2 & 1 \\ 4 & 5 & 1 & 2 \\ 1 & 2 & 5 & 5 \end{pmatrix} \begin{array}{l} \text{The Dark Knight} \\ \text{Spiderman 3} \\ \text{Love Actually} \\ \text{Bridget Jones's Diary} \\ \text{Pretty Woman} \\ \text{Superman 2} \end{array}$$

Simple model

Assumptions:

- Some movies are more popular in general
- Some users are more generous in general

$$y[i,j] \approx a[i]b[j]$$

- $a[i]$ quantifies popularity of movie i
- $b[j]$ quantifies generosity of user j

Rank-1 model

Assume m movies are all rated by n users

Model becomes

$$Y \approx \vec{a}\,\vec{b}^{\,T}$$

We can fit it by solving

$$\min_{\vec{a}\in\mathbb{R}^m,\,\vec{b}\in\mathbb{R}^n} \left|\left| Y - \vec{a}\,\vec{b}^{\,T} \right|\right|_{\text{F}} \qquad \text{subject to} \quad \left|\left|\vec{a}\right|\right|_2 = 1$$

Equivalent to

Rank-1 model

Assume m movies are all rated by n users

Model becomes

$$Y \approx \vec{a}\,\vec{b}^{\,T}$$

We can fit it by solving

$$\min_{\vec{a} \in \mathbb{R}^m,\, \vec{b} \in \mathbb{R}^n} \left|\left| Y - \vec{a}\,\vec{b}^{\,T} \right|\right|_{\mathrm{F}} \quad \text{subject to} \quad ||\vec{a}||_2 = 1$$

Equivalent to

$$\min_{X \in \mathbb{R}^{m \times n}} ||Y - X||_{\mathrm{F}} \quad \text{subject to} \quad \operatorname{rank}(X) = 1$$

Best rank-k approximation

Let USV^T be the SVD of a matrix $A \in \mathbb{R}^{m \times n}$

The truncated SVD $U_{:,1:k} S_{1:k,1:k} V_{:,1:k}^T$ is the best rank-k approximation

$$U_{:,1:k} S_{1:k,1:k} V_{:,1:k}^T = \arg\min_{\{\widetilde{A} \,|\, \operatorname{rank}(\tilde{A}) = k\}} \left|\left| A - \widetilde{A} \right|\right|_{\mathrm{F}}$$

Rank-1 model

$$\sigma_1 \vec{u}_1 \vec{v}_1^{\,T} = \arg \min_{X \in \mathbb{R}^{m \times n}} ||Y - X||_{\text{F}} \qquad \text{subject to} \quad \text{rank}\,(X) = 1$$

The solution to

$$\min_{\vec{a} \in \mathbb{R}^m,\, \vec{b} \in \mathbb{R}^n} \left|\left| Y - \vec{a}\,\vec{b}^{\,T} \right|\right|_{\text{F}} \qquad \text{subject to} \quad ||\vec{a}||_2 = 1$$

is

$$\vec{a}_{\text{min}} =$$
$$\vec{b}_{\text{min}} =$$

Rank-1 model

$$\sigma_1 \vec{u}_1 \vec{v}_1^{\,T} = \arg \min_{X \in \mathbb{R}^{m \times n}} ||Y - X||_{\mathrm{F}} \qquad \text{subject to} \quad \operatorname{rank}(X) = 1$$

The solution to

$$\min_{\vec{a} \in \mathbb{R}^m,\, \vec{b} \in \mathbb{R}^n} \left|\left| Y - \vec{a}\,\vec{b}^{\,T} \right|\right|_{\mathrm{F}} \qquad \text{subject to} \quad ||\vec{a}||_2 = 1$$

is

$$\vec{a}_{\min} = \vec{u}_1$$
$$\vec{b}_{\min} = \sigma_1 \vec{v}_1$$

Rank-r model

Certain people like certain movies: r factors

$$y[i,j] \approx \sum_{l=1}^{r} a_l[i] b_l[j]$$

For each factor l

- $a_l[i]$: movie i is positively (> 0), negatively (< 0) or not (≈ 0) associated to factor l
- $b_l[j]$: user j likes (> 0), hates (< 0) or is indifferent (≈ 0) to factor l

Rank-r model

Equivalent to

$$Y \approx AB, \qquad A \in \mathbb{R}^{m \times r}, \quad B \in \mathbb{R}^{r \times n}$$

SVD solves

$$\min_{A \in \mathbb{R}^{m \times r}, B \in \mathbb{R}^{r \times n}} \left|\left| Y - A\,B \right|\right|_{\mathrm{F}} \qquad \text{subject to} \quad \left|\left|\vec{a}_1\right|\right|_2 = 1, \ldots, \left|\left|\vec{a}_r\right|\right|_2 = 1$$

Problem: Many possible ways of choosing $\vec{a}_1, \ldots, \vec{a}_r, \vec{b}_1, \ldots, \vec{b}_r$

SVD constrains them to be orthogonal

Collaborative filtering

$$Y := \begin{pmatrix} 1 & 1 & 5 & 4 \\ 2 & 1 & 4 & 5 \\ 4 & 5 & 2 & 1 \\ 5 & 4 & 2 & 1 \\ 4 & 5 & 1 & 2 \\ 1 & 2 & 5 & 5 \end{pmatrix}$$

	Bob	Molly	Mary	Larry
The Dark Knight	1	1	5	4
Spiderman 3	2	1	4	5
Love Actually	4	5	2	1
Bridget Jones's Diary	5	4	2	1
Pretty Woman	4	5	1	2
Superman 2	1	2	5	5

SVD

$$A - \mu \vec{1}\,\vec{1}^{\,T} = USV^T = U \begin{bmatrix} 7.79 & 0 & 0 & 0 \\ 0 & 1.62 & 0 & 0 \\ 0 & 0 & 1.55 & 0 \\ 0 & 0 & 0 & 0.62 \end{bmatrix} V^T$$

$$\mu := \frac{1}{n} \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij}$$

Rank 1 model

$\bar{A} + \sigma_1 \vec{u}_1 \vec{v}_1^T =$

Bob	Molly	Mary	Larry	
1.34 (1)	1.19 (1)	4.66 (5)	4.81 (4)	The Dark Knight
1.55 (2)	1.42 (1)	4.45 (4)	4.58 (5)	Spiderman 3
4.45 (4)	4.58 (5)	1.55 (2)	1.42 (1)	Love Actually
4.43 (5)	4.56 (4)	1.57 (2)	1.44 (1)	B.J.'s Diary
4.43 (4)	4.56 (5)	1.57 (1)	1.44 (2)	Pretty Woman
1.34 (1)	1.19 (2)	4.66 (5)	4.81 (5)	Superman 2

Movies

$$\vec{a}_1 = \begin{pmatrix} \overset{\text{D. Knight}}{-0.45} & \overset{\text{Sp. 3}}{-0.39} & \overset{\text{Love Act.}}{0.39} & \overset{\text{B.J.'s Diary}}{0.39} & \overset{\text{P. Woman}}{0.39} & \overset{\text{Sup. 2}}{-0.45} \end{pmatrix}$$

Coefficients cluster movies into action (+) and romantic (-)

Users

$$\vec{b}_1 = \begin{array}{cccc} \text{Bob} & \text{Molly} & \text{Mary} & \text{Larry} \\ (3.74 & 4.05 & -3.74 & -4.05) \end{array}$$

Coefficients cluster people into action (-) and romantic (+)

Netflix Prize

Matrix completion

	Bob	Molly	Mary	Larry
The Dark Knight	1	?	5	4
Spiderman 3	?	1	4	5
Love Actually	4	5	2	?
Bridget Jones's Diary	5	4	2	1
Pretty Woman	4	5	1	2
Superman 2	1	2	?	5

Isn't this completely ill posed?

Can't we fill in the missing entries arbitrarily?

Isn't this completely ill posed?

Can't we fill in the missing entries arbitrarily?

Yes, but not if matrix is low rank

Isn't this completely ill posed?

Can't we fill in the missing entries arbitrarily?

Yes, but not if matrix is low rank

Then it depends on $\approx r\,(m+n)$ parameters

As long as data > parameters recovery is possible (in principle)

$$\begin{bmatrix} 1 & 1 & 1 & 1 & ? & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ ? & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

Matrix cannot be sparse

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 23 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Singular vectors cannot be sparse

$$\begin{bmatrix}1\\1\\1\\1\end{bmatrix}\begin{bmatrix}1 & 1 & 1 & 1\end{bmatrix} + \begin{bmatrix}0\\0\\0\\1\end{bmatrix}\begin{bmatrix}1 & 2 & 3 & 4\end{bmatrix} = \begin{bmatrix}1 & 1 & 1 & 1\\1 & 1 & 1 & 1\\1 & 1 & 1 & 1\\2 & 3 & 4 & 5\end{bmatrix}$$

Incoherence

The matrix must be incoherent: its singular vectors must be spread out

For $1/\sqrt{n} \leq \mu \leq 1$

$$\max_{1 \leq i \leq r, 1 \leq j \leq m} |U_{ij}| \leq \mu$$

$$\max_{1 \leq i \leq r, 1 \leq j \leq n} |V_{ij}| \leq \mu$$

for the left $U_1, \ldots, U_r$ and right $V_1, \ldots, V_r$ singular vectors

Measurements

We must see an entry in each row/column at least

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ ? & ? & ? & ? \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 \\ ? \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}$$

Assumption: Random sampling (usually does not hold in practice!)

Low-rank matrix estimation

First idea:

$$\min_{X \in \mathbb{R}^{m \times n}} \operatorname{rank}(X) \quad \text{such that } X_\Omega = y$$

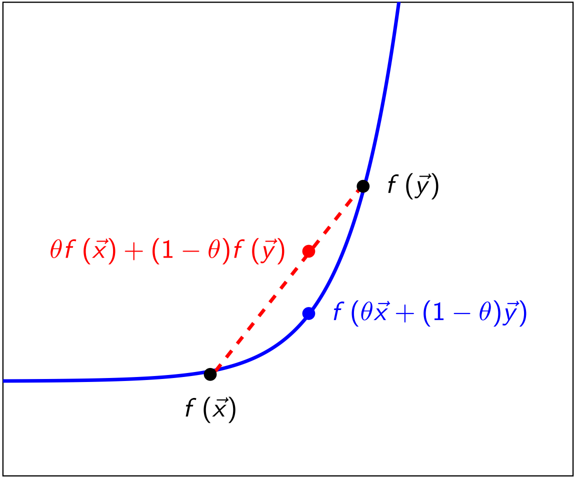
Ω: indices of revealed entries
y: revealed entries

Convex functions

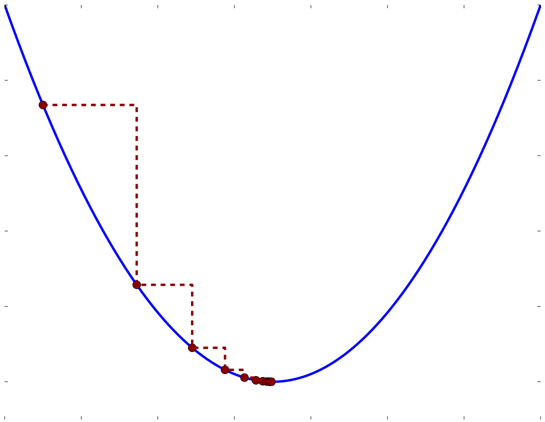
A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if for any $\vec{x}, \vec{y} \in \mathbb{R}^n$ and any $\theta \in (0, 1)$

$$\theta f(\vec{x}) + (1 - \theta) f(\vec{y}) \geq f(\theta \vec{x} + (1 - \theta) \vec{y})$$

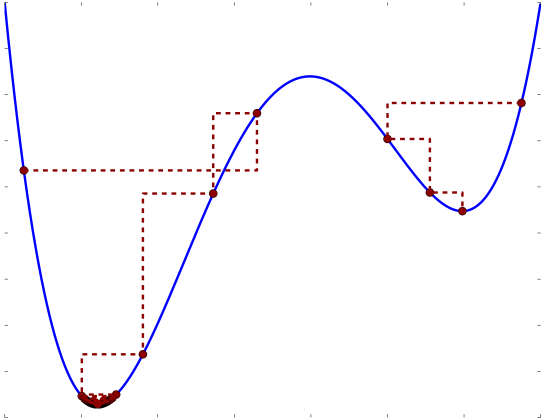
Convex functions



Minimizing convex functions



Minimizing nonconvex functions



The rank is not convex

The rank of matrices in $\mathbb{R}^{n \times n}$ interpreted as a function from $\mathbb{R}^{n \times n}$ to $\mathbb{R}$ is not convex

The rank is not convex

The rank of matrices in $\mathbb{R}^{n\times n}$ interpreted as a function from $\mathbb{R}^{n\times n}$ to $\mathbb{R}$ is not convex

$$X := \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \qquad Y := \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

For any $\theta \in (0,1)$

$$\operatorname{rank}\left(\theta X + (1-\theta)\, Y\right)$$

$$\theta \operatorname{rank}(X) + (1-\theta) \operatorname{rank}(Y)$$

The rank is not convex

The rank of matrices in $\mathbb{R}^{n \times n}$ interpreted as a function from $\mathbb{R}^{n \times n}$ to $\mathbb{R}$ is not convex

$$X := \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \qquad Y := \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

For any $\theta \in (0,1)$

$$\operatorname{rank}\left(\theta X + (1-\theta)\, Y\right) = 2$$

$$\theta \operatorname{rank}(X) + (1-\theta) \operatorname{rank}(Y)$$

The rank is not convex

The rank of matrices in $\mathbb{R}^{n\times n}$ interpreted as a function from $\mathbb{R}^{n\times n}$ to $\mathbb{R}$ is not convex

$$X := \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \qquad Y := \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

For any $\theta \in (0,1)$

$$\operatorname{rank}\left(\theta X + (1-\theta)\, Y\right) = 2$$

$$\theta \operatorname{rank}(X) + (1-\theta) \operatorname{rank}(Y) = 1$$

Norms are convex

For any $\vec{x}, \vec{y} \in \mathbb{R}^n$ and any $\theta \in (0, 1)$

$$||\theta \vec{x} + (1 - \theta) \vec{y}||$$

Norms are convex

For any $\vec{x}, \vec{y} \in \mathbb{R}^n$ and any $\theta \in (0,1)$

$$||\theta\vec{x} + (1-\theta)\,\vec{y}|| \leq ||\theta\vec{x}|| + ||(1-\theta)\,\vec{y}||$$

Norms are convex

For any $\vec{x}, \vec{y} \in \mathbb{R}^n$ and any $\theta \in (0, 1)$

$$\begin{aligned} ||\theta \vec{x} + (1-\theta)\,\vec{y}|| &\leq ||\theta \vec{x}|| + ||(1-\theta)\,\vec{y}|| \\ &= \theta\,||\vec{x}|| + (1-\theta)\,||\vec{y}|| \end{aligned}$$

Promoting low-rank structure

Toy problem: Find t such that

$$M(t) := \begin{bmatrix} 0.5 + t & 1 & 1 \\ 0.5 & 0.5 & t \\ 0.5 & 1 - t & 0.5 \end{bmatrix},$$

is low rank

Strategy: Minimize

$$f(t) := ||M(t)||$$

Matrix norms

Frobenius norm

$$||A||_{\mathsf{F}} := \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}^2} = \sqrt{\sum_{i=1}^{\min\{m,n\}} \sigma_i^2}$$

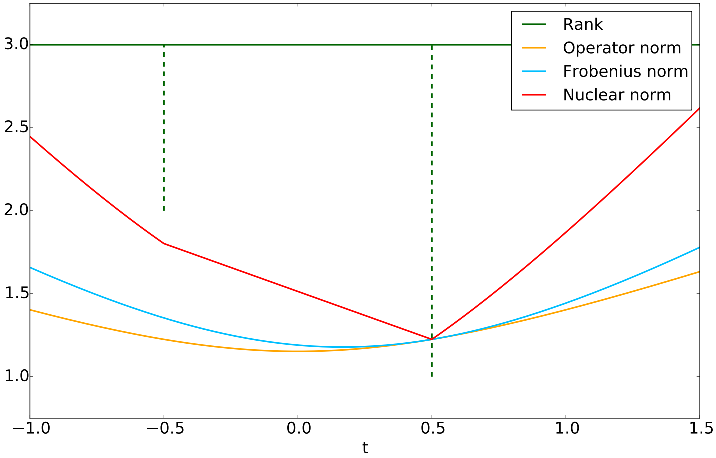
Operator norm

$$||A|| := \max_{\left\{||\vec{x}||_2=1 \mid \vec{x}\in\mathbb{R}^n\right\}} ||A\,\vec{x}||_2 = \sigma_1$$

Nuclear norm

$$||A||_* := \sum_{i=1}^{\min\{m,n\}} \sigma_i$$

Promoting low-rank structure



Exact recovery

Guarantees by Gross 2011, Candès and Recht 2008, Candès and Tao 2009

$$\min_{X \in \mathbb{R}^{m \times n}} ||X||_* \quad \text{such that } X_\Omega = y$$

achieves exact recovery with high probability as long as the number of samples is proportional to $r(n+m)$ up to log terms

Low-rank matrix estimation

If data are noisy

$$\min_{X \in \mathbb{R}^{m \times n}} \left|\left|X_{\Omega} - \vec{y}\right|\right|_2^2 + \lambda \left|\left|X\right|\right|_*$$

where $\lambda > 0$ is a regularization parameter

Matrix completion

Bob	Molly	Mary	Larry	
1	?	5	4	The Dark Knight
?	1	4	5	Spiderman 3
4	5	2	?	Love Actually
5	4	2	1	Bridget Jones's Diary
4	5	1	2	Pretty Woman
1	2	?	5	Superman 2

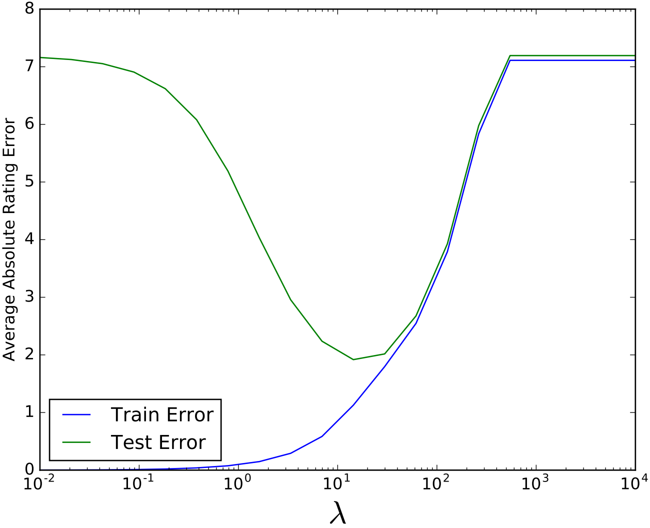
Matrix completion via nuclear-norm minimization

Bob	Molly	Mary	Larry	
1	2 (1)	5	4	The Dark Knight
2 (2)	1	4	5	Spiderman 3
4	5	2	2 (1)	Love Actually
5	4	2	1	Bridget Jones's Diary
4	5	1	2	Pretty Woman
1	2	5 (5)	5	Superman 2

Real data

- Movielens database
- 671 users
- 300 movies
- Training set: 9 135 ratings
- Test set: 1 016

Real data



Average Absolute Rating Error
0
1
2
3
4
5
6
7
8
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
10^{4}
λ
Train Error
Test Error

Low-rank matrix completion

Intractable problem

$$\min_{X \in \mathbb{R}^{m \times n}} \operatorname{rank}(X) \quad \text{such that } X_\Omega \approx \vec{y}$$

Nuclear norm: convex but computationally expensive

Alternative

- Fix rank k beforehand
- Parametrize the matrix as AB where $A \in \mathbb{R}^{m \times r}$ and $B \in \mathbb{R}^{r \times n}$
- Solve

$$\min_{\widetilde{A} \in \mathbb{R}^{m \times r}, \widetilde{B} \in \mathbb{R}^{r \times n}} \left| \left| \left(\widetilde{A}\widetilde{B} \right)_{\Omega} - \vec{y} \right| \right|_2$$

by alternating minimization

Alternating minimization

Sequence of least-squares problems (much faster than computing SVDs)

- To compute $A^{(k)}$ fix $B^{(k-1)}$ and solve

$$\min_{\widetilde{A} \in \mathbb{R}^{m \times r}} \left|\left| \left(\widetilde{A} B^{(k-1)} \right)_{\Omega} - \vec{y} \right|\right|_2$$

- To compute $B^{(k)}$ fix $A^{(k)}$ and solve

$$\min_{\widetilde{B} \in \mathbb{R}^{r \times n}} \left|\left| \left(A^{(k)} \widetilde{B} \right)_{\Omega} - \vec{y} \right|\right|_2$$

Theoretical guarantees: Jain, Netrapalli, Sanghavi 2013

Background

Low-rank models

Matrix completion

Structured low-rank models

Data-driven Analysis of Infant Sleep Patterns

Nonnegative matrix factorization

Nonnegative atoms/coefficients can make results easier to interpret

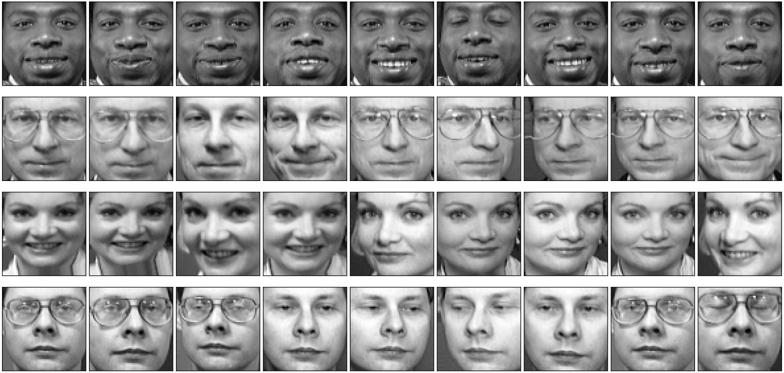
$$X \approx A\,B, \quad A_{i,j} \geq 0,\; B_{i,j} \geq 0, \text{ for all } i, j$$

Nonconvex optimization problem:

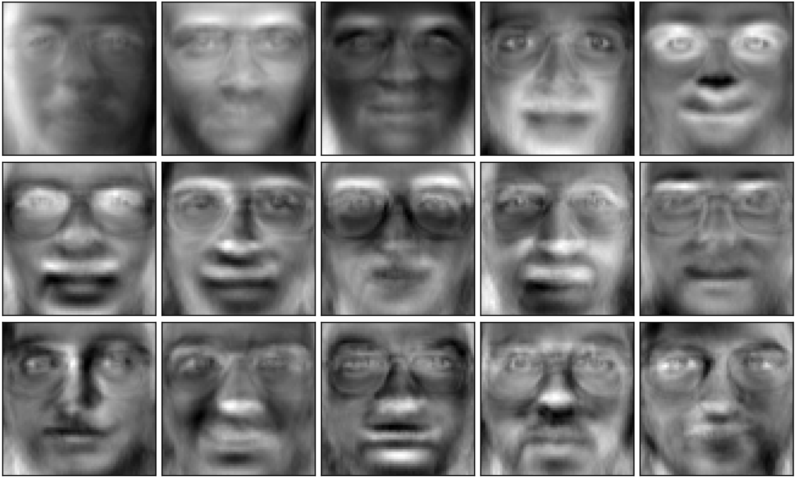
$$\begin{aligned} \text{minimize} \quad & \left|\left| X - \tilde{A}\,\tilde{B} \right|\right|_F^2 \\ \text{subject to} \quad & \tilde{A}_{i,j} \geq 0, \\ & \tilde{B}_{i,j} \geq 0, \qquad \text{for all } i, j \end{aligned}$$

$\tilde{A} \in \mathbb{R}^{m \times r}$ and $\tilde{B} \in \mathbb{R}^{r \times n}$

Face dataset



Faces dataset: Principal component analysis



Faces dataset: Nonnegative matrix factorization

Topic modeling

	singer	GDP	senate	election	vote	stock	bass	market	band	Articles
	6	1	1	0	0	1	9	0	8	a
	1	0	9	5	8	1	0	1	0	b
$A :=$	8	1	0	1	0	0	9	1	7	c
	0	7	1	0	0	9	1	7	0	d
	0	5	6	7	5	6	0	7	2	e
	1	0	8	5	9	2	0	0	1	f

SVD

$$A = USV^T = U \begin{bmatrix} 23.64 & 0 & 0 & 0 & & \\ 0 & 18.82 & 0 & 0 & 0 & 0 \\ 0 & 0 & 14.23 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3.63 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2.03 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1.36 \end{bmatrix} V^T$$

Left singular vectors

$$
\begin{array}{lcl}
 & & \quad \text{a} \qquad \text{b} \qquad \text{c} \qquad \text{d} \qquad \text{e} \qquad \text{f} \\
U_1 & = & (-0.24 \quad -0.47 \quad -0.24 \quad -0.32 \quad -0.58 \quad -0.47) \\
U_2 & = & (\ 0.64 \quad -0.23 \quad 0.67 \quad -0.03 \quad -0.18 \quad -0.21) \\
U_3 & = & (-0.08 \quad -0.39 \quad -0.08 \quad 0.77 \quad 0.28 \quad -0.40)
\end{array}
$$

Right singular vectors

	singer	GDP	senate	election	vote	stock	bass	market	band
V_1 =	-0.18	-0.24	-0.51	-0.38	-0.46	-0.34	-0.2	-0.3	-0.22
V_2 =	0.47	0.01	-0.22	-0.15	-0.25	-0.07	0.63	-0.05	0.49
V_3 =	-0.13	0.47	-0.3	-0.14	-0.37	0.52	-0.04	0.49	-0.07

Nonnegative matrix factorization

$$X \approx W\ H$$

$$W_{i,j} \geq 0,\ H_{i,j} \geq 0,\ \text{for all } i, j$$

Right nonnegative factors

	singer	GDP	senate	election	vote	stock	bass	market	band
$H_1 =$	(0.34	0	3.73	2.54	3.67	0.52	0	0.35	0.35)
$H_2 =$	(0	2.21	0.21	0.45	0	2.64	0.21	2.43	0.22)
$H_3 =$	(3.22	0.37	0.19	0.2	0	0.12	4.13	0.13	3.43)

Interpretations:

- Count atom: Counts for each doc are weighted sum of H_1, H_2, H_3
- Coefficients: They cluster words into politics, music and economics

Left nonnegative factors

$$
\begin{array}{lcllllll}
 & & \text{a} & \text{b} & \text{c} & \text{d} & \text{e} & \text{f} \\
W_1 & = & (0.03 & 2.23 & 0 & 0 & 1.59 & 2.24) \\
W_2 & = & (\ 0.1 & 0 & 0.08 & 3.13 & 2.32 & 0\ ) \\
W_3 & = & (2.13 & 0 & 2.22 & 0 & 0 & 0.03)
\end{array}
$$

Interpretations:

- Count atom: Counts for each word are weighted sum of W_1, W_2, W_3
- Coefficients: They cluster docs into politics, music and economics

Sparse PCA

Sparse atoms can make results easier to interpret

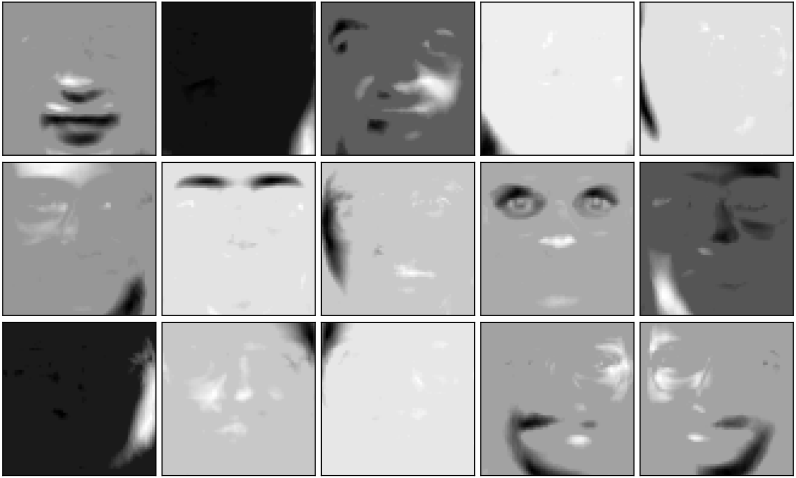
$$X \approx A\, B, \quad A \text{ sparse}$$

Nonconvex optimization problem:

$$\begin{aligned} \text{minimize} \quad & \left|\left| X - \tilde{A}\, \tilde{B} \right|\right|_2^2 + \lambda \sum_{i=1}^{k} \left|\left| \tilde{A}_i \right|\right|_1 \\ \text{subject to} \quad & \left|\left| \tilde{A}_i \right|\right|_2 = 1, \qquad 1 \leq i \leq k \end{aligned}$$

$\tilde{A} \in \mathbb{R}^{m \times r}$ and $\tilde{B} \in \mathbb{R}^{r \times n}$

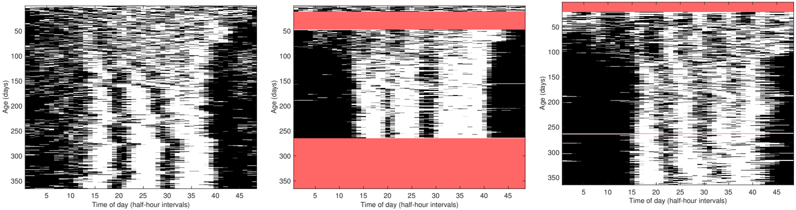
Faces dataset



Acknowledgements

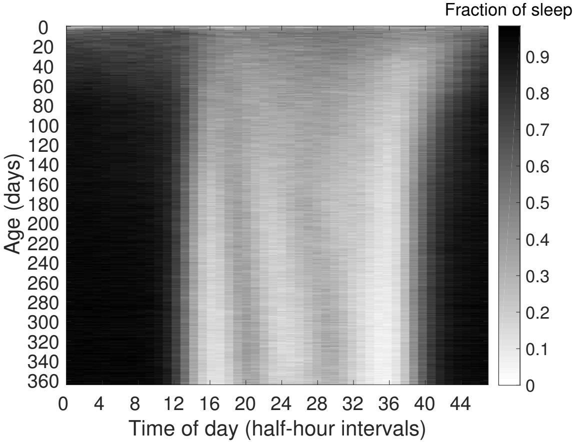
Joint work with Mark Cheng, David Heeger and Sheng Liu

Data



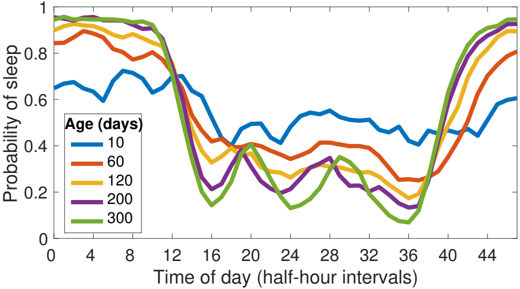
Age (days)
50
100
150
200
250
300
350
5
10
15
20
25
30
35
40
45
Time of day (half-hour intervals)
Age (days)
50
100
150
200
250
300
350
5
10
15
20
25
30
35
40
45
Time of day (half-hour intervals)
Age (days)
50
100
150
200
250
300
350
5
10
15
20
25
30
35
40
45
Time of day (half-hour intervals)

Sample mean



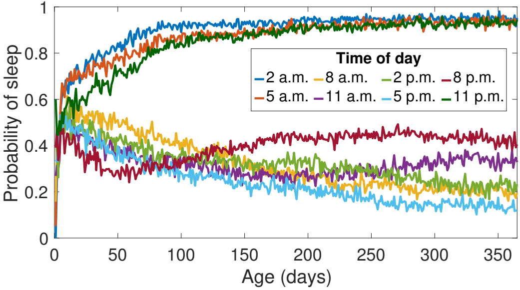
Fraction of sleep
Age (days)
0
20
40
60
80
100
120
140
160
180
200
220
240
260
280
300
320
340
360
0
4
8
12
16
20
24
28
32
36
40
44
Time of day (half-hour intervals)
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0

Sample mean



Probability of sleep
Time of day (half-hour intervals)
Age (days)
10
60
120
200
300
0
0.2
0.4
0.6
0.8
1
0
4
8
12
16
20
24
28
32
36
40
44

Sample mean

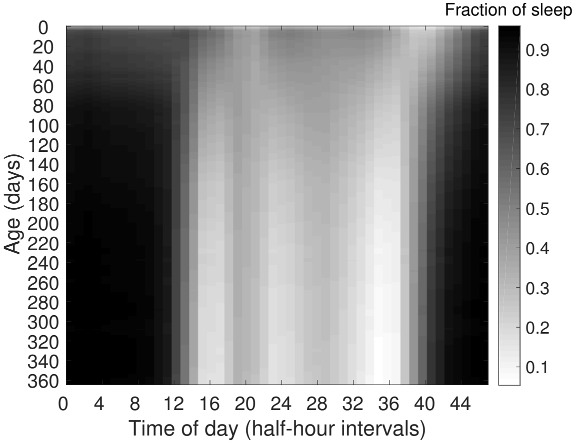


Time of day
2 a.m.
8 a.m.
2 p.m.
8 p.m.
5 a.m.
11 a.m.
5 p.m.
11 p.m.
Probability of sleep
0
0.2
0.4
0.6
0.8
1
0
50
100
150
200
250
300
350
Age (days)

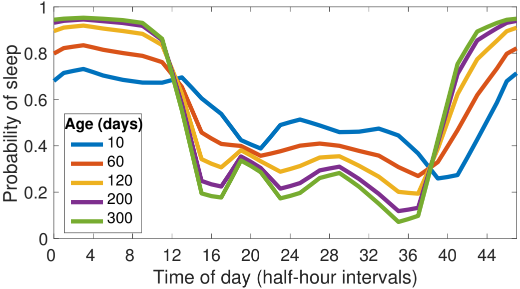
Low-rank model

$$\text{minimize} \sum_{d=1}^{365} \sum_{h=1}^{48} \sum_{b \in \mathcal{B}_{d,t}} \left(S\left(d, t, b\right) - \sum_{i=1}^{k} D_i\left(d\right) T_i\left(t\right) \right)^2$$

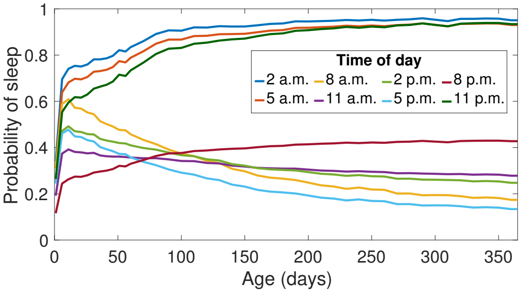
Low-rank model



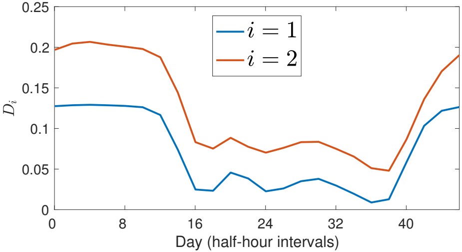
Low-rank model



Low-rank model

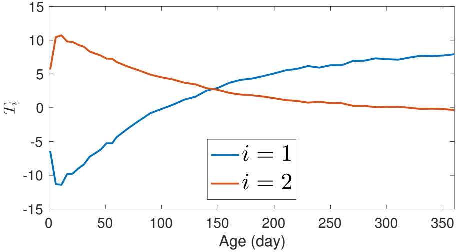


Factors



0.25
0.2
0.15
0.1
0.05
0
D_i
0
8
16
24
32
40
Day (half-hour intervals)
$i = 1$
$i = 2$

Factors

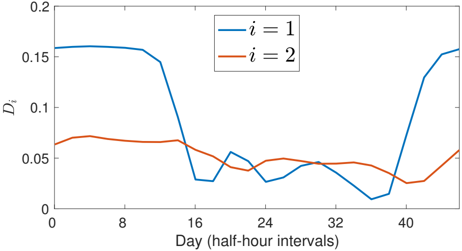


Low-rank model with nonnegative factors

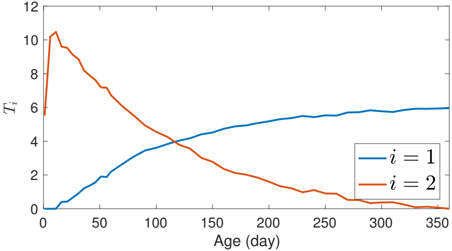
$$\text{minimize} \sum_{d=1}^{365} \sum_{h=1}^{48} \sum_{b \in \mathcal{B}_{d,t}} \left(S\left(d,t,b\right) - \sum_{i=1}^{k} D_i\left(d\right) T_i\left(t\right) \right)^2$$

$$\text{subject to } D_i\left(d\right) \geq 0, \;\; T_i\left(t\right) \geq 0 \quad \text{for all } i, d, t$$

Factors



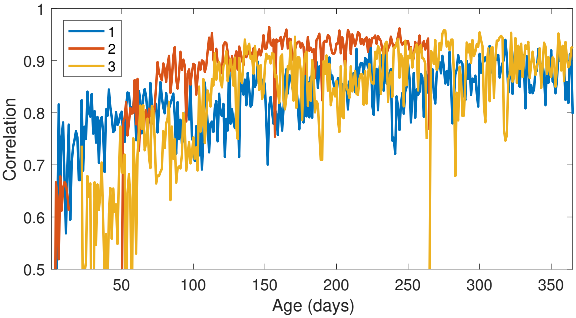
Factors



RMSE

	Mean	Low-rank model				Nonnegative low-rank model			
		k=1	k=2	k=3	k=4	k=1	k=2	k=3	k=4
Training	0.3586	0.3663	0.3596	0.3593	0.3591	0.3663	0.3596	0.3593	0.3593
Test	0.4282	0.3640	0.3585	0.3581	0.3579	0.3640	0.3585	0.3581	0.3582

Emergence of circadian rhythm



Emergence of circadian rhythm