Data Mining and Matrices

01 – Introduction

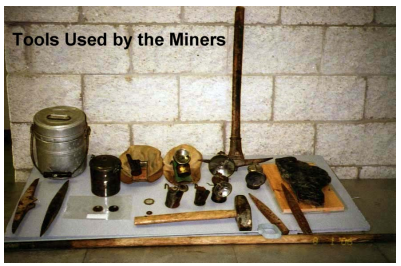
Rainer Gemulla, Pauli Miettinen

April 18, 2013

Outline

What is data mining?

"Data mining is the process of discovering knowledge or patterns from massive amounts of data." (Encyclopedia of Database Systems)

Rock	Gold	Tools	Miners
			
Data	Knowledge	Software	Analysts

Estimated $100 billion industry around managing and analyzing data.

Data, Data everywhere. The Economist, 2010.

What is data mining?

"Data mining is the process of discovering knowledge or patterns from massive amounts of data." (Encyclopedia of Database Systems)

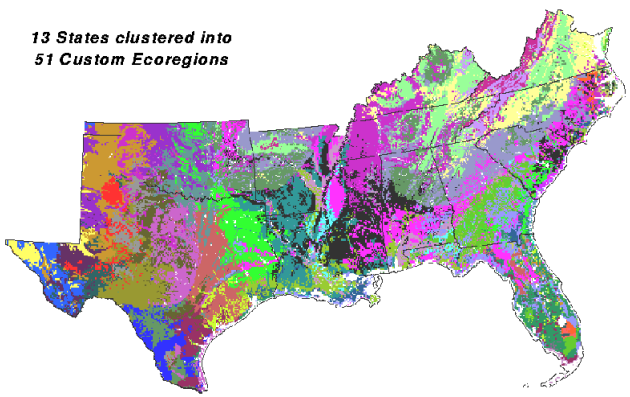
- Science
  - The Sloan Digital Sky Survey gathered 140TB of information
  - NASA Center for Climate Simulation stores 32PB of data
  - 3B base pairs exist in the human genome
  - LHC registers 600M particle collisions per second, 25PB/year
- Social data
  - 1M customer transactions are performed at Walmart per hour
  - 25M Netflix customers view and rate hundreds of thousands of movies
  - 40B photos have been uploaded to Facebook
  - 200M active Twitter users write 400M tweets per day
  - 4.6B mobile-phone subscriptions worldwide
- Government, health care, news, stocks, books, web search, ...

Data, Data everywhere. The Economist, 2010.

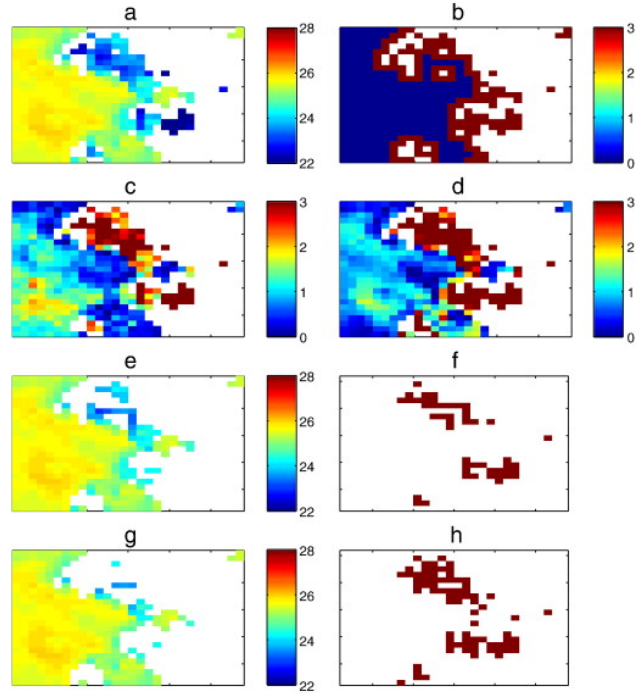
What is data mining?

"Data mining is the process of discovering knowledge or patterns from massive amounts of data." (Encyclopedia of Database Systems)

Prediction



Clustering



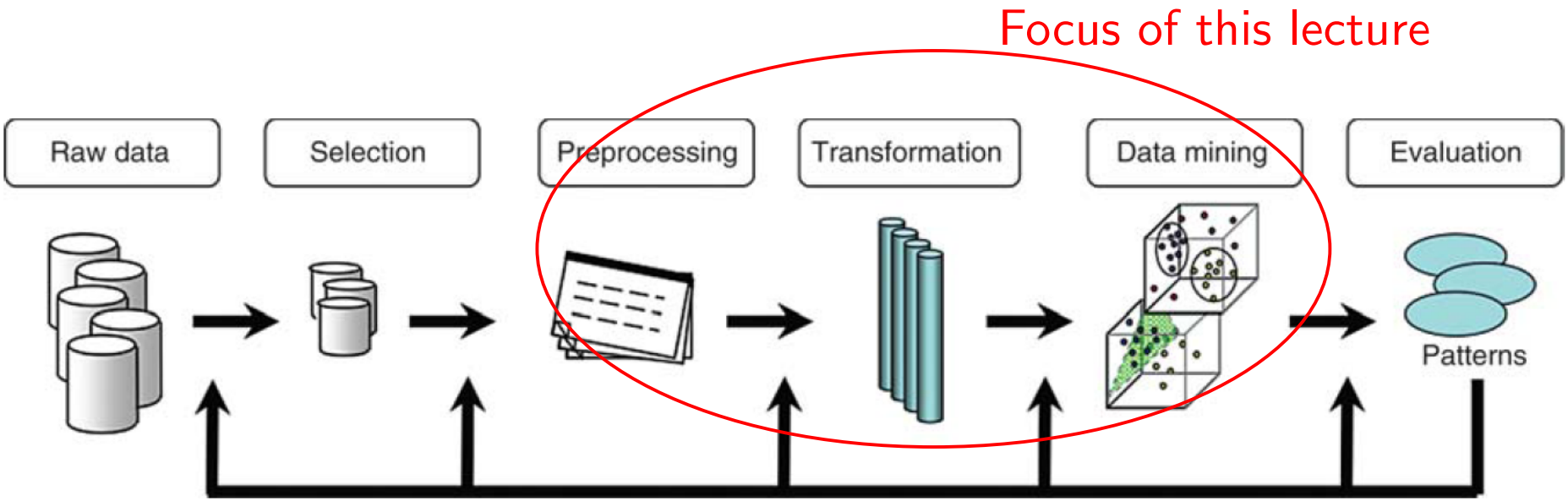
Outlier detection

"Regnet es am Siebenschläfertag, der Regen sieben Wochen nicht weichen mag." (German folklore)

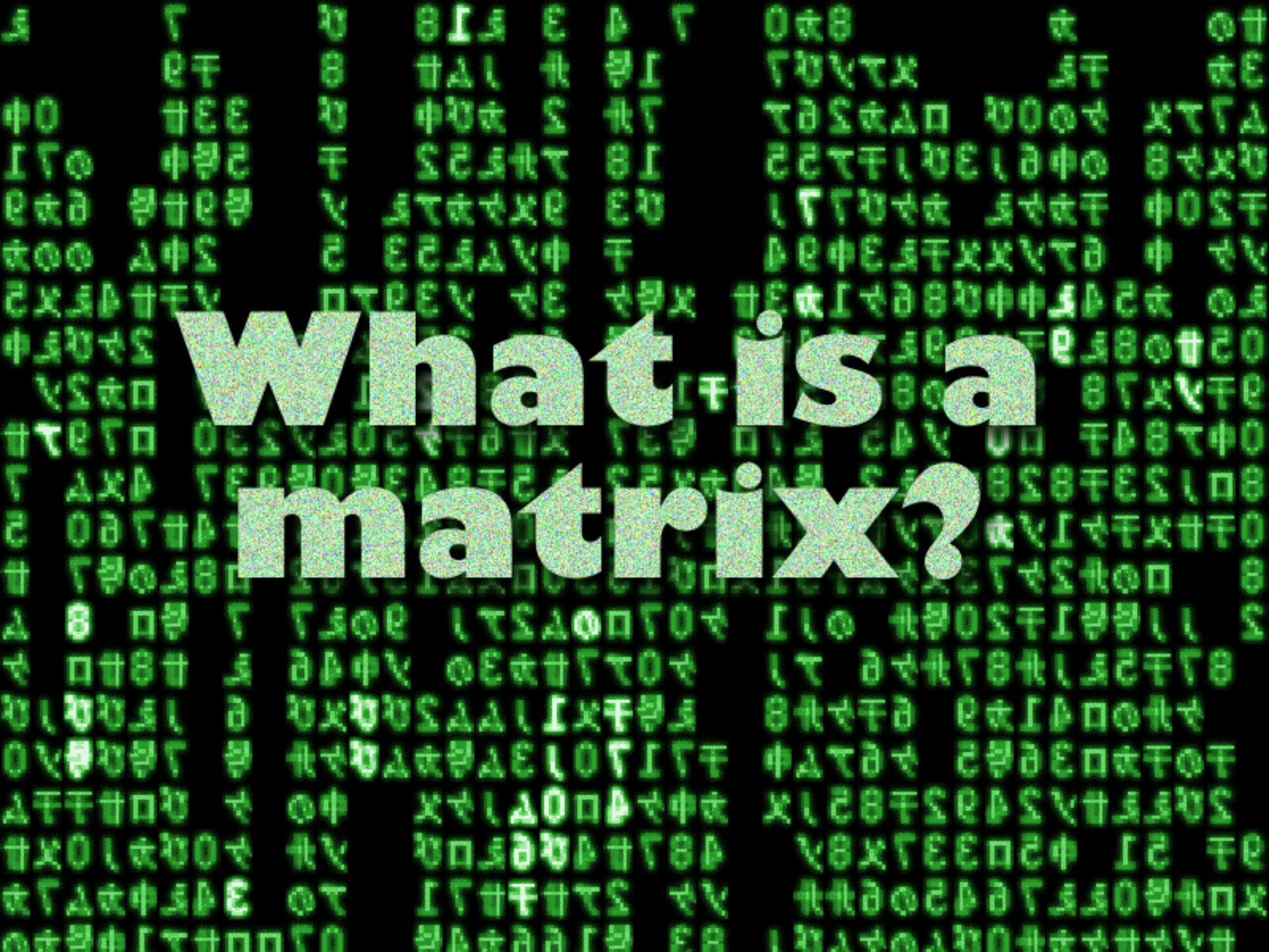
Pattern mining

What is data mining?

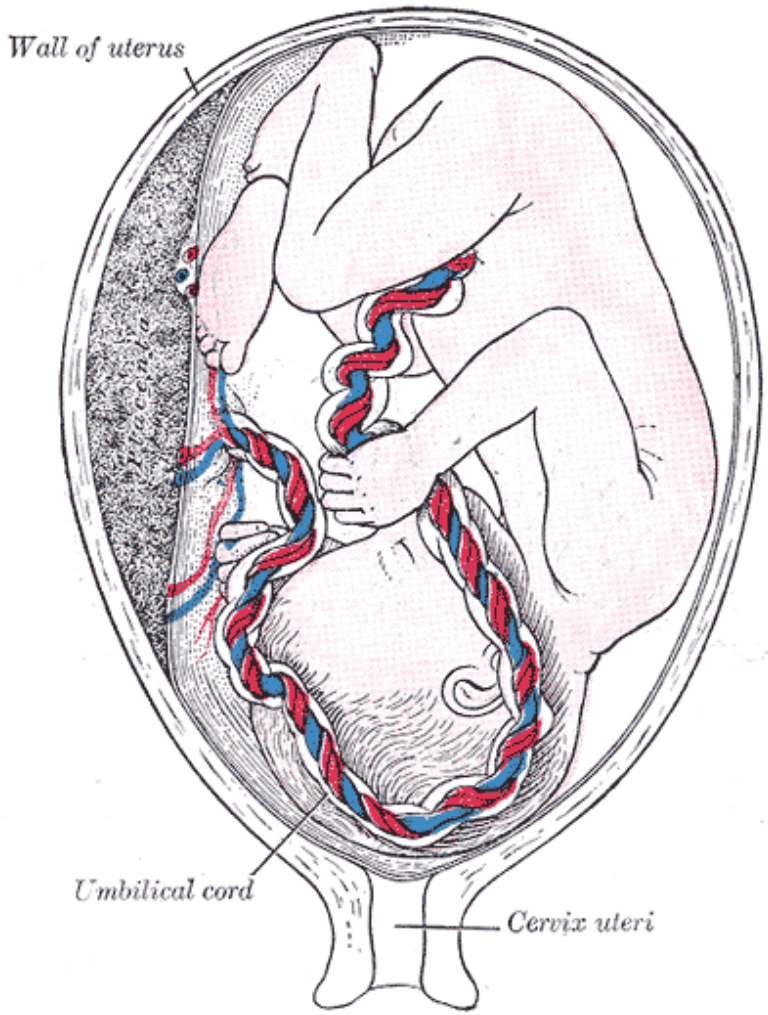
"Data mining is the process of discovering knowledge or patterns from massive amounts of data." (Encyclopedia of Database Systems)



Knowledge discovery pipeline


What is a matrix?

Womb



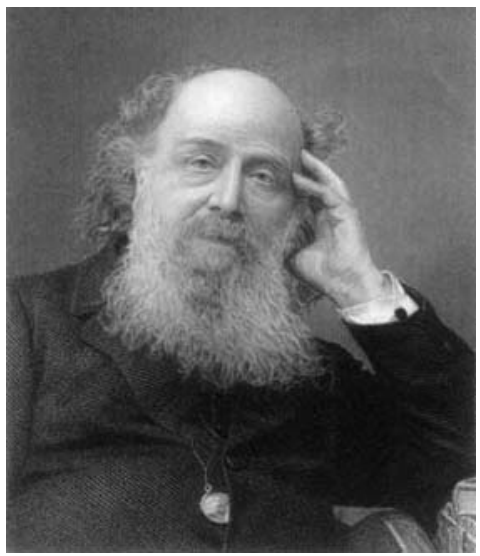
- *mater* (Latin) = *mother*
- *matrix* (Latin) = *pregnant animal*
- *matrix* (Late Latin) = *womb*
  also *source*, *origin*
- Since 1550s: *place or medium where something is developed*
- Since 1640s: *embedding or enclosing mass*

Rectangular arrays of numbers

- "Rectangular arrays" known in ancient China (*rod calculus*, estimated as early as 300BC)

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

- Term "matrix" coined by J.J. Sylvester in 1850



System of linear equations

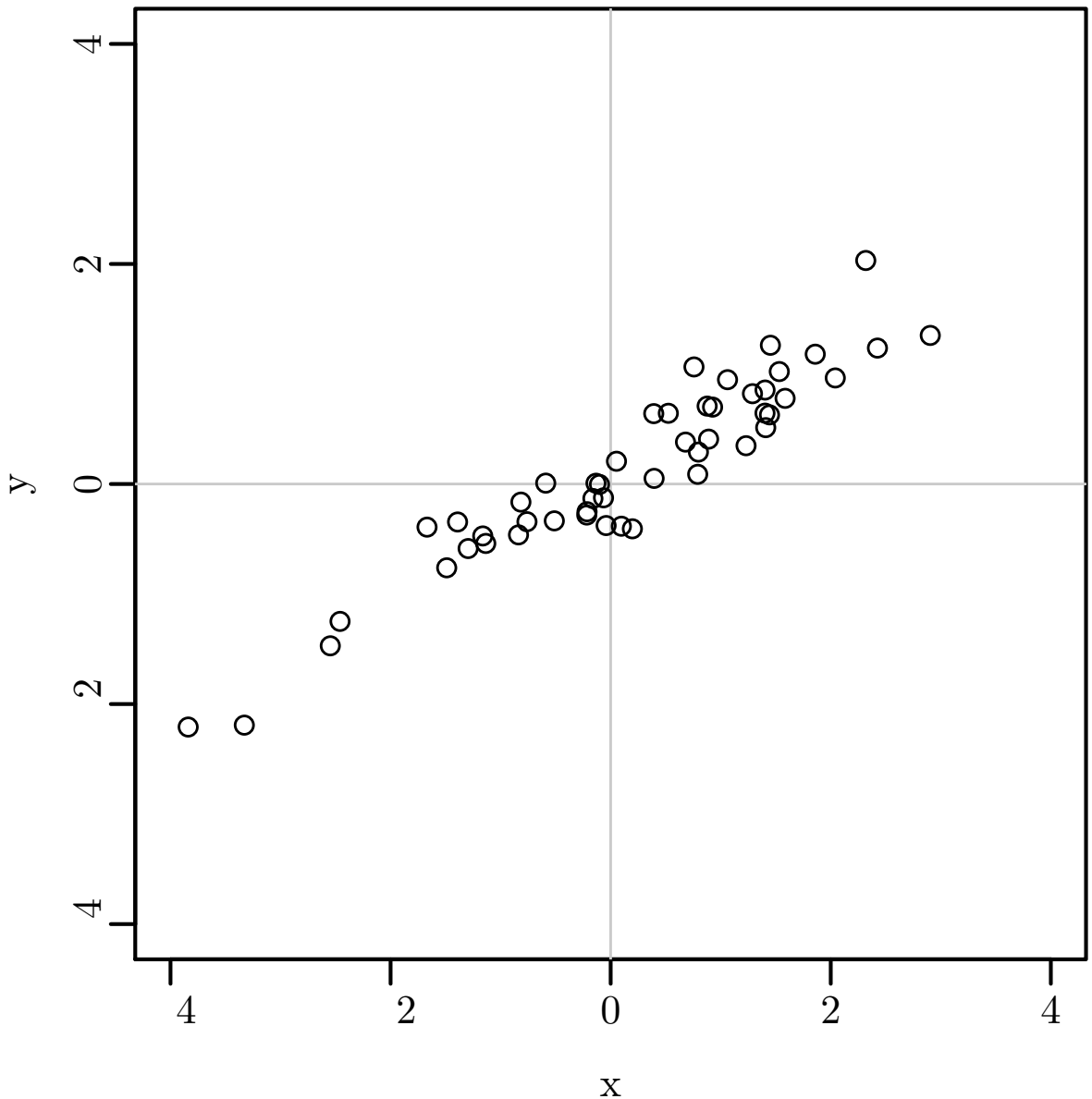
- Systems of linear equations can be written as matrices

$$\begin{aligned} 3x + 2y + z &= 39 \\ 2x + 3y + z &= 34 \\ x + 2y + 3z &= 26 \end{aligned} \quad \rightarrow \quad \left(\begin{array}{ccc|c} 3 & 2 & 1 & 39 \\ 2 & 3 & 1 & 34 \\ 1 & 2 & 3 & 26 \end{array}\right)$$

- and then be solved using linear algebra methods

$$\left(\begin{array}{ccc|c} 3 & 2 & 1 & 39 \\ & 5 & 1 & 24 \\ & & 12 & 33 \end{array}\right) \implies \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 9.25 \\ 4.25 \\ 2.75 \end{pmatrix}$$

Set of data points



$$\begin{array}{cc} x & y \\ \left(\begin{array}{cc} -3.84 & -2.21 \\ -3.33 & -2.19 \\ -2.55 & -1.47 \\ -2.46 & -1.25 \\ -1.49 & -0.76 \\ -1.67 & -0.39 \\ -1.3 & -0.59 \\ \vdots & \vdots \\ 1.59 & 0.78 \\ 1.53 & 1.02 \\ 1.45 & 1.26 \\ 1.86 & 1.18 \\ 2.04 & 0.96 \\ 2.42 & 1.24 \\ 2.32 & 2.03 \\ 2.9 & 1.35 \end{array}\right) \end{array}$$

Linear maps

- Linear maps from $\mathbb{R}^3$ to $\mathbb{R}$

$$f_1(x, y, z) = 3x + 2y + z$$
$$f_2(x, y, z) = 2x + 3y + z$$
$$f_3(x, y, z) = x + 2y + 3z$$
$$f_4(x, y, z) = x$$

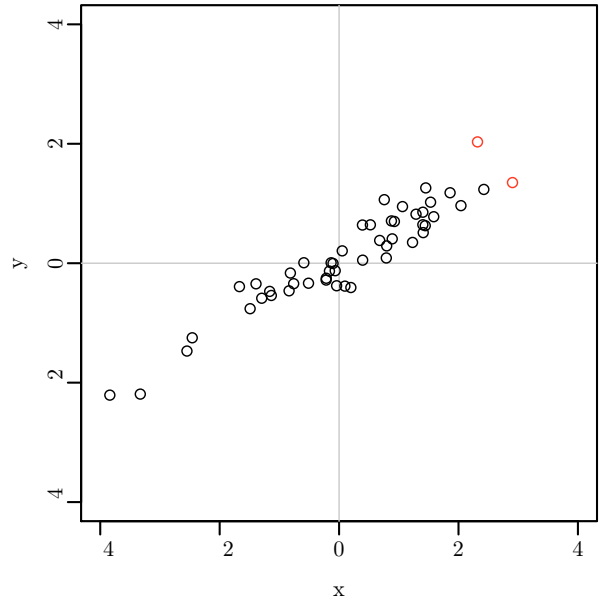
- Linear map f_1 written as a matrix

$$\begin{pmatrix} 3 & 2 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = f_1(x, y, z)$$

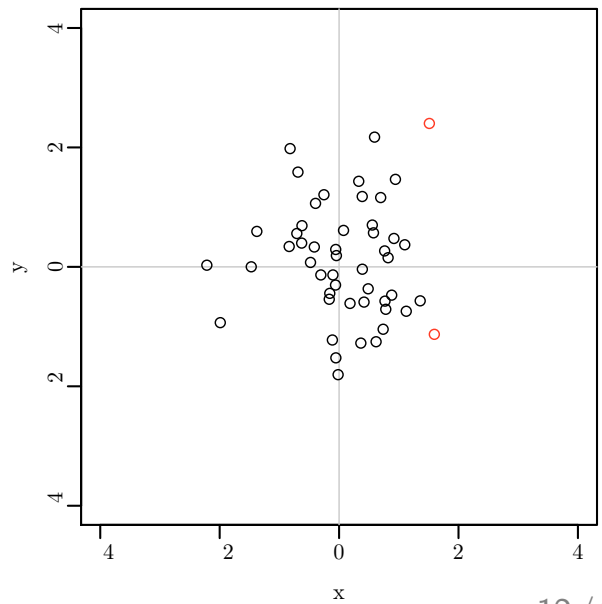
- Linear map from $\mathbb{R}^3$ to $\mathbb{R}^4$

$$\begin{pmatrix} 3 & 2 & 1 \\ 2 & 3 & 1 \\ 1 & 2 & 3 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} f_1(x, y, z) \\ f_2(x, y, z) \\ f_3(x, y, z) \\ f_4(x, y, z) \end{pmatrix}$$

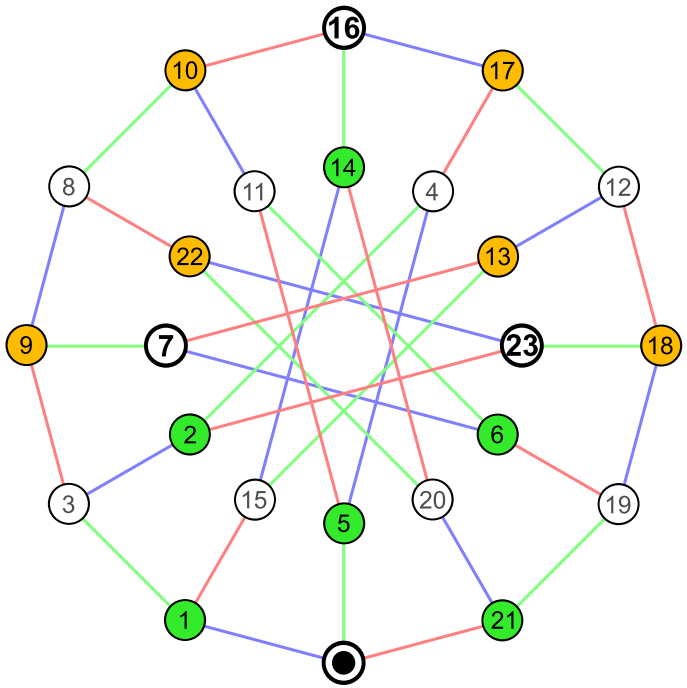
Original data

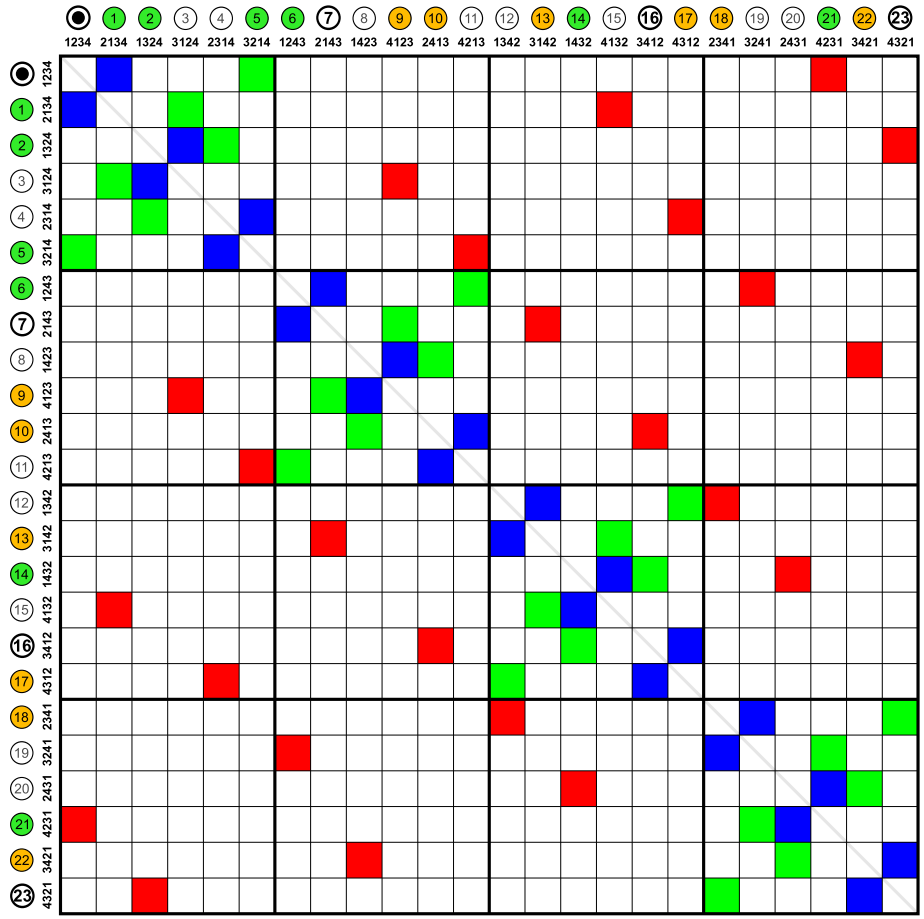


Rotated and stretched



Graphs





Adjacency matrix

Objects and attributes

Anna, Bob, and Charlie went shopping

- Anna bought butter and bread
- Bob bought butter, bread, and beer
- Charlie bought bread and beer

	Bread	Butter	Beer
Anna	1	1	0
Bob	1	1	1
Charlie	0	1	1

Customer transactions

	Data	Matrix	Mining
Book 1	5	0	3
Book 2	0	0	7
Book 3	4	6	5

Document-term matrix

	Avatar	The Matrix	Up
Alice		4	2
Bob	3	2	
Charlie	5		3

Incomplete rating matrix

	Jan	Jun	Sep
Saarbrücken	1	11	10
Helsinki	6.5	10.9	8.7
Cape Town	15.7	7.8	8.7

Cities and monthly temperatures

Many different kinds of data fit this object-attribute viewpoint.

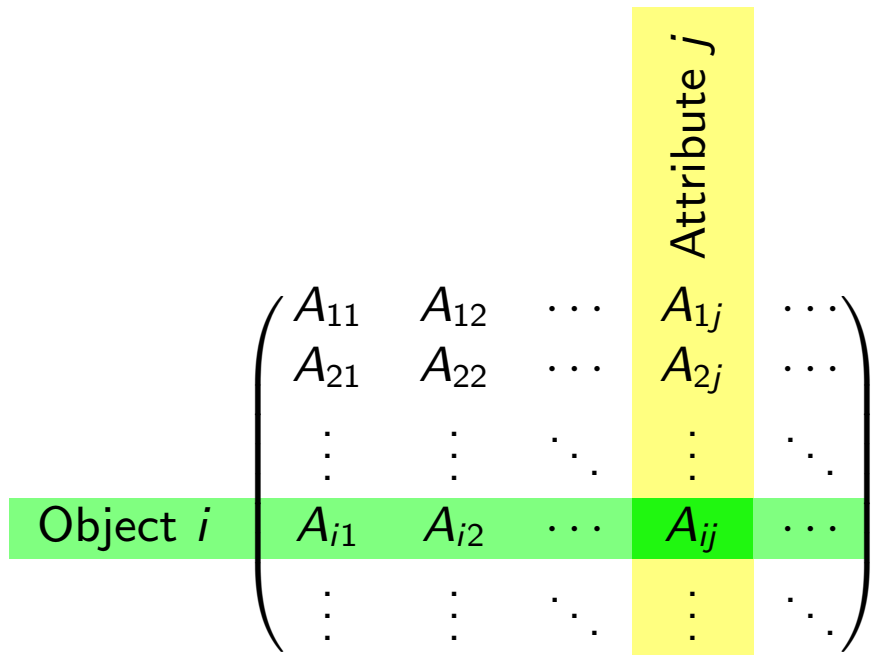
What is a matrix?

- A means to describe *computation*
  - Rotation
  - Rescaling
  - Permutation
  - Projection
  - ...

  } Linear operators

- A means to describe *data*

Rows	Columns	Entries
Objects	*Attributes*	*Values*
Equations	Variables	Coefficients
Data points	Axes	Coordinates
Vertices	Vertices	Edges
⋮	⋮	⋮



In data mining, we make use of both viewpoints simultaneously.

Outline

Key tool: Matrix decompositions

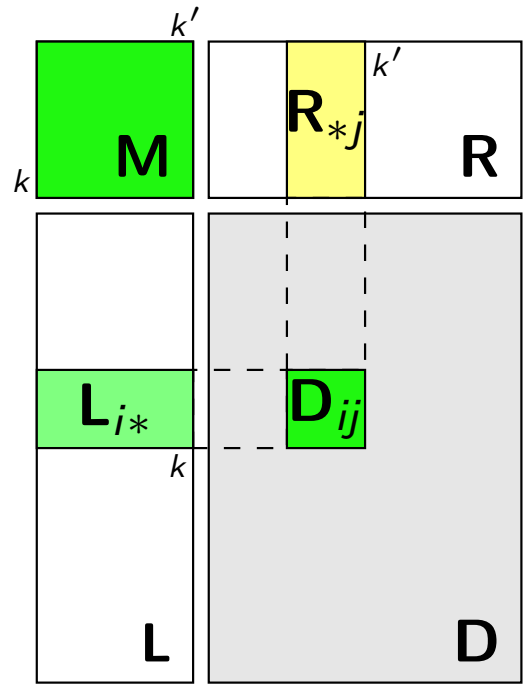
A *matrix decomposition* of a data matrix $\mathbf{D}$ is given by three matrices $\mathbf{L}$, $\mathbf{M}$, $\mathbf{R}$ such that

$$\mathbf{D} = \mathbf{LMR},$$

where

- $\mathbf{D}$ is an $m \times n$ data matrix,
- $\mathbf{L}$ is an $m \times r$ matrix,
- $\mathbf{M}$ is an $r \times r$ matrix,
- $\mathbf{R}$ is an $r \times n$ matrix, and
- r is an integer ≥ 1.

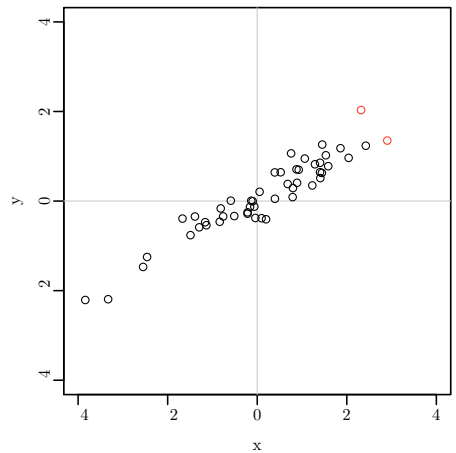
$$\mathbf{D}_{ij} = \sum_{k,k'} \mathbf{L}_{ik} \mathbf{M}_{kk'} \mathbf{R}_{k'j}$$



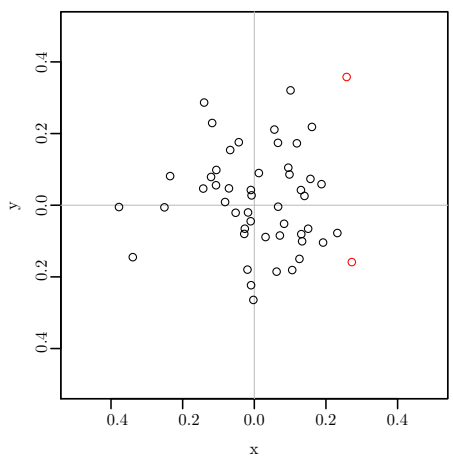
There are many different kinds of matrix decompositions, each putting certain *constraints* on matrices $\mathbf{L}$, $\mathbf{M}$, $\mathbf{R}$ (which may not be easy to find).

Example: Singular value decomposition

$\mathbf{D}_{50\times 2}$



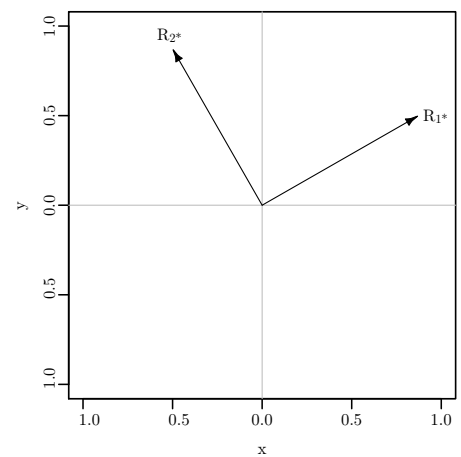
$\mathbf{L}_{50\times 2}$



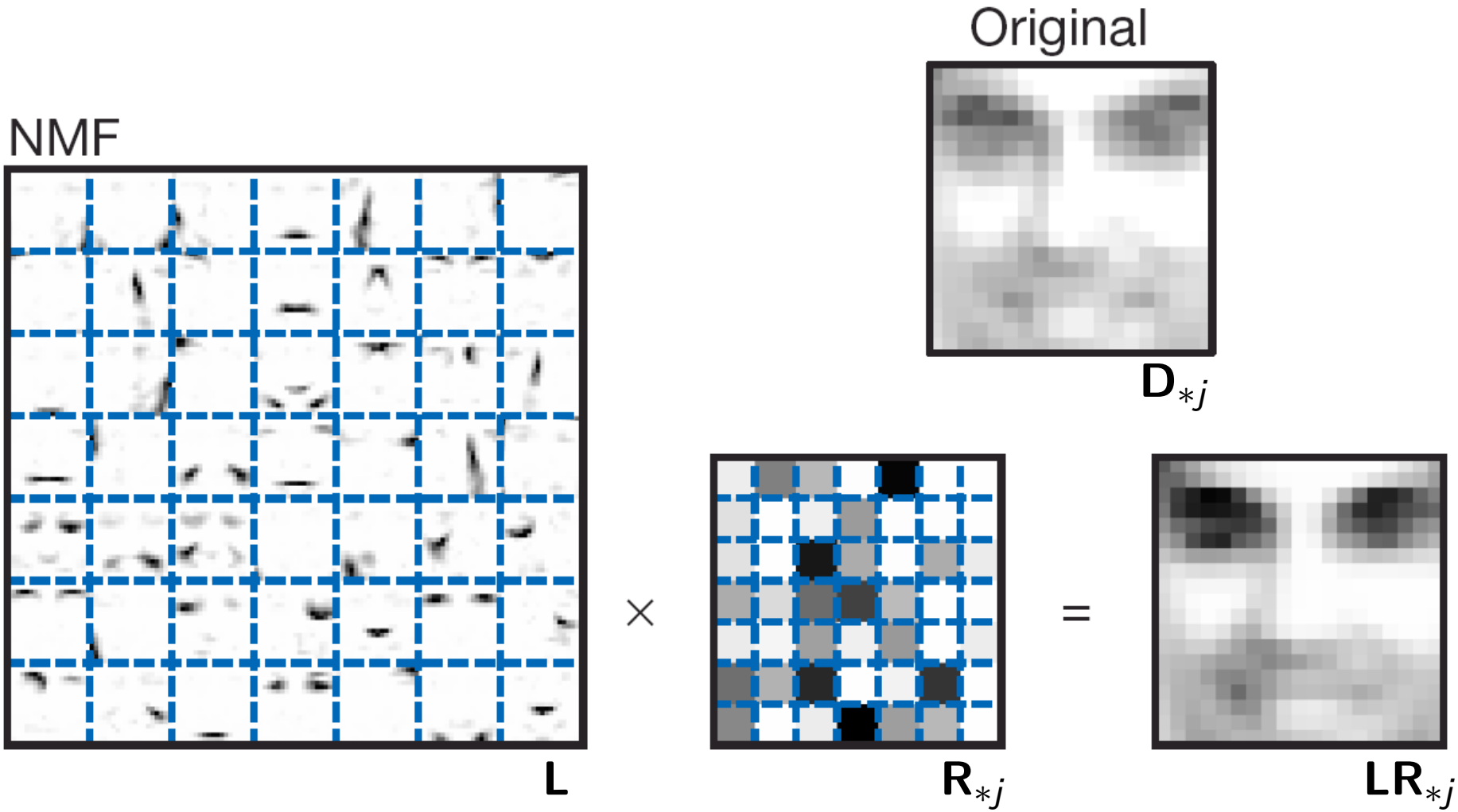
$\mathbf{M}_{2\times 2}$

$$\begin{pmatrix} 11.73 & 0 \\ 0 & 1.71 \end{pmatrix}$$

$\mathbf{R}_{2\times 2}$



Example: Non-negative matrix factorization



Lee and Seung. Learning the parts of objects by non-negative matrix factorization. Nature, 1999.

Example: Latent Dirichlet allocation

"Arts"	"Budgets"	"Children"	"Education"
NEW	MILLION	CHILDREN	SCHOOL
FILM	TAX	WOMEN	STUDENTS
SHOW	PROGRAM	PEOPLE	SCHOOLS
MUSIC	BUDGET	CHILD	EDUCATION
MOVIE	BILLION	YEARS	TEACHERS
PLAY	FEDERAL	FAMILIES	HIGH
MUSICAL	YEAR	WORK	PUBLIC
BEST	SPENDING	PARENTS	TEACHER
ACTOR	NEW	SAYS	BENNETT
FIRST	STATE	FAMILY	MANIGAT
YORK	PLAN	WELFARE	NAMPHY
OPERA	MONEY	MEN	STATE
THEATER	PROGRAMS	PERCENT	PRESIDENT
ACTRESS	GOVERNMENT	CARE	ELEMENTARY
LOVE	CONGRESS	LIFE	HAITI

R

The William Randolph Hearst Foundation will give $1.25 million to Lincoln Center, Metropolitan Opera Co., New York Philharmonic and Juilliard School. "Our board felt that we had a real opportunity to make a mark on the future of the performing arts with these grants an act every bit as important as our traditional areas of support in health, medical research, education and the social services," Hearst Foundation President Randolph A. Hearst said Monday in announcing the grants. Lincoln Center's share will be $200,000 for its new building, which will house young artists and provide new public facilities. The Metropolitan Opera Co. and New York Philharmonic will receive $400,000 each. The Juilliard School, where music and the performing arts are taught, will get $250,000. The Hearst Foundation, a leading supporter of the Lincoln Center Consolidated Corporate Fund, will make its usual annual $100,000 donation, too.

(L)

Other matrix decompositions

- Singular value decomposition (SVD)
- k-means
- Non-negative matrix factorization (NMF)
- Semi-discrete decomposition (SDD)
- Boolean matrix decomposition (BMF)
- Independent component analysis (ICA)
- Matrix completion
- Probabilistic matrix factorization
- ...

What can we do with matrix decompositions?

- Separate data from multiple processes
- Remove noise from the data
- Remove redundancy from the data
- Reveal latent structure and similarities in the data
- Fill in missing entries
- Find local patterns
- Reduce space consumption
- Reduce computational cost
- Aid visualization

Matrix decompositions can make data mining algorithms more effective. They may also provide insight into the data by themselves.

Factor interpretion of matrix decompositions

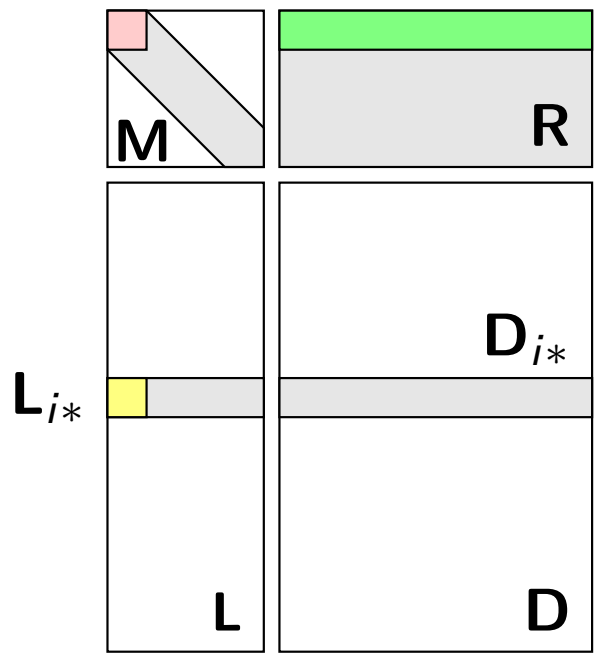
Assume that $\mathbf{M}$ is diagonal. Consider object i.

- Row of $\mathbf{R}$ = part (or piece), called *latent factor* ("latent object")
- Entry of $\mathbf{M}$ = weight of corresponding part
- Row of $\mathbf{MR}$ = weighted part
- Row of $\mathbf{L}$ = "view" of corresponding row of $\mathbf{D}$ in terms of the weighted parts (r pieces of information)
- r forces "compactness" (often $r < n$)

Each object can be viewed as a combination of r (weighted) "latent objects" (or "prototypical objects"). Similarly, each attribute can be viewed as a combination of r (weighted) "latent attributes."

(e.g., latent attribute = "body size"; latent object relates body size to real attributes such as "height", "weight", "shoe size")

$\mathbf{D}_{i*} = \sum_k \mathbf{L}_{ik}\mathbf{M}_{kk}\mathbf{R}_{k*}$



Other interpretions

- Geometric interpretation
  - Transformation of n-dimensional space in r-dimensional space
  - Row of $\mathbf{R}$ = axis
  - Row of $\mathbf{C}$ = coordinates
- Component interpretation
  - $\mathbf{D}$ is viewed as consisting of r layers (of same shape as $\mathbf{D}$)
  - k-th layer described by $\mathbf{L}_{*k}\mathbf{M}_{kk}\mathbf{R}_{k*}$
  - $\mathbf{D} = \sum_k \mathbf{L}_{*k}\mathbf{M}_{kk}\mathbf{R}_{k*}$
- Graph interpretation
  - $\mathbf{D}$ is thought of as a bipartite graph with object and attribute vertexes
  - Edge weights measure association b/w objects and attributes
  - Decomposition thought of as a tripartite graph with row, waypoint, and column vertexes

All interpretations are useful (more later).

Outline

Lessons learned

- Data mining = from data to knowledge
  $\rightarrow$ Prediction, clustering, outlier detection, local patterns
- Many different data types can be represented with a matrix
  $\rightarrow$ Linear equations, data points, maps, graphs, relational data, ...
- Common interpretation: rows = objects, columns = attributes
- Matrix decompositions reveal structure in the data
  $\rightarrow$ $\mathbf{D} = \mathbf{LMR}$
- Many different decompositions with different applications exist
  $\rightarrow$ SVD, k-means, NMF, SDD, BMF, ICA, completion, ...
- Factor interpretation: objects described by "latent attributes"

Suggested reading

- David Skillicorn
  Understanding Complex Datasets: Data Mining with Matrix Decompositions (Chapters 1–2)
  Chapman and Hall, 2007