# Applicazioni della Ricerca Operativa al Data Mining e all’Information Retrieval

*Carla Michini, Università di Roma “Sapienza”*

Fondazione Ugo Bordoni

# Sommario

- Introduzione alla Ricerca Operativa

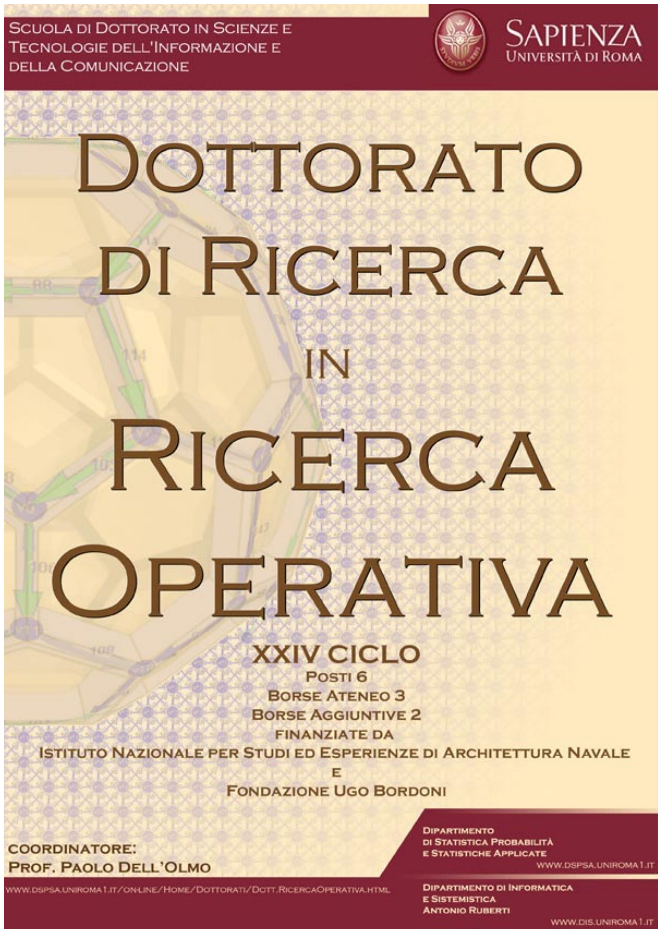
**TEORIA:**

- Cenni di Programmazione Lineare Intera
- Set Covering

**APPLICAZIONI:**

- Classificatori binari
- Previsione degli ascolti degli utenti radiotelevisivi
- Kernel semantico basato su Reticoli Concettuali

# Ricerca Operativa?

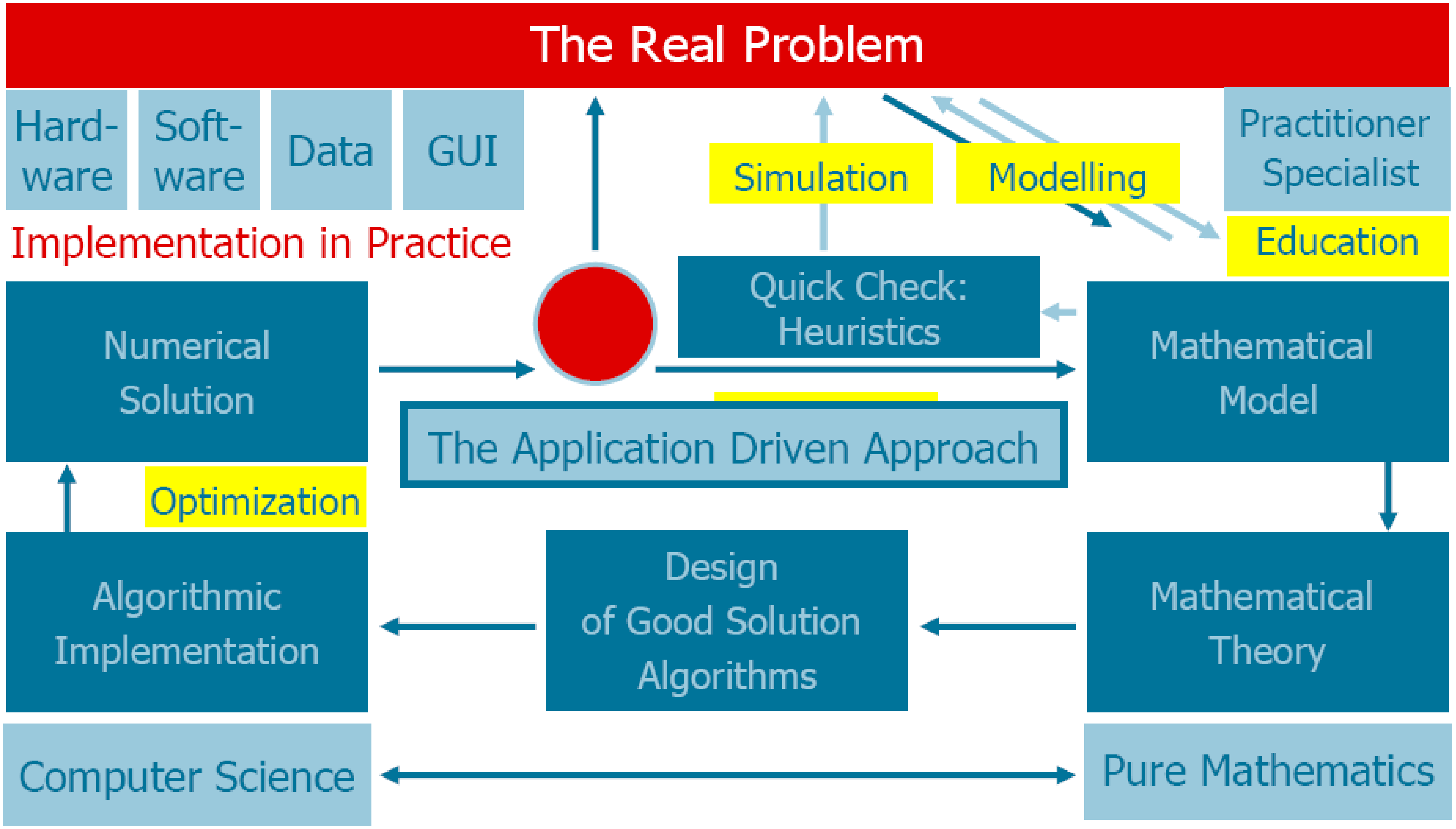



Applicazione di metodi scientifici a problemi decisionali che si presentano in strutture organizzate complesse

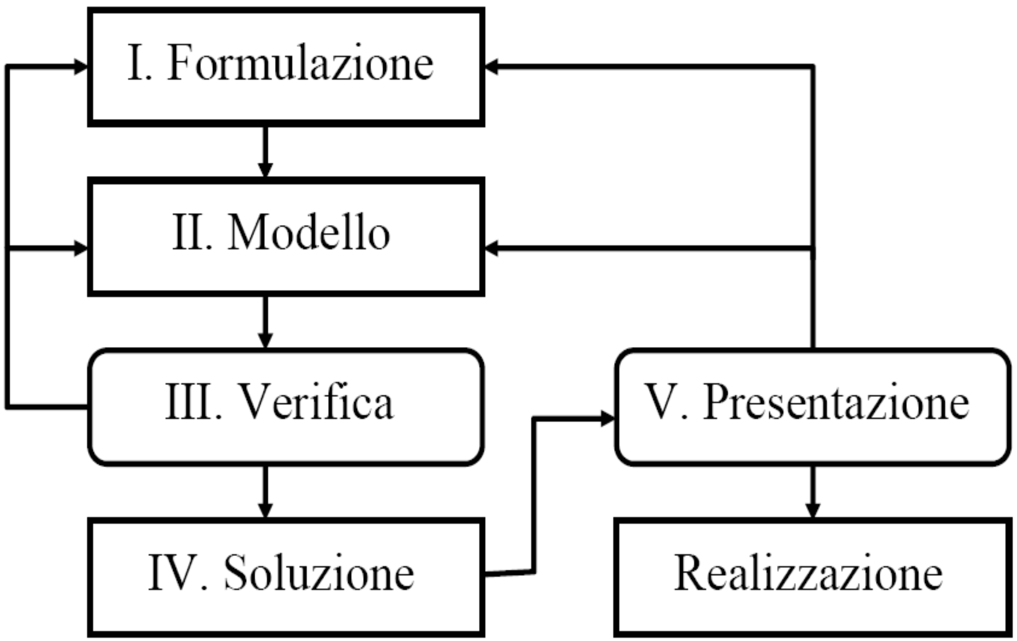
Matematica → Ricerca Operativa ← Informatica

Ricerca Operativa
(Operations Research)

# Matematica... all'opera!



# Metodologia della Ricerca Operativa




# Tipici Problemi di ottimizzazione

$$\begin{aligned} &\min \ f(x) \\ &g_i(x) = 0, \ i = 1,\dots,k \\ &h_j(x) \le 0, \ j = 1,\dots,m \\ &x \in \mathbf{R}^n \end{aligned}$$

General
Nonlinear
Program
NLP

$$\begin{aligned} &\min \ c^T x \\ &Ax = a \\ &Bx \le b \\ &x \ge 0 \\ &x \in \mathbf{R}^n \end{aligned}$$

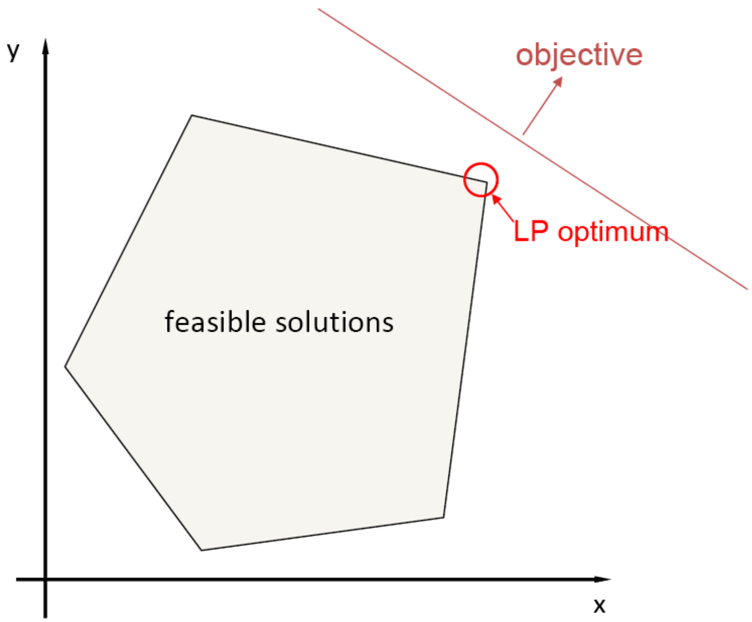
Linear
Program
LP

$$\begin{aligned} &\min \ c^T x \\ &Ax = a \\ &Bx \le b \\ &x \ge 0 \\ &x \in \mathbf{Z}^n \quad [x \in \{0,1\}^n] \end{aligned}$$

(linear)
Integer
Program
IP, MIP

Program = optimization problem

# Linear Programming

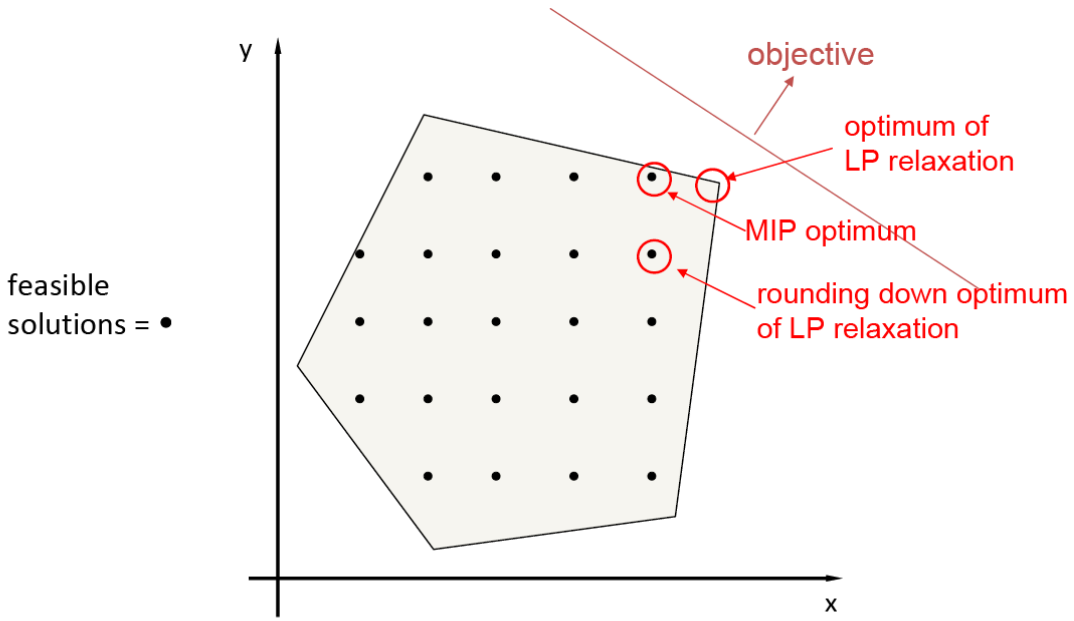



Linear Program

$$
\boxed{\begin{array}{l} \min \ c^T x \\ Ax = a \\ Bx \le b \\ x \ge 0 \\ x \in \mathbf{R}^n \end{array}}
$$

Linear Program LP

# Programmazione Lineare Intera

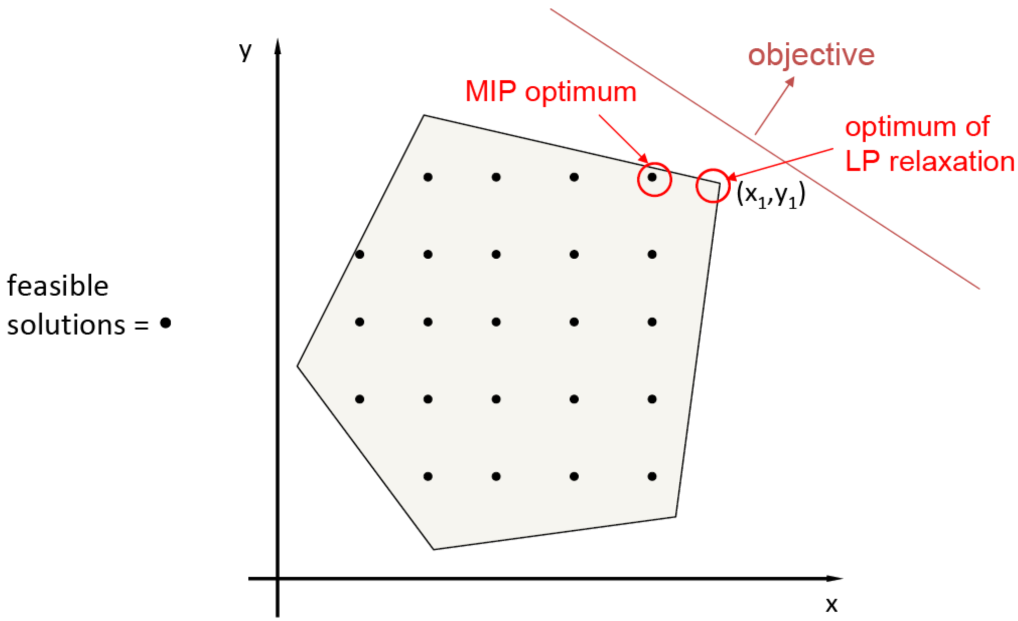



Integer Program

$$
\begin{aligned}
&\min \ c^T x \\
&Ax = a \\
&Bx \leq b \\
&x \geq 0 \\
&x \in \mathbf{Z}^n \qquad [x \in \{0,1\}^n]
\end{aligned}
$$

(linear)
Integer
Program
IP, MIP

# Programmazione Lineare Intera: Rilassamento Lineare

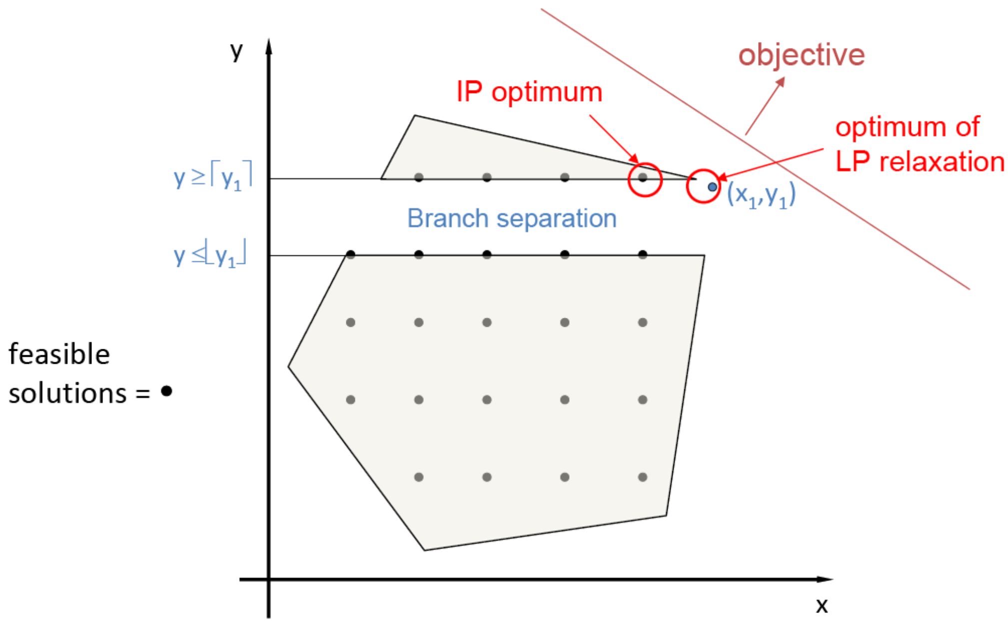



Integer Program

$$\min \; c^T x$$
$$Ax = a$$
$$Bx \leq b$$
$$x \geq 0$$
$$x \in \mathbf{Z}^n \qquad [x \in \{0,1\}^n]$$

(linear)
Integer
Program
IP, MIP

# Programmazione Lineare Intera: Branch and Bound



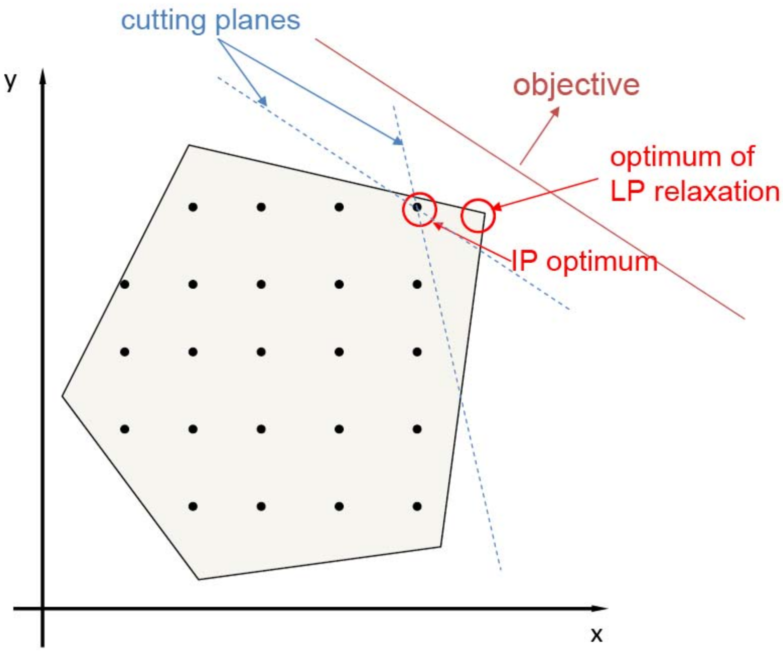

Integer Program: Branch and Bound

$$\min \; c^T x$$
$$Ax = a$$
$$Bx \leq b$$
$$x \geq 0$$
$$x \in \mathbf{Z}^n \qquad [x \in \{0,1\}^n]$$

(linear)
Integer
Program
IP, MIP

# Programmazione Lineare Intera: Piani di Taglio

feasible solutions = •




Integer Program: Cutting Planes

$$
\begin{aligned}
&\min\ c^T x \\
&Ax = a \\
&Bx \le b \\
&x \ge 0 \\
&x \in \mathbf{Z}^n \qquad [x \in \{0,1\}^n]
\end{aligned}
$$

(linear)
Integer
Program
IP, MIP

# Set Covering ed Euristiche

$$\min \sum_{j \in N} c_j x_j$$

$A = (a_{ij})$ $\quad$ $a_{ij} = 1$ se l'elemento la colonna j copre la riga i

$$\sum_{j \in N} a_{ij} x_j \geq 1 \quad \forall i \in M$$

$$x_j \in \{0,1\} \quad \forall j \in N$$

**Obiettivo: determinare un insieme di colonne di costo minimo, tale che ogni riga sia coperta da almeno una colonna**

**NP-HARD**

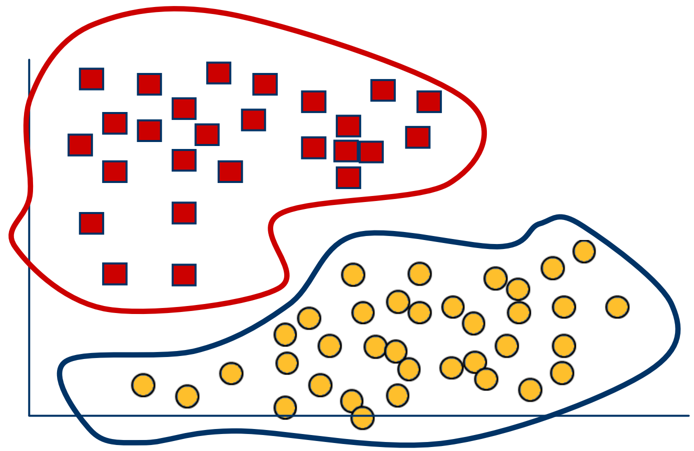
EURISTICHE = "buone" soluzioni ammissibili, che possono essere determinate attraverso metodi efficienti

- Polyhedral Theory
- Concave Programming
- Matroids and Independent Systems
- …

# Classificatore binario

$x \in \Re^n$ $\longrightarrow$ **LEARNING MACHINE** $f : \Re^n \to \{-1,1\}$ $\longrightarrow$ $y = f(x)$

- Apprendimento Automatico Supervisionato
- Classificazione binaria



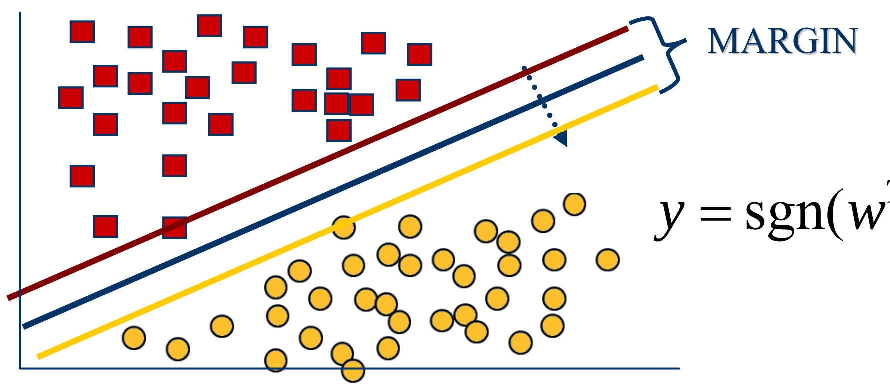
$$TS = \{(x^i, y_i) \in \Re^n \times \{-1,1\}, \quad i = 1, \ldots, m\}$$

# Support Vector Machine

$x \in \Re^n$ $\longrightarrow$ **SVM** $f(x) = sign(w^T x + b)$, $w \in \Re^n$ $b \in \Re$ $\longrightarrow$ $y = f(x)$

- La superficie di separazione è un iperpiano
- L'iperpiano di separazione migliore è quello che massimizza il margine

$$y = \mathrm{sgn}(w^T x + b) = -1$$




$$y = \mathrm{sgn}(w^T x + b) = 1$$

$$TS = \{(x^i, y_i) \in \Re^n \times \{-1,1\}, \quad i = 1, \ldots, m\}$$

# Addestramento e test della SVM

- L'iperpiano di separazione ottimo viene determinato risolvendo un problema di ottimizzazione

**Addestramento**

$$\min_{w \in \Re^n, b \in \Re} \tfrac{1}{2} \|w\|_p^{\,p}$$

$$y_i \left[ w^T x^i + b \right] \geq 1 \qquad i = 1, \ldots, m$$

- Classificatore:

**Test**

$$f(d) = sign\left( \sum_{i=1}^{m} \alpha_i^{\,*} y_i k(d, d^i) + b^* \right)$$

KERNEL FUNCTION

# Il marketing radiotelevisivo


**Dati Auditel & Dati programmi:**

- Campione: 5163 famiglie, 14000 individui
- 24 ore su 24, minuto per minuto
- 30/03/08-5/04/08
- RAI 1, RAI 2, RAI 3, Canale 5, Italia 1, Rete 4, La 7

- pianificare gli investimenti delle aziende utenti di pubblicità
- valutare le "performance" dei programmi
- analizzare i comportamenti del pubblico
- fornire elementi per migliorare l'offerta televisiva

**CONTESTO ATTUALE**

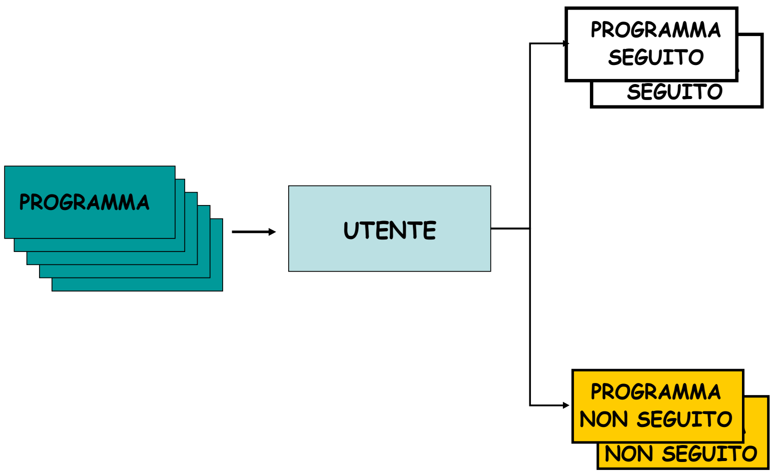
**PREVISIONE MANUALE**

**CONTESTO FUTURO**

**DATA MINING**

# Previsione degli ascolti

## GENERALIZZAZIONE PER SINGOLO UTENTE




Fissato utente *u*

Componenti istanze:

- Caqratteristiche programma *x*
- Etichetta (visto/non visto)

Addestramento SVM

**Output:** Classificatore per prevedere la visione programma *x* da parte dell'utente *u*

# Classificazione di documenti di testo

- **Documenti** come vettori di n componenti

$$documents = \{d_1,\dots,d_m\}$$
$$dictionary = \{t_1,\dots,t_n\}$$

$$d_j^{\,T} = [\,f_{1j}, f_{2j},\dots, f_{nj}\,]$$

- **Bag-of-words kernel:** $k^{BOW}(d_1,d_2) = d_1^{\,T} d_2 = \sum_{i=1}^{n} f_{i1} f_{i2}$

IGNORA LA SEMANTICA

- **Mapping** dei documenti: $\phi(d_j) = P d_j \quad j = 1,\dots,m$

- Kernel **semantico** $k^{SEM}(d_1,d_2) = \phi(d_1)^T \phi(d_2) = d_1^{\,T} P^T P d_2$

*P deve contenere informazioni SEMANTICHE*

IDEA: Definire *P* tale che $[P]_{ij}$ rappresenti la prossimità tra i termini $t_i$ e $t_j$

# Costruzione del kernel semantico

- Rappresentazione dei termini su un reticolo concettuale
- La distanza tra due termini corrisponde alla lunghezza del cammino minimo che li congiunge sul reticolo concettuale.
- La similarità tra ogni coppia di termini è calcolata come funzione inversa della loro distanza

$$(D, T, TD)$$

context

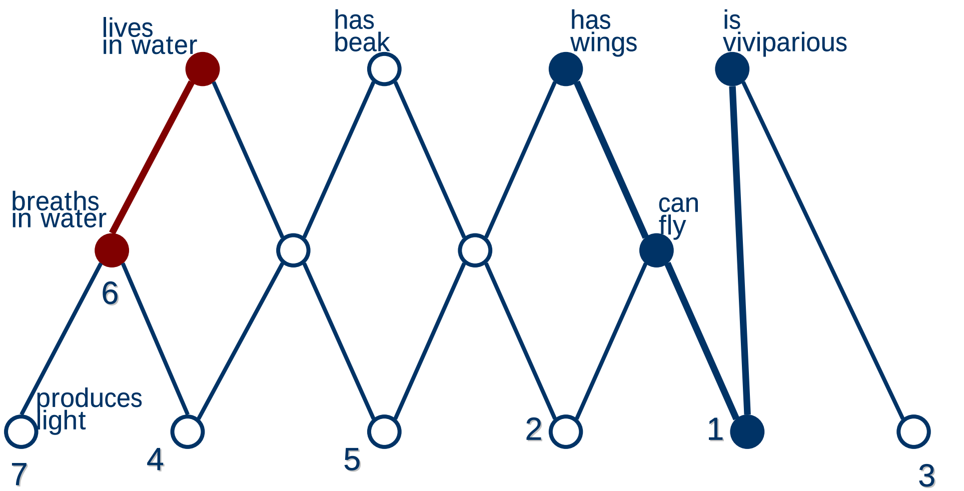
$$\mathcal{C}(D, T, TD)$$

set of *concepts*

Es.

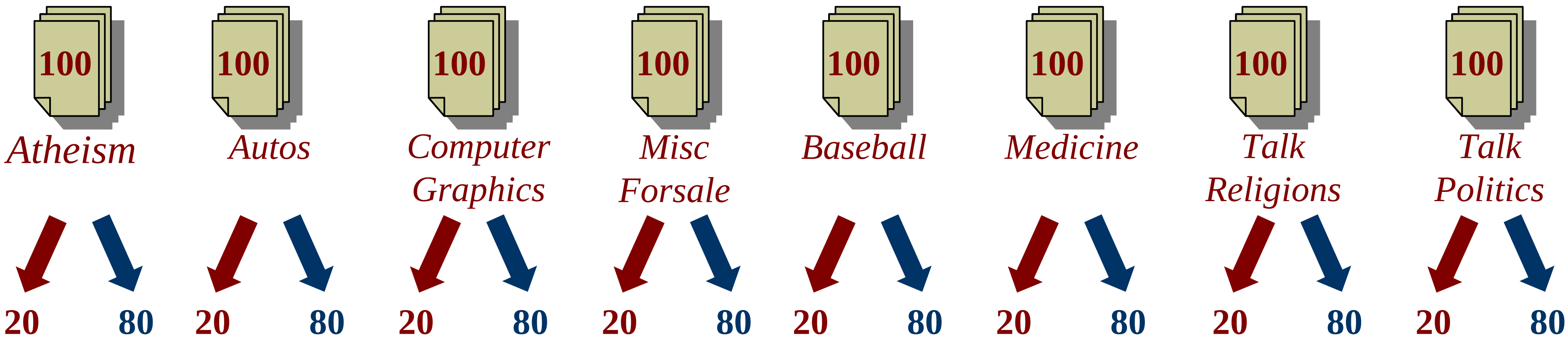
$$d(g,e) = 3$$
$$(13, g) \succ\!\prec (1, beg) \succ\!\prec (12, be) \succ\!\prec (125, e)$$

$$d(a, f) = 1$$
$$(467, af) \succ\!\prec (4567, f)$$



# Esperimenti (1)

Mini-20NG collection




**TRAINING** 160 $\supset$ 80 $\supset$ 40

**TEST** 640

# Esperimenti (2)

**SIZE OF DATASET**

**KERNEL FUNCTION**

| | 160 training docs | 80 training docs | 40 training docs |
|---|---|---|---|
| linear BOW-SVM | 58,516 | 59,8651 | 58,516 |
| linear CL-SVM | 68,1282 | 65,43 | 64,5868 |
| gaussian BOW-SVM | 61,8887 | 59,1906 | 58,516 |
| gaussian CL-SVM | 62,3946 | 62,3946 | 60,7083 |

- **CL-SVM outperformed BOW-SVM**
- As the size of the training set decreased, the performance of each method decreased very gently, while the gain in performance due to the use of the conceptual method did not grow
- gaussian kernel together with the conceptual document similarity metric seemed to hurt performance

# Thank you!