Numerical Methods and Optimization

- Unique course (12 cfu/ects)
- Unique program
- Unique exam (oral test[s])
- 2 instructors

Webpage: http://www.di.unipi.it/˜bigig/dida/mno.html

(linked also on http://www.di.unipi.it/˜bevilacq)

News: http://twitter.com/mno_lminf

Office hours

Roberto Bevilacqua: Thursdays 16-19 pm - Room 370, Dip Informatica

email: bevilacq@di.unipi.it

Giancarlo Bigi: Thursdays 14:30-18 pm - Room 323, Dip Informatica

email: giancarlo.bigi@di.unipi.it

Not to be blocked by spam filters always write: Subject: MNO ....

Syllabus

- Linear algebra and calculus background
- Unconstrained optimization and systems of equations
- Direct and iterative methods for linear systems
- Iterative methods for nonlinear systems
- Numerical methods for unconstrained optimization
- The least-squares problem
- Iterative methods for computing eigenvalues
- Constrained optimization and systems of equations
- Numerical methods for constrained optimization
- Applications: regression, parameter estimation, approximation and data fitting
- Applications: machine learning, data mining, image and signal reconstruction
- Applications: economic equilibria and finance
- Software tools for numerical problems

Connections

Many optimization problems can be turned into [appropriate] systems of equations and/or inequalities

Vice versa:

Systems of equations and inequalities can be turned into [appropriate] optimization problems

Interactions between the corresponding solution methods

Basic bibliography

1. Demmel, *Applied Numerical Linear Algebra*, SIAM, 1997
2. Burden, Faires, *Numerical Analysis*, Brooks/Cole, 2011
3. Trefethen, Bau III, *Numerical Linear Algebra*, SIAM, 1997
4. Nocedal, Wright, *Numerical optimization*, Springer, 1999 (2nd edition, 2006)
5. Bazaraa, Sherali, Shetty, *Nonlinear Programming: Theory and Algorithms*, Wiley, 1993
6. Bertsekas, *Nonlinear programming*, Athena, 2004

   in Italian

7. Bini, Capovani, Menchi, *Metodi numerici per l'algebra lineare*, Zanichelli, 1988
8. Bevilacqua, Bini, Capovani, Menchi, *Metodi numerici*, Zanichelli, 1992
9. Grippo, Sciandrone, *Metodi di ottimizzazione non vincolata*, Springer, 2011

Selected chapters of 7 and 8 plus some lecture notes will be available on the webpage (login [*********] and password [***********] required)

Syllabus

- Linear algebra and calculus background
- Unconstrained optimization and systems of equations
- Direct and iterative methods for linear systems
- Iterative methods for nonlinear systems
- Numerical methods for unconstrained optimization
- The least-squares problem
- Iterative methods for computing eigenvalues
- Constrained optimization and systems of equations
- Numerical methods for constrained optimization
- Applications: regression, parameter estimation, approximation and data fitting
- Applications: machine learning, data mining, image and signal reconstruction
- Applications: economic equilibria and finance
- Software tools for numerical problems

Syllabus

- Linear algebra and calculus background
- Unconstrained optimization and systems of equations
- Direct and iterative methods for linear systems
- Iterative methods for nonlinear systems
- Numerical methods for unconstrained optimization
- The least-squares problem
- Iterative methods for computing eigenvalues
- Constrained optimization and systems of equations
- Numerical methods for constrained optimization
- Applications: regression, parameter estimation, approximation and data fitting
- Applications: machine learning, data mining, image and signal reconstruction
- Applications: economic equilibria and finance
- Software tools for numerical problems

Syllabus

- Linear algebra and calculus background
- Unconstrained optimization and systems of equations
- Direct and iterative methods for linear systems
- Iterative methods for nonlinear systems
- Numerical methods for unconstrained optimization
- The least-squares problem
- Iterative methods for computing eigenvalues
- Constrained optimization and systems of equations
- Numerical methods for constrained optimization
- Applications: regression, parameter estimation, approximation and data fitting
- Applications: machine learning, data mining, image and signal reconstruction
- Applications: economic equilibria and finance
- Software tools for numerical problems

Evaluating scientific journals

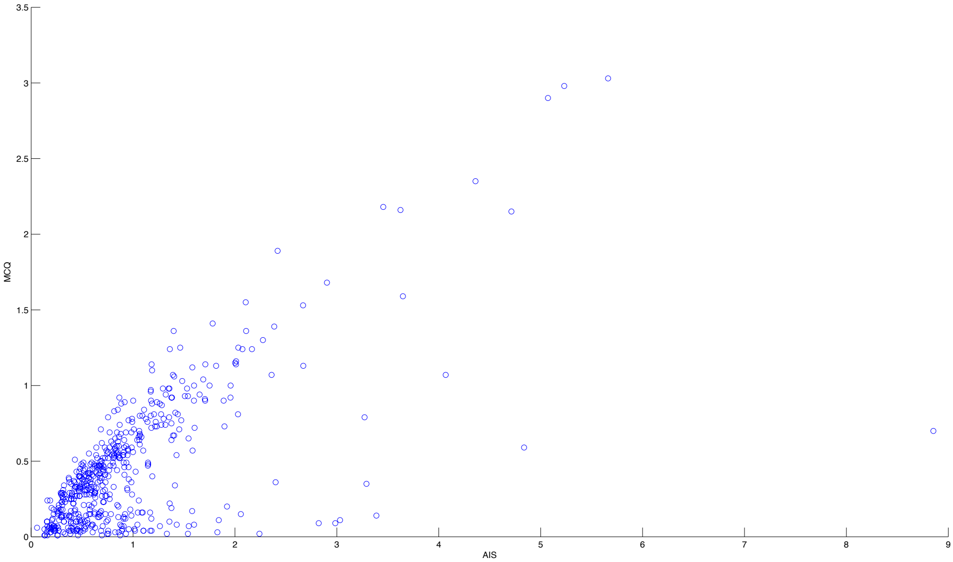
Two quality indices for mathematical journals:

- Article Influence Score [ISI]
- Mathematical Citation Quotient [AMS]

Are the two indices actually related?

$n = 509$ journals: $(AIS_i, MCQ_i) \qquad i = 1, ..., n$

journal data



Evaluating scientific journals

Two quality indices for mathematical journals:

- Article Influence Score [ISI]
- Mathematical Citation Quotient [AMS]

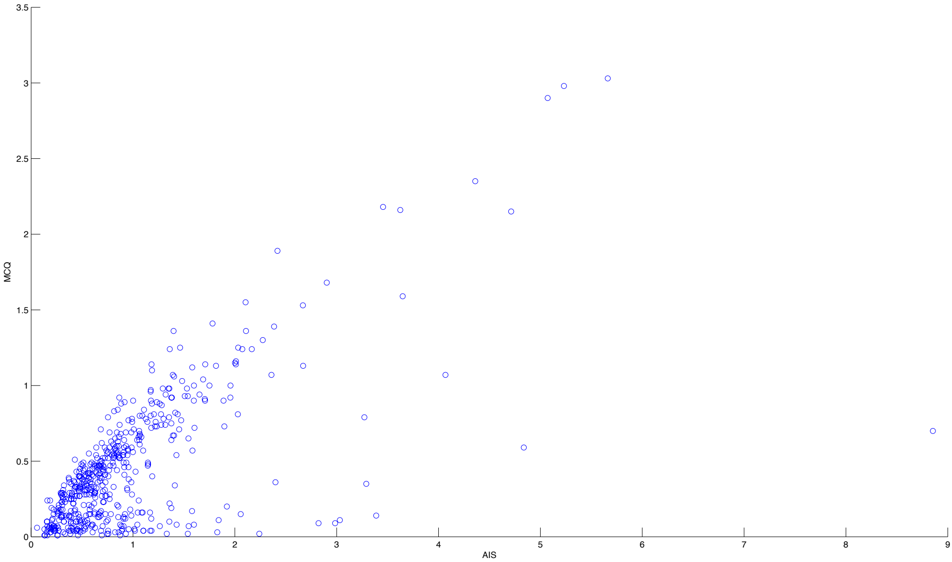
Are the two indices actually related? $MCQ \approx p\, AIS + q$?

$n = 509$ journals: $(AIS_i, MCQ_i) \qquad i = 1, ..., n$

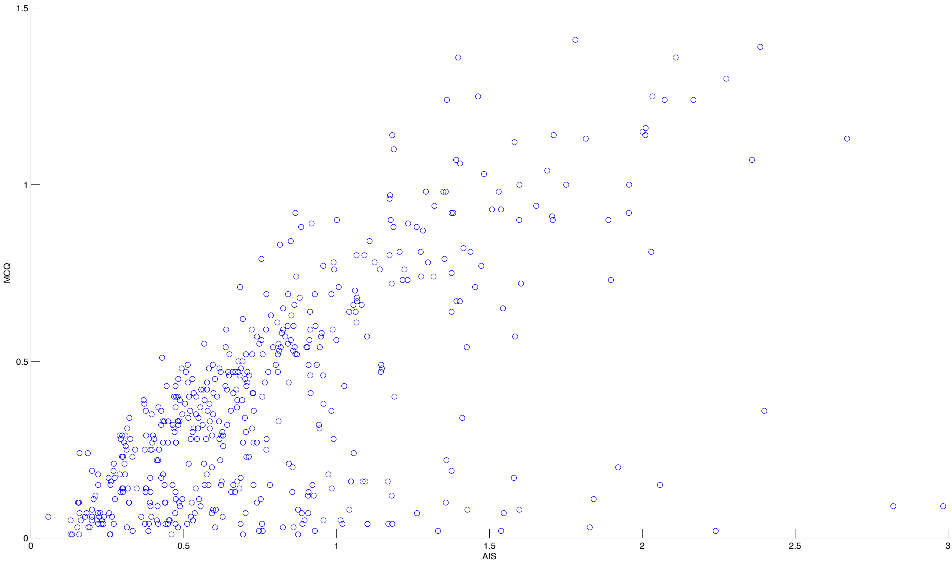
Linear least squares:

$$\min \left\{ \sum_{i=1}^{n} (MCQ_i - p\, AIS_i - q)^2 \;:\; p, q \in \mathbb{R} \right\}$$

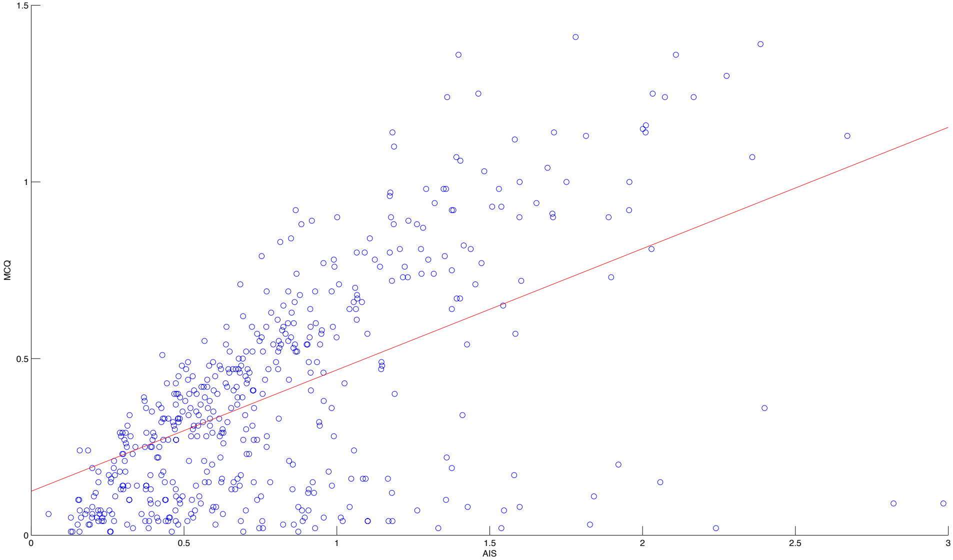
journal data



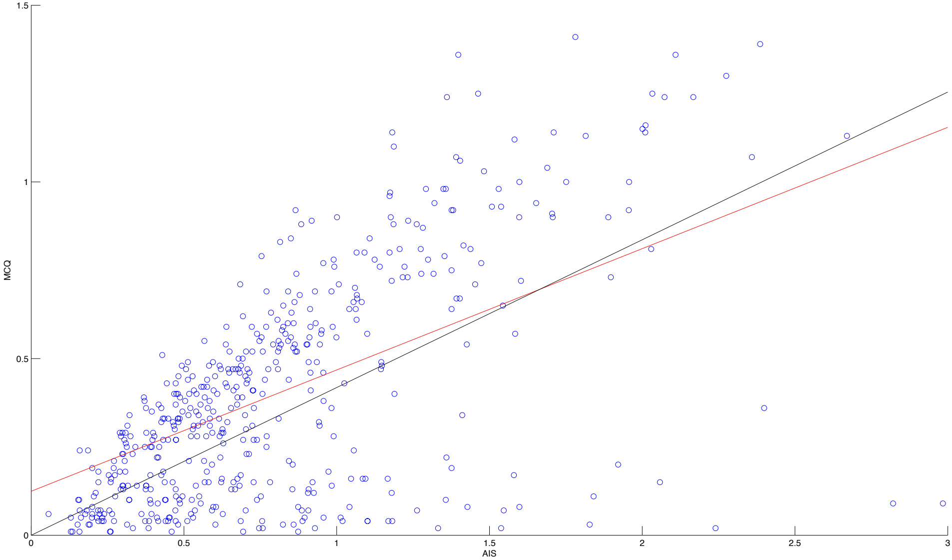
journal data



data fitting



data fitting with $q = 0$



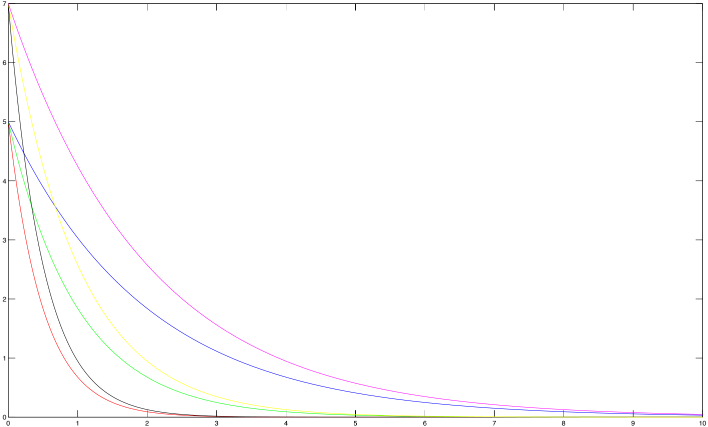
Exponential decay

$$y(t) = \theta e^{-\lambda t}$$

- physics
- chemistry
- pharmacology
- economics
- computer science (BGP: routing flapping damping)

Exponential decay

$$y(t) = \theta e^{-\lambda t}$$

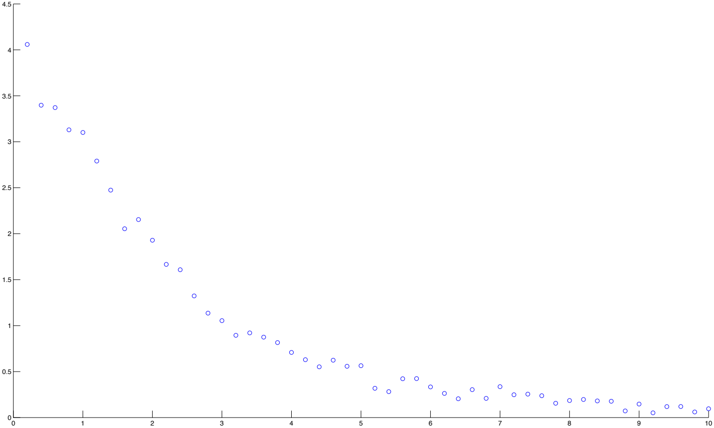


$\theta = 7 : \lambda = 2$ (black) $\lambda = 1$ (yellow) $\lambda = 0.5$ (magenta)

$\theta = 5 : \lambda = 2$ (red) $\lambda = 1$ (green) $\lambda = 0.5$ (blue)

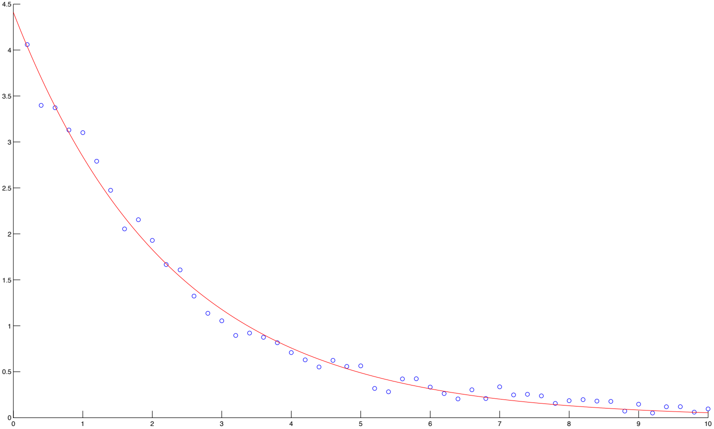
Parameter estimation: observations

experimental observations: $(t_i, y_i) \qquad i = 1, ..., n$



Parameter estimation: curve fitting

least squares: $\min \left\{ \sum_{i=1}^{n} (y_i - \theta e^{-\lambda t_i})^2 \ : \ \theta, \lambda \geq 0 \right\}$



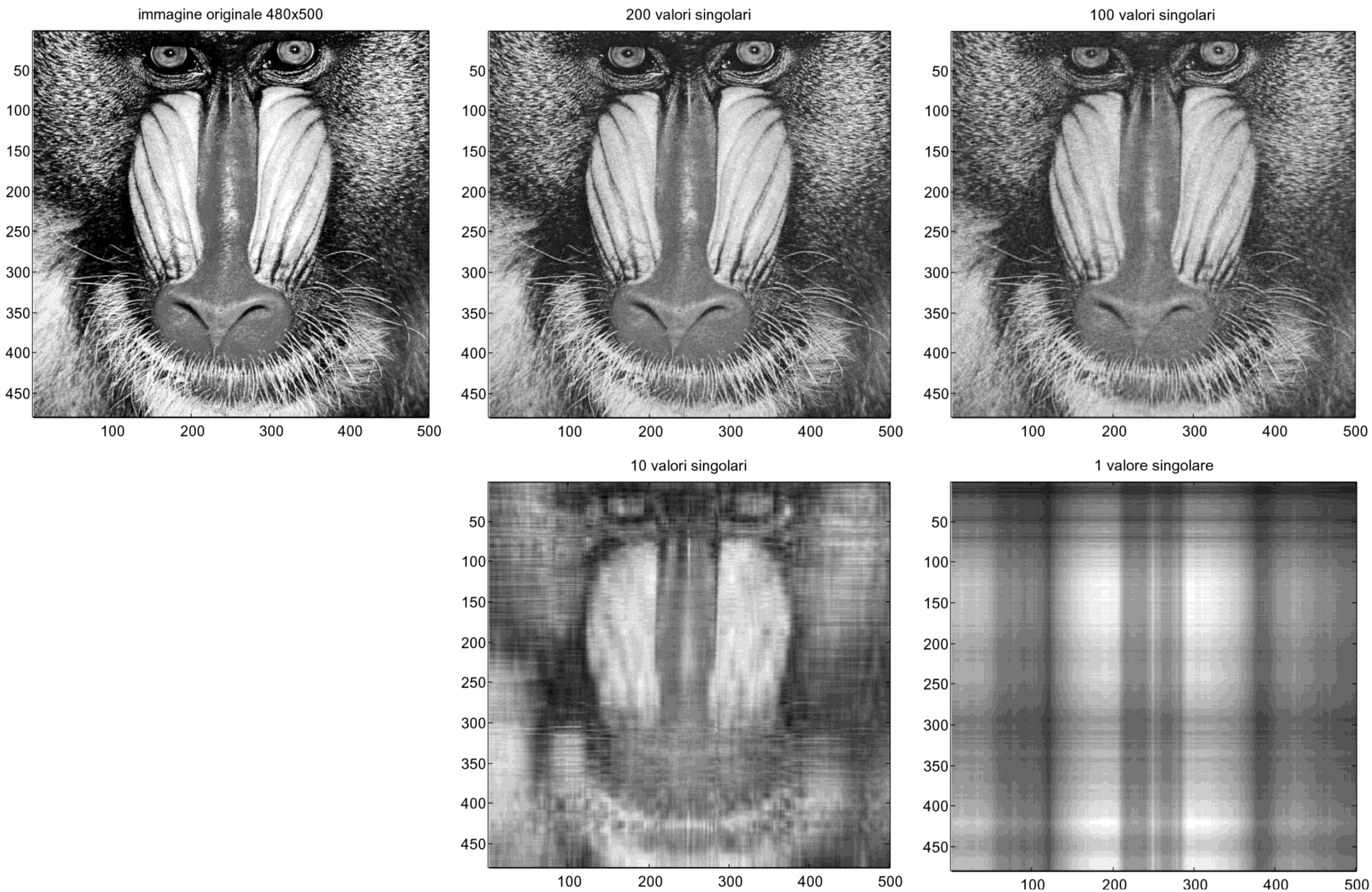
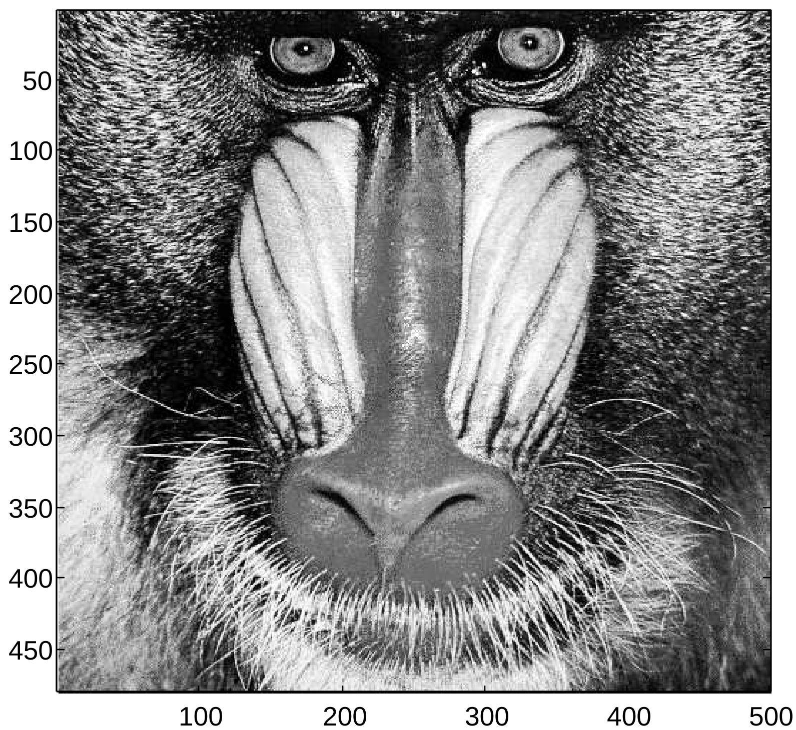
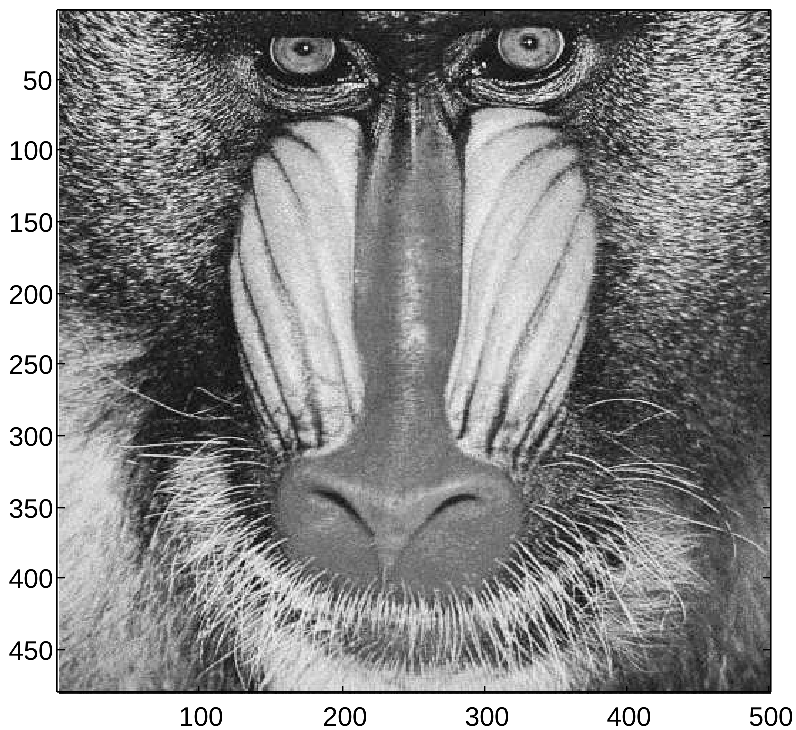


immagine originale 480x500
200 valori singolari
100 valori singolari
10 valori singolari
1 valore singolare
50
100
150
200
250
300
350
400
450
100
200
300
400
500

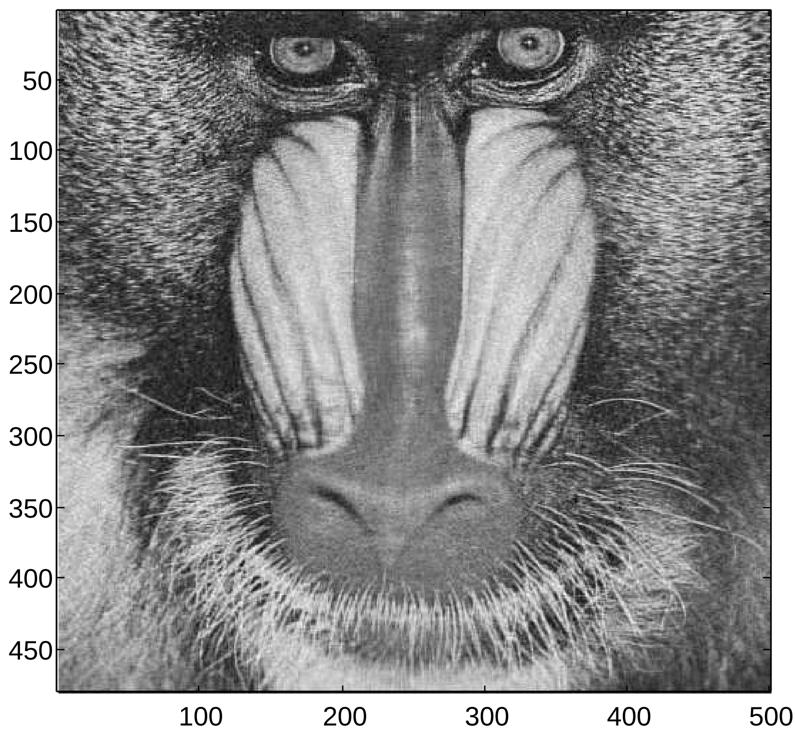
immagine originale 480x500

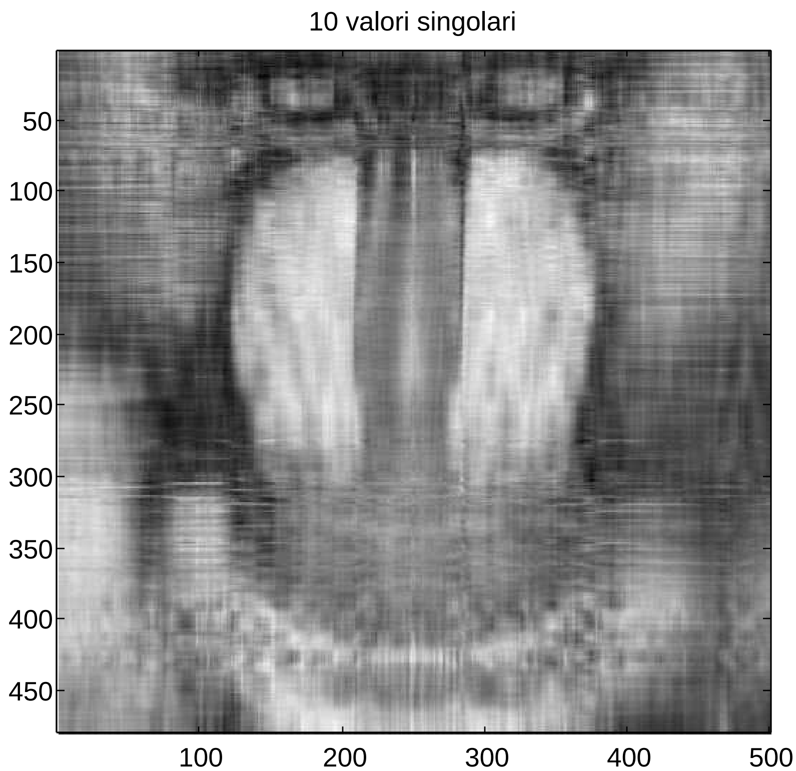


200 valori singolari

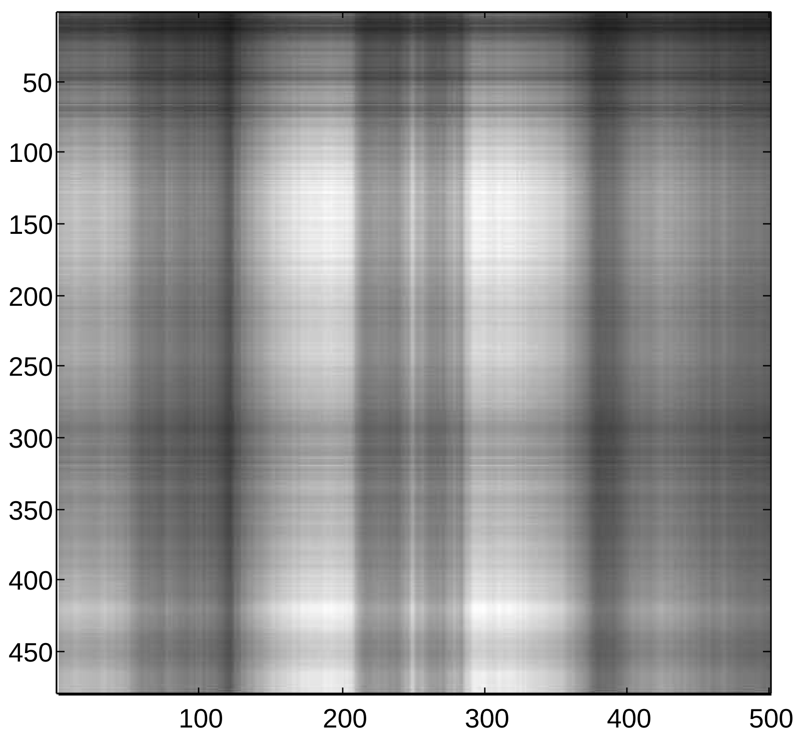


100 valori singolari




10 valori singolari
50
100
150
200
250
300
350
400
450
100
200
300
400
500

1 valore singolare



Supervised classification

$\{x_1, ..., x_n\}, \{y_1, ..., y_m\} \subseteq \mathbb{R}^n$: x_i "positive" data, y_j "negative" data

Is it possible to discriminate the two sets?

Classification: find $f : \mathbb{R}^n \to \mathbb{R}$ such that

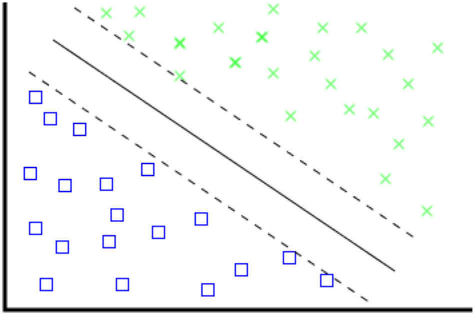
- $f(x_i) > 0 \qquad i = 1, ..., n$
- $f(y_j) < 0 \qquad j = 1, ..., m$

Search space: $f(x) = a^T x - b \qquad a \in \mathbb{R}^n, b \in \mathbb{R}$

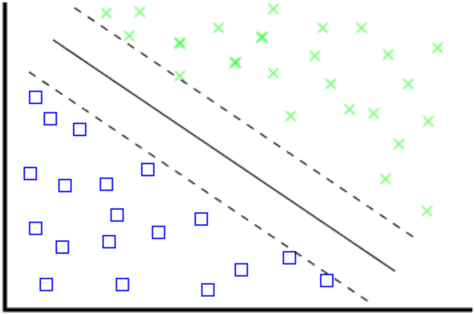
Linear classification: find $a \in \mathbb{R}^n, b \in \mathbb{R}$ such that

- $a^T x_i > b \qquad i = 1, ..., n$
- $a^T y_j < b \qquad j = 1, ..., m$

Supervised classification



Classification margin



Maximize the classification margin t such that

- $a^T x_i \geq b + t \qquad i = 1, ..., n$
- $a^T y_j \leq b - t \qquad j = 1, ..., m$

Classification margin

$$\max \quad t$$
$$a^T x_i - b \geq t \qquad i = 1, ..., n$$
$$a^T y_j - b \leq -t \qquad j = 1, ..., m$$
$$||a||_2 \leq 1$$

If the optimal value is positive, that is the two sets can be actually discriminated, (just setting $a' = a/t$ and $b' = b/t$) the problem is equivalent to:

$$\min \quad ||a'||_2$$
$$a'^T x_i - b' \geq 1 \qquad i = 1, ..., n$$
$$a'^T y_j - b' \leq -1 \qquad j = 1, ..., m$$

The Google Matrix

September 20, 2012

Page Rank: the relevance of a web page

Problema:

How can all the pages in the web be ordered according to their relevance?

How can the relevance of a web page (page rank) be defined?

A possible choice for a computational model:

the relevance of a page depends on the relevance of the pages pointing to the page itself.

Page Rank: the relevance of a web page

Problema:

How can all the pages in the web be ordered according to their relevance?

How can the relevance of a web page (page rank) be defined?

A possible choice for a computational model:

the relevance of a page depends on the relevance of the pages pointing to the page itself.

The mathematical model

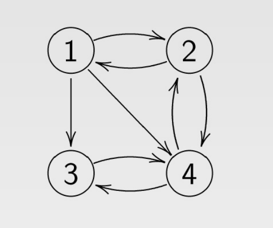
Number all web pages from 1 up to n.
Define the adjacency matrix H:

$$H = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1n} \\ h_{21} & h_{22} & \cdots & h_{2n} \\ \vdots & \vdots & & \vdots \\ h_{n1} & h_{n2} & \cdots & h_{nn} \end{bmatrix},$$

$h_{ij} = 1$ if page i contains at least one link to page j,
$h_{ij} = 0$ otherwise.

An example

For $n = 4$, consider the web set described by the following graph and by the adjacency matrix H:



$$H = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}$$

The mathematical model (continued)

Page i transmits its relevance to the pages it points to, which are as many as the nonzero entries in the i-th row of H.
The relevance inherited from page i has to be uniformly distributed over all the pages linked by page i:

$$\text{the link } i \to j \text{ is weighted by } 1/r_i,$$

where r_i is the total number of links present in page i:

$$r_i = \sum_{j=1}^{n} h_{ij}$$

An example (continued)

From matrix H the following matrix is obtained by row-scaling:

$$\text{diag}\,(r_1^{-1}\ r_2^{-1}\ r_3^{-1}\ r_4^{-1})H = \begin{bmatrix} 0 & 1/3 & 1/3 & 1/3 \\ 1/2 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 1 \\ 0 & 1/2 & 1/2 & 0 \end{bmatrix}.$$

In general, all the rows of the matrix $\text{diag}(r_1^{-1} \cdots r_n^{-1})H$ have sum 1. A matrix satisfying this property is called row stochastic.

A (row or column) stochastic matrix has eigenvalue 1.

The mathematical model (continued)

If x_j is the rank of page j, then the following equations hold:

$$x_j = \frac{h_{1j}}{r_1} x_1 + \frac{h_{2j}}{r_2} x_2 + \cdots + \frac{h_{nj}}{r_n} x_n, \quad \text{for } j = 1, 2, \ldots, n,$$

that is

$$\mathbf{x}^T = \mathbf{x}^T DH,$$

where $D = \text{diag}(r_1^{-1} \cdots r_n^{-1})$.
This is a linear system of equations. Is it consistent?

The mathematical model (continued)

The linear system can be rewritten in the following forms:

$$\mathbf{x} = H^T D \mathbf{x},$$

$$(H^T D - I)\mathbf{x} = \mathbf{0}.$$

The matrix $H^T D - I$ is singular, since $H^T D$ is column stochastic, therefore has eigenvalue 1. It can be shown that

1 is a dominant eigenvalue , that is $\rho(H^T D) = 1$,
1 admits nonnegative eigenvectors.

The mathematical model (continued)

In the space of solutions of $(H^T D - I)\mathbf{x} = \mathbf{0}$ let us consider the nonegative ones, obeying the condition $||x||_1 = 1$

(if the space has dimension 1, then this solution is unique).

The list of pages ordered by rank is obtained by rearranging the entries of an eigenvector of $H^T D$.

The mathematical model (continued)

The mathematical model can be improved.

The entries of DH can be seen as probability values:

h_{ij}/r_i is the probability that the surfer

jumps from page i to page j.

It is allowed that, under a small prescribed probability value, the surfer jumps from page i to any othe page. In the model:

the matrix DH is replaced by the matrix $G = cDH + \frac{1-c}{n}E$,

where E is the matrix with all entries 1, and $0 < c < 1$

An example (continued)

In the example given above

$$G = \begin{bmatrix} (1-c)/4 & c/3+(1-c)/4 & c/3+(1-c)/4 & c/3+(1-c)/4 \\ c/2+(1-c)/4 & (1-c)/4 & (1-c)/4 & 1/2+(1-c)/4 \\ (1-c)/4 & (1-c)/4 & (1-c)/4 & 1+(1-c)/4 \\ (1-c)/4 & 1/2+(1-c)/4 & 1/2+(1-c)/4 & (1-c)/4 \end{bmatrix}.$$

Usually the value $c = 0.85$ is chosen.

The mathematical model (continued)

The matrix G:

- is row stochastic;
- is irreducible;
- has eigenvalue 1 with multiplicity 1, and the normalized corresponding eigenvector positive.

Moreover, the matrix E can be modified in such a way that personalized preferences are introduced.

The mathematical model (continued)

The linear system is now:

$$\mathbf{x} = G^T\mathbf{x} = (cH^TD + \frac{1-c}{n}E)\mathbf{x} = cH^TD\mathbf{x} + \frac{1-c}{n}\mathbf{e},$$

where $\mathbf{e}$ is the vector with all entries 1.

$$G = \begin{bmatrix} 0.0375 & 0.3208 & 0.3208 & 0.3208 \\ 0.4625 & 0.0375 & 0.0375 & 0.4625 \\ 0.0375 & 0.0375 & 0.0375 & 0.8875 \\ 0.0375 & 0.4625 & 0.4625 & 0.0375 \end{bmatrix}.$$

Iterations of the power method applied to G^T:

0.2500	0.1437	0.1287	0.1479	0.1325	0.1444	0.1352	0.1423	0.1368	0.1411
0.2500	0.2146	0.2597	0.2235	0.2516	0.2299	0.2466	0.2337	0.2437	0.2360
0.2500	0.2146	0.2597	0.2235	0.2516	0.2299	0.2466	0.2337	0.2437	0.2360
0.2500	0.4271	0.3518	0.4051	0.3644	0.3958	0.3715	0.3903	0.3758	0.3870

The pages rearranged according to the rank are in this order: 4, 2 e 3 (ex aequo), 1.

Molecular configuration

Lennard-Jones potential:

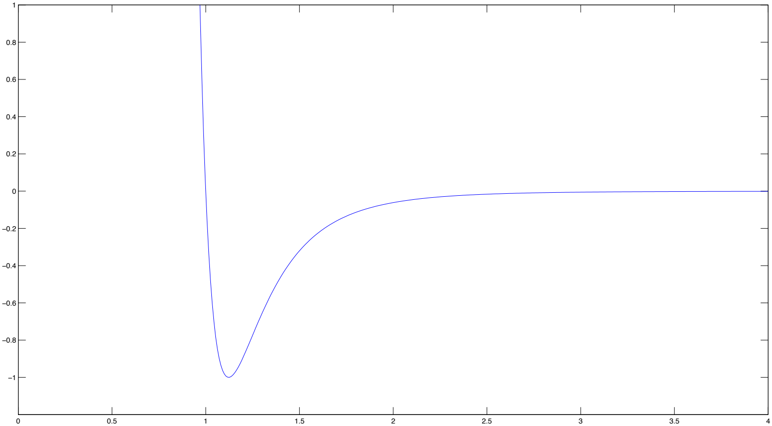
$$V(r) = 4\varepsilon\left[(\sigma/r)^{12} - (\sigma/r)^{6}\right]$$

modeling the potential energy of 2 atoms at distance r (noble gases)

Molecular configuration

Lennard-Jones potential:

$$V(r) = 4\varepsilon\left[(\sigma/r)^{12} - (\sigma/r)^{6}\right]$$



$\varepsilon = \sigma = 1$

Molecular configuration

Molecule with n atoms: $x_i \in \mathbb{R}^3$ spacial position of atom i

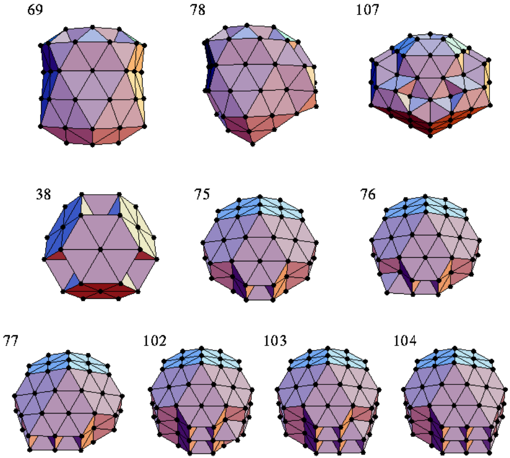
$$E[x_1, .., x_n] = \sum_{i<j} 4\varepsilon \left[(\sigma / r_{ij})^{12} - (\sigma / r_{ij})^6 \right]$$

where $r_{ij} = ||x_i - x_j||_2$ is the distance between atoms i and j

Molecular configuration with minimal potential energy

$$\min\{ \ E[x_1, .., x_n] \ : \ x_1, ..., x_n \in \mathbb{R}^3 \ \}$$

Some configurations



http://www-wales.ch.cam.ac.uk/CCD.html
http://physchem.ox.ac.uk/~doye/jon/structures/LJ.html