Numerical Methods and Optimization: introduction

Mauro Passacantando

Department of Computer Science, University of Pisa
mauro.passacantando@unipi.it

Numerical Methods and Optimization
Master in Computer Science – University of Pisa

Course Organization

- 1 course (12 cfu/ects)
- 1 program
- 1 exam
- 2 lecturers

Federico Poloni (Numerical methods)

Dipartimento di Informatica
Largo B. Pontecorvo 3, Pisa
Building C, 2nd floor, room 343
Tel. 050 2213143, e-mail: federico.poloni@unipi.it

Office hours: Still to decide

Mauro Passacantando (Optimization)

Dipartimento di Informatica
Largo B. Pontecorvo 3, Pisa
Building C, 2nd floor, room 323
Tel. 050 2212756, e-mail: mauro.passacantando@unipi.it

Office hours: Wednesday 9:00-11:00

Information

Course Schedule

- Tuesday 9:00-11:00 room L1
- Tuesday 16:00-18:00 room H-lab
- Wednesday 11:00-13:00 room C (except September 21: room C1)
- Thursday 11:00-13:00 room C

Web page

https://elearning.di.unipi.it/course/view.php?id=77

Exam

Written and oral exam

Bibliography

- S. Boyd, L. Vandenberghe, Convex optimization, Cambridge University Press, 2004 (available on `http://web.stanford.edu/~boyd/cvxbook/`)
- M.S. Bazaraa, H.D. Sherali, C.M. Shetty, Nonlinear programming: theory and algorithms, Wiley & Sons, 2006 (Chapters 1-6, 8-9)
- J. Nocedal, S. Wright, Numerical Optimization, Springer Series in Operations Research and Financial Engineering, 2006 (Chapters 1-3, 5, 12, 16, 17)
- A.R. Conn, K. Scheinberg, L.N. Vicente, Introduction to Derivative-Free Optimization, SIAM series on Optimization, 2009 (Chapters 1, 7)

- J. Demmel, Applied Numerical Linear Algebra, SIAM press, 1996
- L. N. Trefethen, D. Bau, Numerical Linear Algebra, SIAM press, 1997

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 linear least-squares problem
- Iterative methods for computing eigenvalues
- Constrained optimization and systems of equations
- Lagrangian duality
- Numerical methods for constrained optimization
- The fast Fourier transform
- Applications: regression, parameter estimation, approximation and data fitting, support vector machines, image and signal reconstruction
- Software tools for numerical and optimization problems (Matlab).

Support Vector Machines for Data Classification

Given training data in different classes (labels known)

Predict test data (labels unknown)

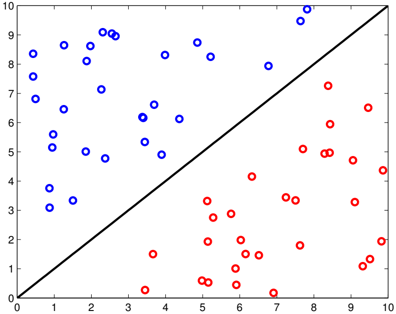
Examples:

- handwritten digits recognition
- spam filtering
- credit card fraud detection
- marketing
- medical diagnosis
- image processing
- ...

Support Vector Machines for Data Classification

Given two sets $A, B \subset \mathbb{R}^n$ (training set).
Assume that A and B are linearly separable, i.e., there is an hyperplane $H = \{x \in \mathbb{R}^n : \; w^\mathsf{T} x + b = 0\}$ such that

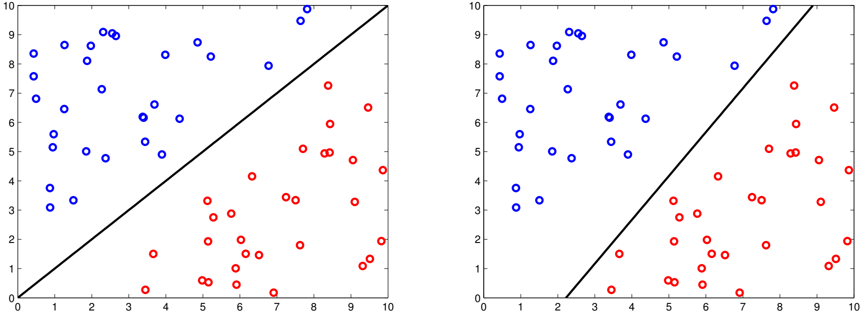
$$\begin{array}{ll} w^\mathsf{T} x^i + b > 0 & \forall\, x^i \in A \\ w^\mathsf{T} x^j + b < 0 & \forall\, x^j \in B \end{array}$$



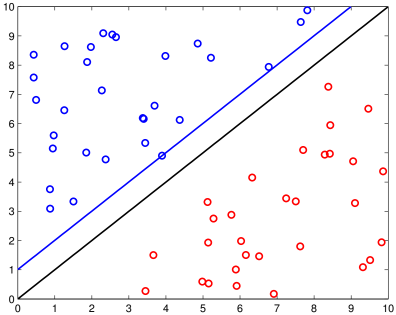
Test data x:
decision function $f(x) = \text{sign}(w^\mathsf{T} x + b)$

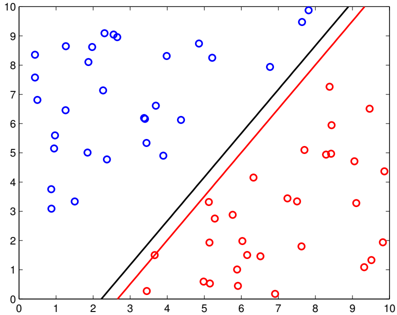
Support Vector Machines for Data Classification

Many possible choices of w and b. Which hyperplane do we choose?



Support Vector Machines for Data Classification





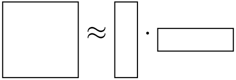
We look for the separating hyperplane with the maximum margin of separation. This problem can be modeled as

$$\begin{cases} \min \ \|w\|^2 \\ w^{\mathsf{T}} x^i + b \geq 1 \qquad \forall \ x^i \in A \\ w^{\mathsf{T}} x^j + b \leq -1 \qquad \forall \ x^j \in B \end{cases}$$

Low-rank approximation with the Singular Value Decomposition

Problem

Given a matrix $M \in \mathbb{R}^{n \times m}$, approximate it with the product of a 'tall thin' $A \in \mathbb{R}^{n \times k}$ and a 'fat large' $B \in \mathbb{R}^{k \times m}$ ($k \ll n, m$):



$$\min_{A,B} \|M - AB\|$$

Why?

- Reduce data dimension
- Reveal features

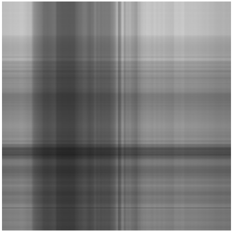
Applications:

- Text and data mining
- Neural networks
- Web search
- Graph / community analysis...

An example

An example easy to visualize: image compression by approximation.
Black/white image: matrix with color intensities $\in [0,1]$.

Original (512×512)



$k = 1$

$k = 10$

$k = 25$

$k = 50$

$k = 100$

Eigenvalue problems

Problem

Given a matrix $M \in \mathbb{R}^{n\times m}$, approximate it with the product of a 'tall thin' $A \in \mathbb{R}^{n\times k}$ and a 'fat large' $B \in \mathbb{R}^{k\times m}$ ($k \ll n$): $\min_{A,B} \|M - AB\|$.

Optimal A, B can be obtained from eigenvectors of $M^T M$ and MM^T.

- How can we solve this problem?
- What if M is very large and sparse?
- What if M has some special structure?
- Can we avoid forming $M^T M$ and MM^T?
- Is the solution numerically stable, or should we expect large errors?