

Facoltà di Scienze MM. FF. NN.

Università di Verona

A.A. 2014-15

Teorie e Tecniche del Riconoscimento

Introduzione alla
Pattern Recognition

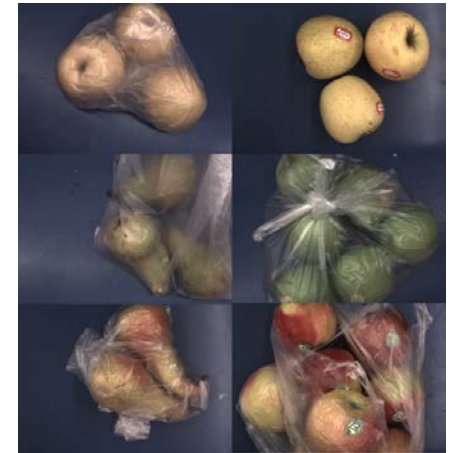
Inquadramento

- Sistemi di Pattern Recognition nell'uomo:
 - riconoscere la faccia di una persona conosciuta, anche se questo cambia pettinatura, ha gli occhiali da sole, ...
 - capire quello che una persona sta dicendo, anche se varia il tono della voce;
 - leggere una lettera scritta a mano;
 - ...
- Attività che l'uomo risolve in modo molto naturale, mentre per un calcolatore sono compiti complicati e complessi

Alcune possibili definizioni

- *Pattern recognition*
 - studio delle problematiche connesse all'utilizzo dei calcolatori per il riconoscimento automatico di dati, altrimenti detti *pattern*.
- Studio di come le macchine possono osservare l'ambiente, imparare a distinguere i pattern di interesse dall'informazione di sfondo e prendere decisioni relative alla categoria dei pattern, a partire da quella più fondamentale, ossia *classificare*
- *Sistema di Pattern Recognition*: il processo che prende in input dati grezzi (*raw*) ed effettua un'azione sulla base della “categoria” dei dati.

Esempio, bilancia multimediale



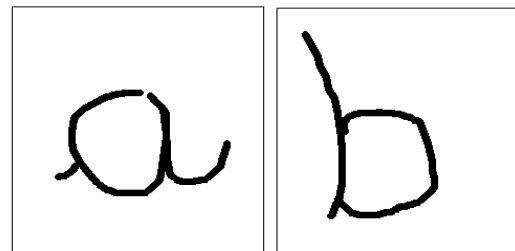
 **Fraunhofer**
IOSB

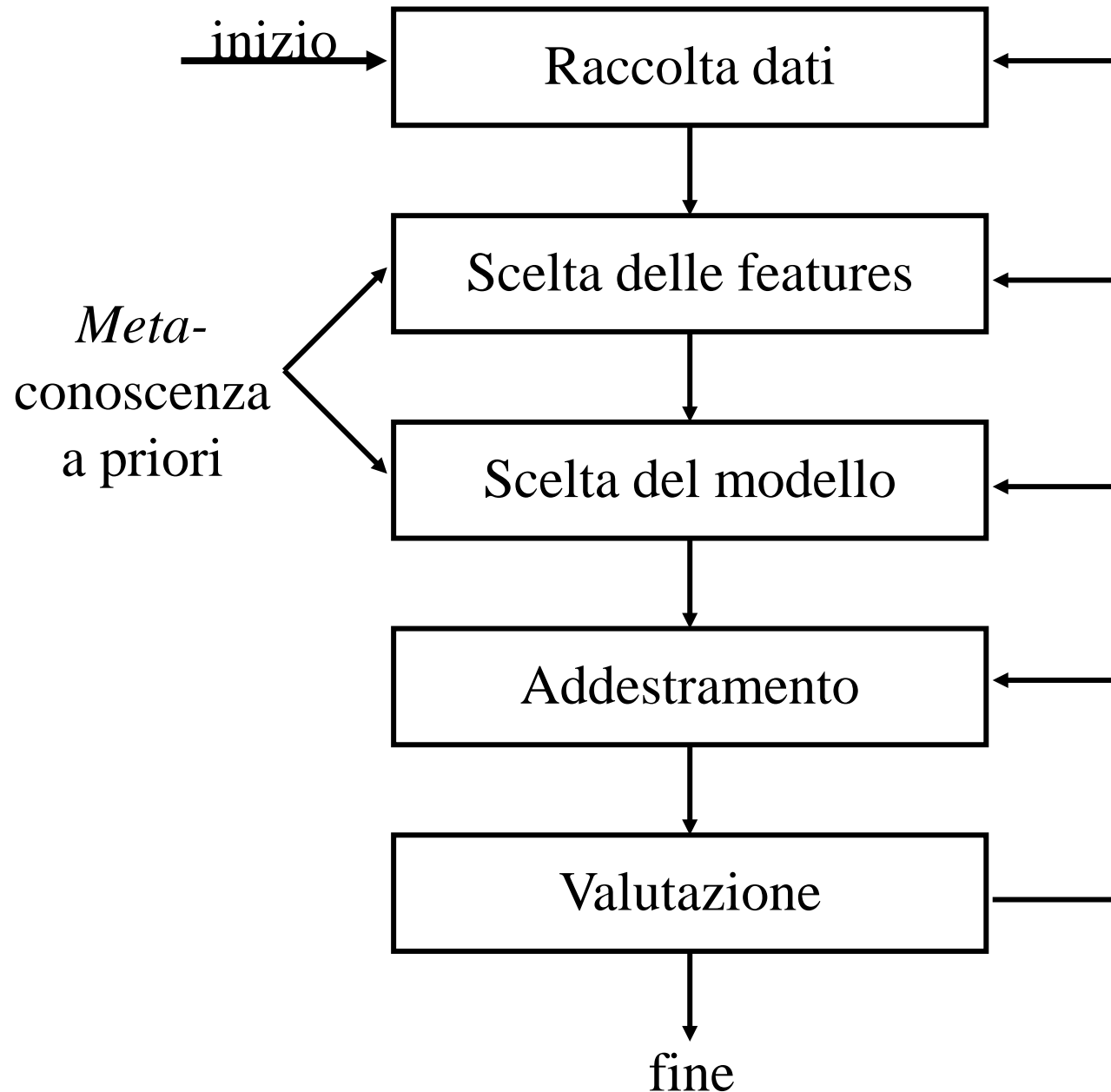
<http://www.iosb.fraunhofer.de/servlet/is/3332>
8/

Sistema di Pattern Recognition

- Raccolta dati
- Scelta delle *feature*
- Scelta del modello
- Addestramento del modello
- Valutazione

Esempio guida: sistema che distingue tra le lettere scritte a mano “a” e “b”.

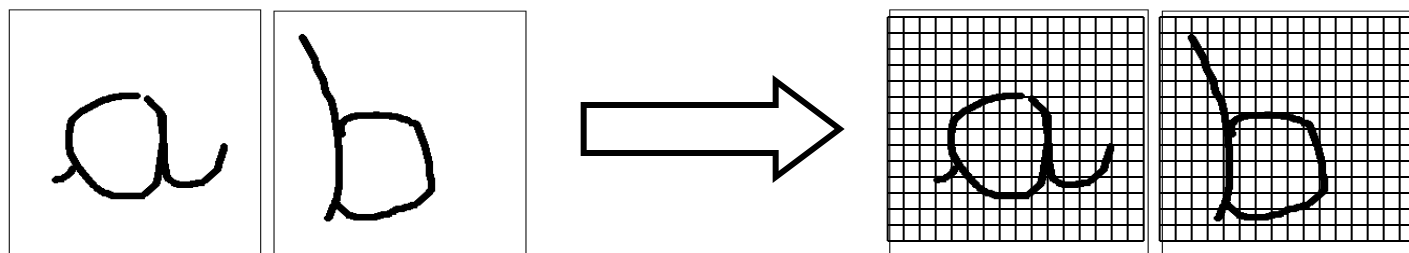




Raccolta Dati

- Collezione di un insieme “sufficiente” e “rappresentativo” di esempi dal problema in esame.
- “sufficiente”?
 - riesce a distinguere un pattern di test e metterlo in una delle classi a disposizione
- “rappresentativo”?
 - copre lo spazio di tutti i possibili esempi
- Problemi di sensoristica (risoluzione, banda, ...)

- *Esempio:* un insieme di immagini contenenti le lettere “a” e “b” viene acquisito tramite una telecamera, e memorizzato nel computer



L'immagine viene rappresentata da un array di pixel, ogni pixel assume valore compreso tra 0 (completamente bianco) e 1 (completamente nero)

Scelta delle feature

- Non si possono utilizzare i dati così come sono (immagine 256x256 sono 65536 pixels)
- *Feature*: caratteristiche misurabili del fenomeno in esame (*pattern* = vettore di features):
 - semplici da calcolare;
 - invarianti a trasformazioni irrilevanti;
 - affidabili;
 - indipendenti;
 - discriminanti;
 - poche (problema della *curse of dimensionality*);
- In questa fase è molto utile l'utilizzo della conoscenza a priori sul problema

- *Esempio:*
- ad esempio una feature potrebbe essere il numero totale di pixel neri:
 - invariante alla rotazione e traslazione dell'oggetto
 - poco discriminante: non tiene conto della forma
- uso di meta-conoscenza a priori: devo distinguere tra “a” e “b”, e so che la lettera “b” è tipicamente più alta e allungata della “a”.
 - uso come feature il rapporto altezza/larghezza

Scelta delle feature: estrazione e selezione

- Il numero di feature deve essere piccolo per limitare il costo della misura e non influire sull'accuratezza del classificatore
- Estrazione di feature:
 - misura sui dati
 - creazione di nuove feature da combinazioni di feature misurate (esempio precedente delle a e b)
- Selezione di feature:
 - migliore sottoinsieme delle feature estratte
- Tali feature possono aver una miglior capacità discriminativa, ma si perde il significato fisico di queste.
- Uso di una funzione criterio per la riduzione: tipicamente l'errore di classificazione di un sottoinsieme di feature.
- Inoltre, è importante determinare la dimensione dello spazio ridotto.

Scelta del modello

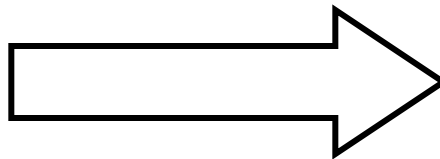
- Scelta della struttura logica e la base matematica delle regole di classificazione.
- Tipicamente, il classificatore stima, per ogni oggetto, un valore che indica il grado di appartenenza ad una o più classi sulla base del vettore di feature che lo caratterizza.
- Problemi:
 - come decidere il modello
 - come decidere la **dimensione del modello** (ossia *il numero dei suoi parametri*)
 - come capire se il modello ottenuto rappresenta effettivamente il fenomeno in esame

- Non esiste un classificatore che vada bene per tutte le applicazioni
- *Esempio:* uso di un classificatore a soglia:
 - data un'immagine I
 - calcolo il rapporto altezza/larghezza $R(I)$;
 - se $R(I)$ è maggiore di una certa soglia θ , allora l'immagine è una “b”, altrimenti è una “a”.

Addestramento del modello

- Sinonimi:
 - *training* del classificatore
 - *learning* del classificatore
- Processo (spesso iterativo, basato su *cicli*, *epoche*) con il quale si utilizzano i dati a disposizione (*training set*) per la costruzione del modello

Esempi tratti dal
problema
(*Training Set*)
Conoscenza a
priori



Regole che
governano il
fenomeno

- *Esempio:*

addestramento del modello = determinazione della soglia θ

- Si ha a disposizione una serie di immagini di esempio per le lettere “a” e per le lettere “b” (*training set*)
- calcolo $R(I)$ per tutte le immagini del training Set
- determino una soglia θ “adatta” a separare i valori $R(I)$ calcolati

Tipologie di addestramento: supervisionato

- Sinonimi: *supervised learning*, *classificazione*
- Idea e scopo:
 - di ogni elemento del training set si conosce l'esatta categoria.
 - L'obiettivo è quello di creare uno strumento in grado di classificare nuovi oggetti.
- Problemi:
 - capire se un algoritmo di training è capace di trovare la soluzione ottimale;
 - capire se converge, e se è sufficientemente scalabile;
 - capire se riesce a prediligere soluzioni semplici.

- *Esempio:*
 - il training set è costituito da un insieme di immagini “a” e “b”.
 - di ogni immagine conosciamo l’esatta classificazione (cioè se è “a” oppure “b”)
 - queste informazioni sono utilizzate per determinare la soglia del classificatore.

Tipologie di addestramento: semi supervisionato

- Sinonimi: semi-*supervised learning*,
- Idea e scopo:
 - di qualche elemento del training set si conosce l'esatta categoria, della maggior parte no
 - L'obiettivo è quello di creare uno strumento in grado di classificare nuovi oggetti.
- Problemi:
 - Come nel caso precedente

Tipologie di addestramento: non supervisionato

- Sinonimi: *unsupervised learning, clustering*
- Idea e scopo:
 - nessuna informazione sulla categorizzazione degli elementi del training set.
 - Il sistema deve trovare i clusters (gruppi) “naturali” all’interno del training set, sulla base della “similarità” tra patterns
- Problemi:
 - intrinsecamente più difficile della classificazione
 - “naturali”?
 - “similarità”?

- *Esempio:*
 - il training set è costituito da un insieme di immagini “a” e “b”.
 - nessuna informazione sulla categorizzazione delle immagini.
 - si cercano di creare due gruppi, mettendo assieme quelle immagini che hanno valore simile di $R(I)$ (la *feature*)

Tipologie di addestramento: Rinforzato

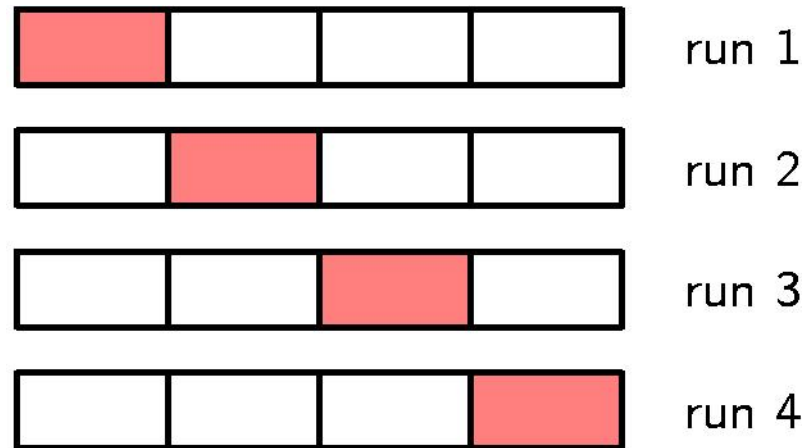
- Sinonimi: *reinforcement learning*, *learning with a critic*
- Idea: a metà strada tra le due: non viene fornita alcuna informazione sulla categoria esatta, viene dato un giudizio sulla correttezza della classificazione
- La strategia di addestramento viene modificata:
 - si presenta un pattern al classificatore
 - il classificatore fa un tentativo di classificazione
 - viene detto *esclusivamente* se il tentativo è corretto o meno
 - sulla base del giudizio si modifica il classificatore

Valutazione e model selection

- Misura delle prestazioni del classificatore
- Prestazioni di *generalizzazione*: capacità del classificatore di classificare correttamente anche esempi non presenti nel data set
- Nessun errore sul training set non implica necessariamente aver ottenuto il classificatore ottimale (pb di *overfitting*, *overtraining*)
- Per evitare situazioni di *overfitting* è sempre meglio utilizzare due insiemi disgiunti in fase di learning, uno per il training e uno per il testing.

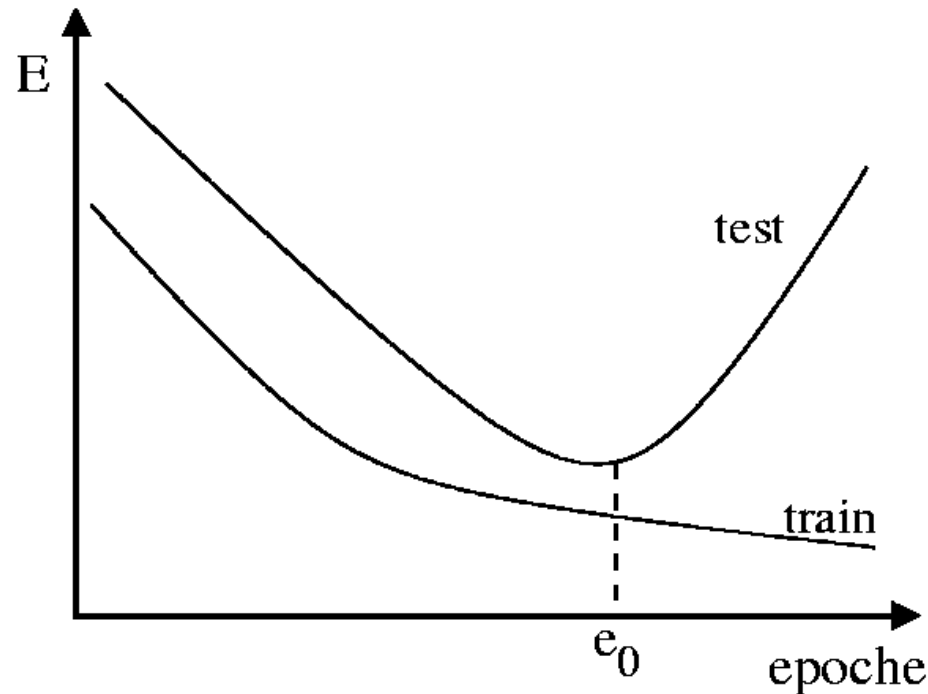
- Tecniche per la scelta del training set e del testing set:
 - *Holdout*: si suddivide casualmente il training set in due parti uguali: una per il training una per il testing
 - *Averaged Holdout*: si effettuano più partizioni holdout, e si media il risultato ottenuto. In questo modo si ha indipendenza dalla particolare partizione scelta
 - *Leave One Out*: per il training vengono utilizzati tutti i patterns tranne uno, utilizzato per il testing. Si ripete per tutte le possibili combinazioni.
 - *Leave P-Out*: come il precedente, utilizza P elementi per il testing, invece che uno.
 - *K-folding*: vedasi slide successiva

- **Cross-Validation: 4-fold**



Problemi training/valutazione: overtraining

- Problema che nasce quando il classificatore impara solo i dati di training e non generalizza più



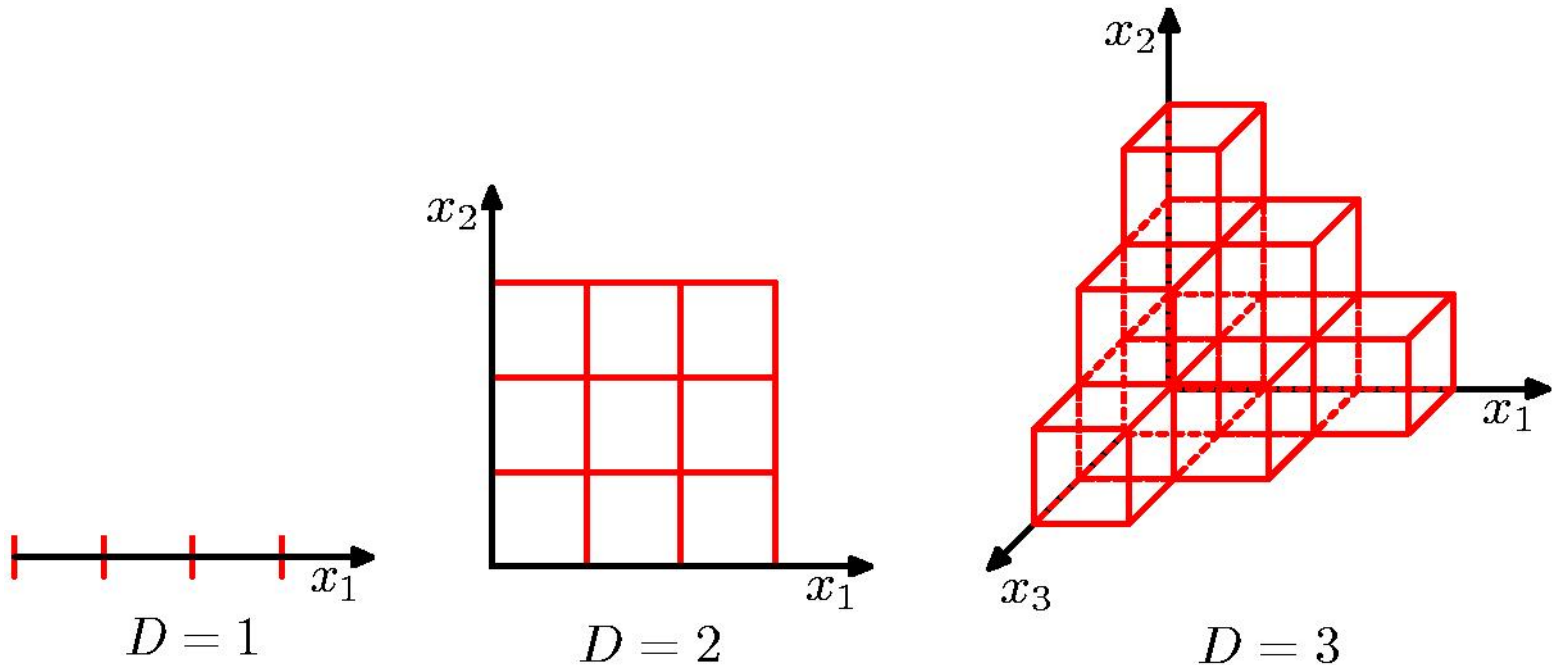
Si ferma l'addestramento prima del verificarsi del fenomeno dell'overtraining (o overfitting) (e_0)

Problemi training/valutazione: Curse of dimensionality

- Le prestazioni dipendono dalle relazioni tra il numero di campioni, numero di feature e dalla complessità del classificatore.
- In teoria, la probabilità di errore non aumenta se si aggiungono feature
- In pratica si riscontrano dei problemi dovuti al fatto che le ipotesi sono solo approssimazioni nei casi reali
- Inoltre, il numero di campioni deve essere in relazione esponenziale rispetto al numero di feature
- Tutti i comuni classificatori soffrono di questo problema ed esistono regole guida

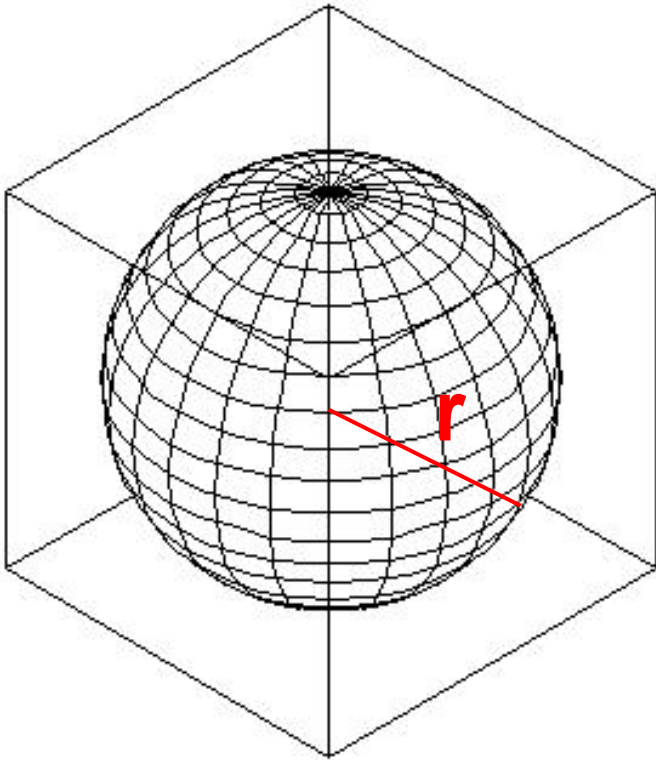
Problemi training/valutazione: Curse of dimensionality (2)

- Vorrei tessellare lo spazio di esistenza dei dati in modo uniforme. Di quanti dati ho bisogno, all'aumentare delle dimensioni?



Problemi training/valutazione: Curse of dimensionality (3)

Mostro che avere i dati nello spazio sferico nel cubo che lo racchiude è non informativo ad alte dimensionalità

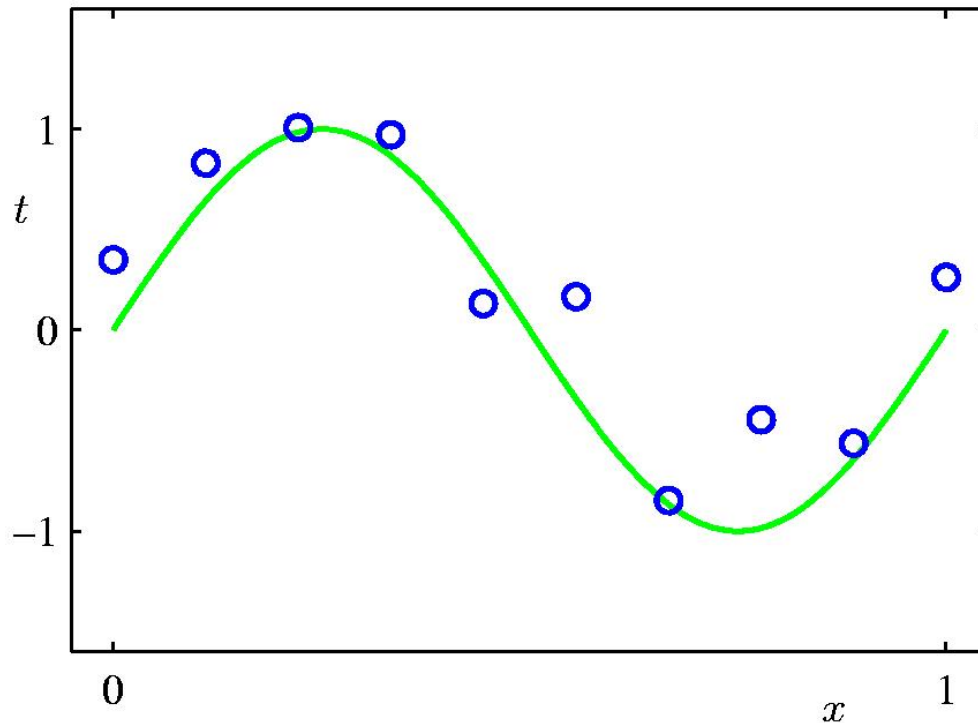


$$\text{Vol. cubo} = (2r)^d$$

$$\text{Vol. sfera} = \frac{2r^d \pi^{d/2}}{d\Gamma(d/2)}$$

Metto a rapporto e al limite ottengo $\lim_{d \rightarrow \infty} \frac{\pi^{d/2}}{d 2^{d-1} \Gamma(d/2)} = 0$

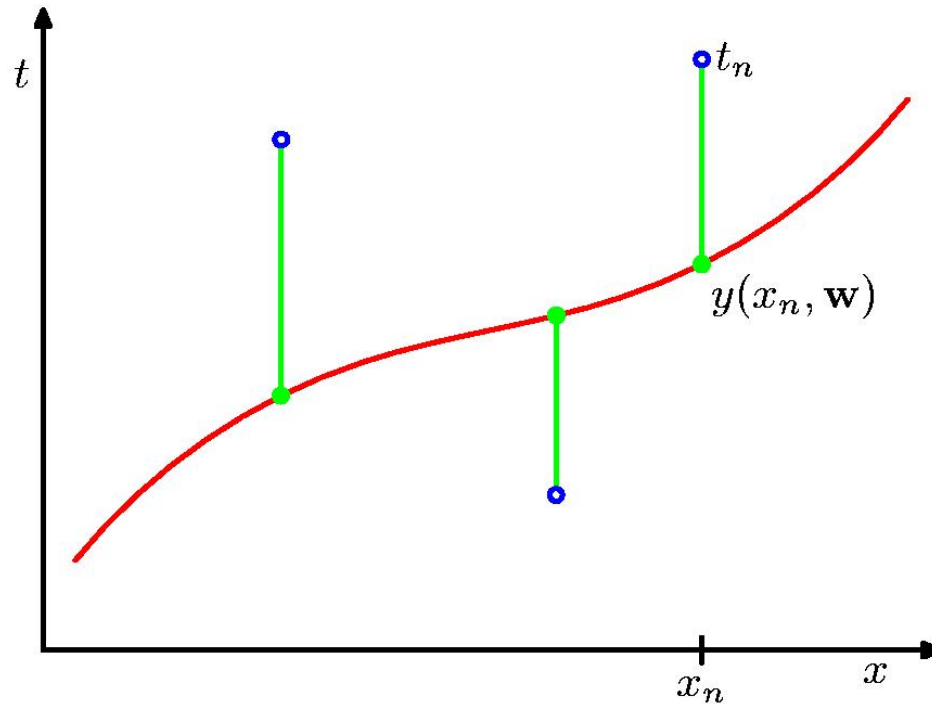
Curse of dimensionality: Fitting di una curva polinomiale



$$y(x, \mathbf{w}) = w_0 + w_1x + w_2x^2 + \dots + w_Mx^M = \sum_{j=0}^M w_jx^j$$

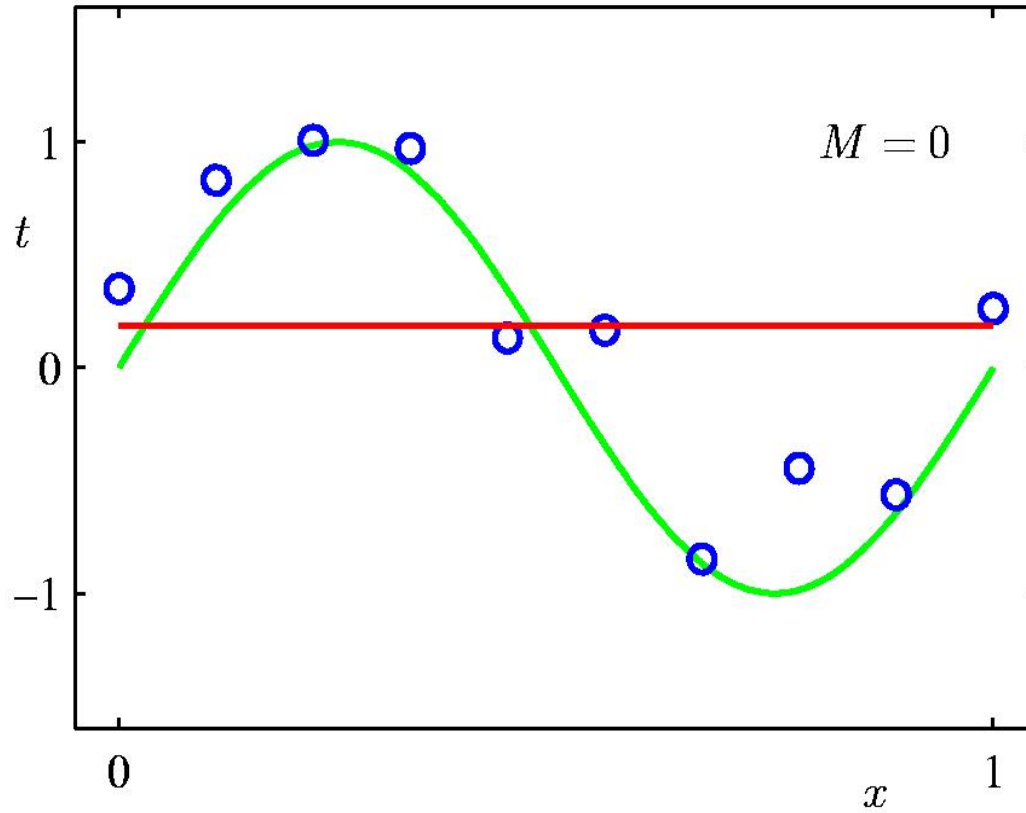
Curse of dimensionality:

Fitting di una curva polinomiale (2)

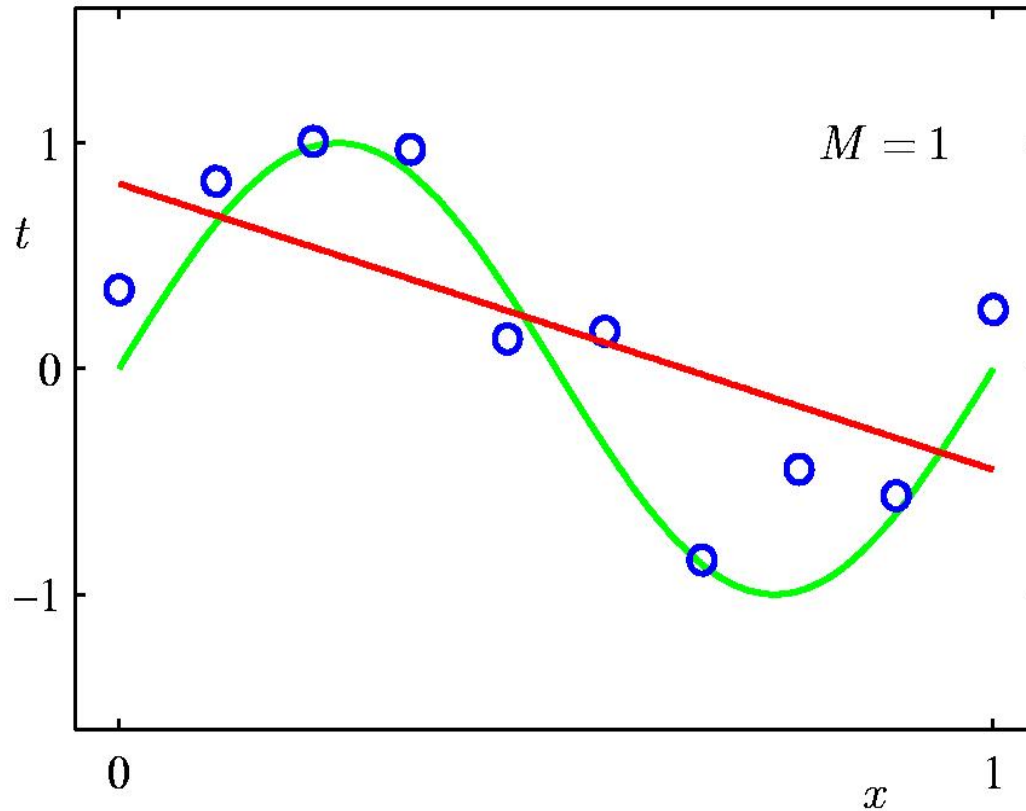


$$E(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^N \{y(x_n, \mathbf{w}) - t_n\}^2$$

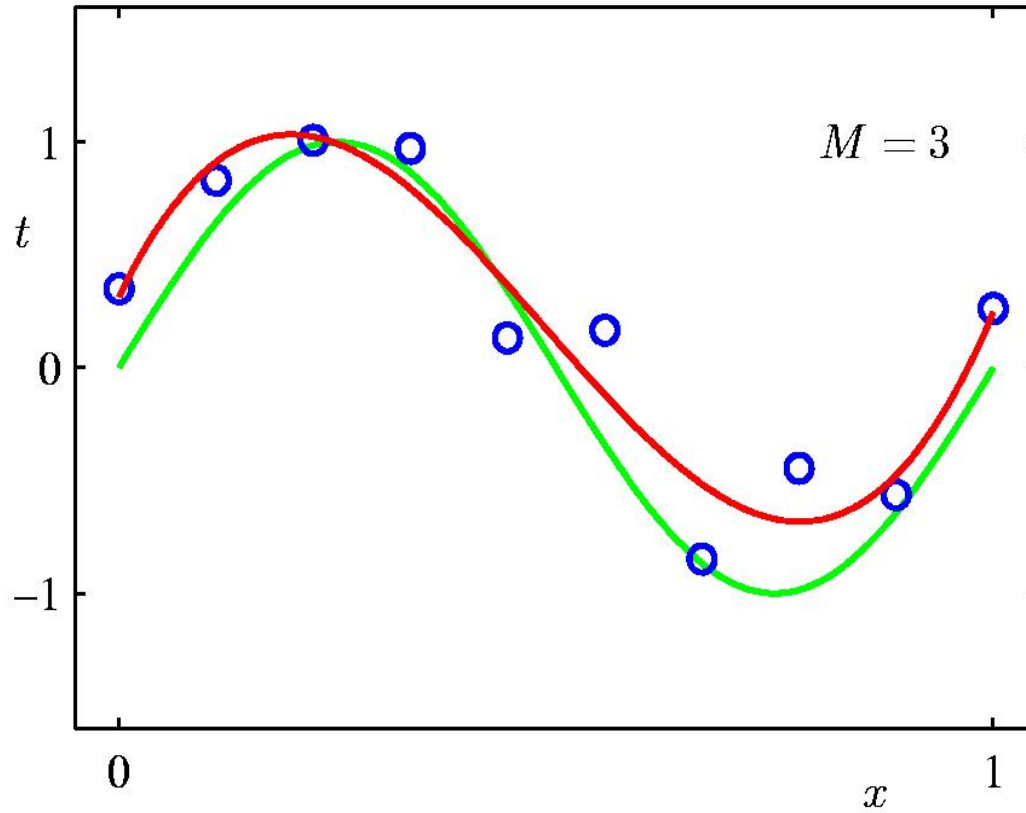
Esempio di overfitting: polinomio di ordine 0



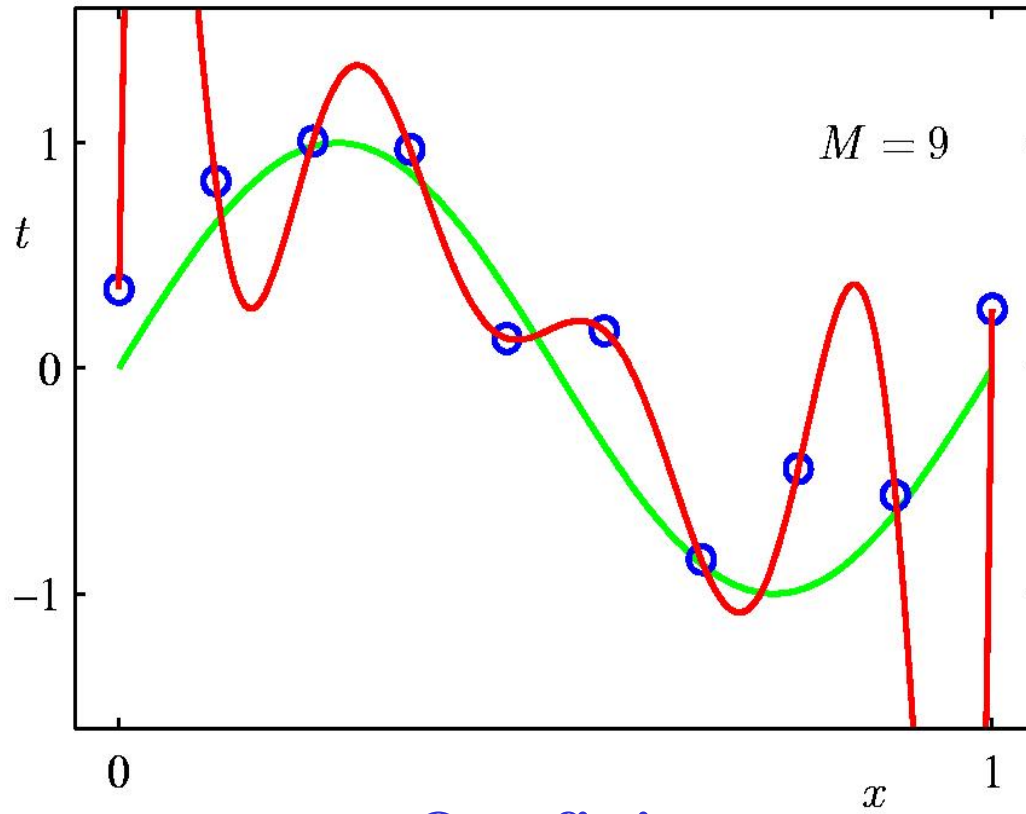
Esempio di overfitting: polinomio di ordine 1



Esempio di overfitting: polinomio di ordine 3



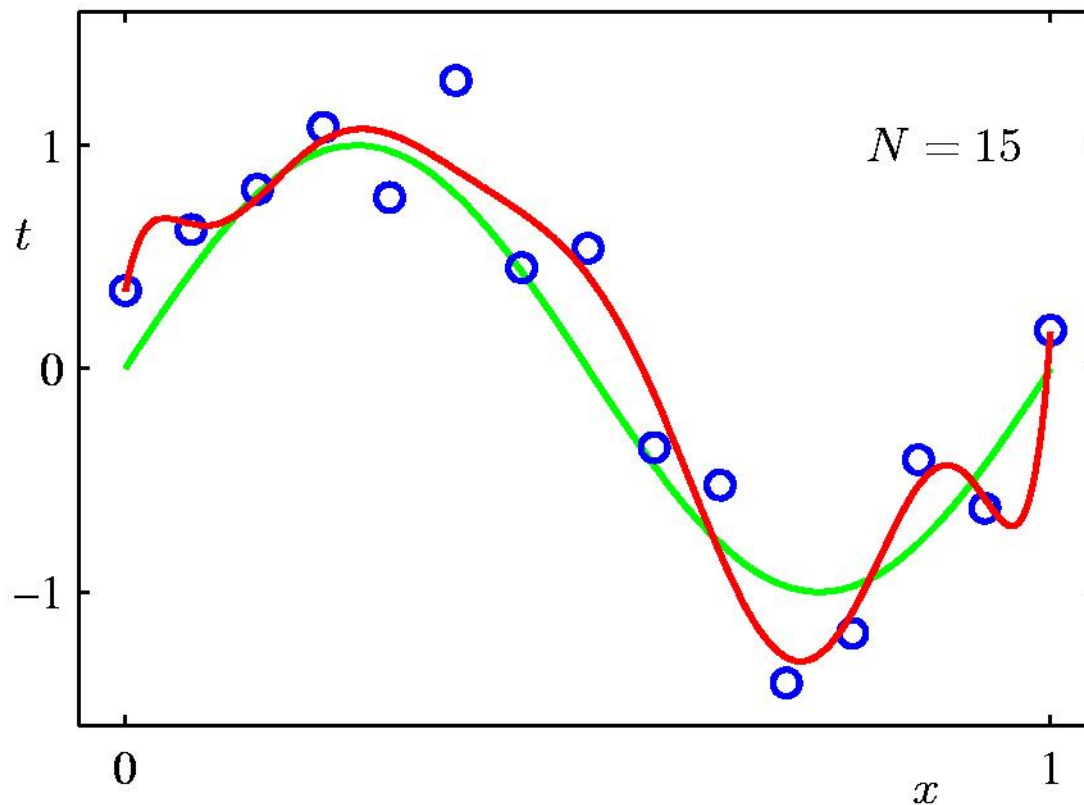
Esempio di overfitting: polinomio di ordine 9



Overfitting

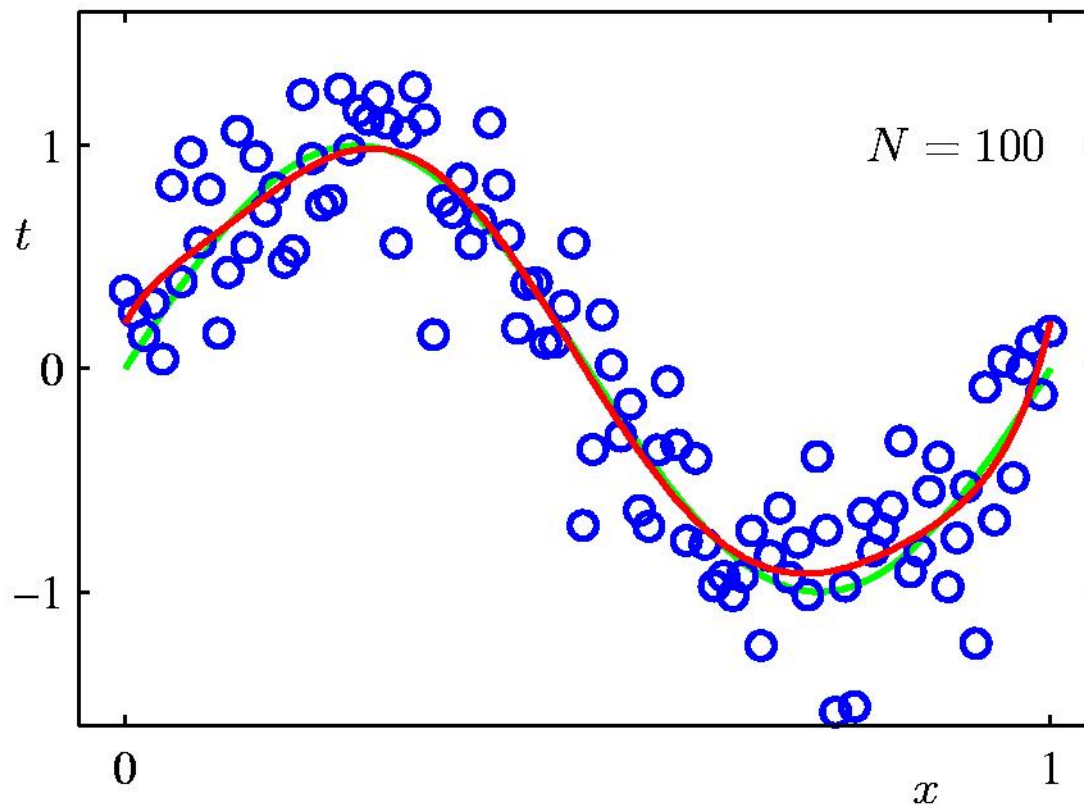
Esempio di overfitting: la cardinalità N entra in gioco

Polinomio di ordine 9th $N = 15$





Esempio di overfitting: la cardinalità N entra in gioco

Polinomio di ordine 9th $N = 100$



Approcci alla Pattern Recognition

- **Approccio sintattico:** approccio gerarchico.
Analogia tra la struttura dei patterns e la sintassi di un linguaggio:
 - patterns  frasi di un linguaggio
 - sottopattern primitivi  alfabeto
- **Approccio statistico:**
 - ad ogni pattern viene associato un vettore di feature che rappresenta un punto nello spazio multidimensionale del problema.
 - L'informazione sul problema, le dipendenze tra i vari fattori e i risultati prodotti sono tutti espressi in termini di probabilità.

Classificazione statistica

- La descrizione statistica di oggetti utilizza descrizioni numeriche elementari chiamate *feature*, che formano i cosiddetti *pattern*, $\mathbf{x} = (x_1, x_2, \dots, x_n)$ o vettori di *feature*.
- L'insieme di tutti i possibili pattern forma lo spazio dei pattern o delle *feature*.
- Nella classificazione statistica, alle entità di interesse sono associate delle distribuzioni di probabilità
- Un pattern è un'entità di interesse, quindi anche ad essa è associata una probabilità!

Classificatore di Bayes

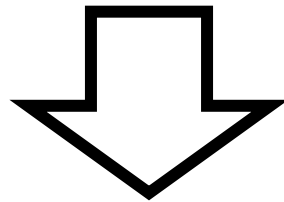
- Framework di base per la classificazione statistica
- Lavagna

Stima delle pdf

- *Classificatori parametrici*
 - si fissa il modello della distribuzione e sulla base del training set se ne stimano i parametri;
 - esempio: classificatore gaussiano.
- *Classificatori non parametrici*
 - nessuna assunzione sulla forma della pdf, la stima si basa esclusivamente sui dati;
 - esempio: K-nearest Neighbor.
- *Classificatori semi-parametrici*
 - si ha una classe molto generale di modelli di pdf, in cui il numero di parametri può essere aumentato in modo sistematico per costruire modelli sempre più flessibili;
 - esempio: reti neurali.

K Nearest Neighbor (KNN)

- Classificatore non parametrico.
- Molto utilizzato per la sua semplicità, flessibilità e ragionevole accuratezza dei risultati prodotti.
- IDEA: due elementi della stessa classe avranno, molto probabilmente, caratteristiche simili, cioè saranno vicini nello spazio dei punti che rappresenta il problema



La classe di un punto può essere determinata analizzando la classe dei punti in un suo intorno

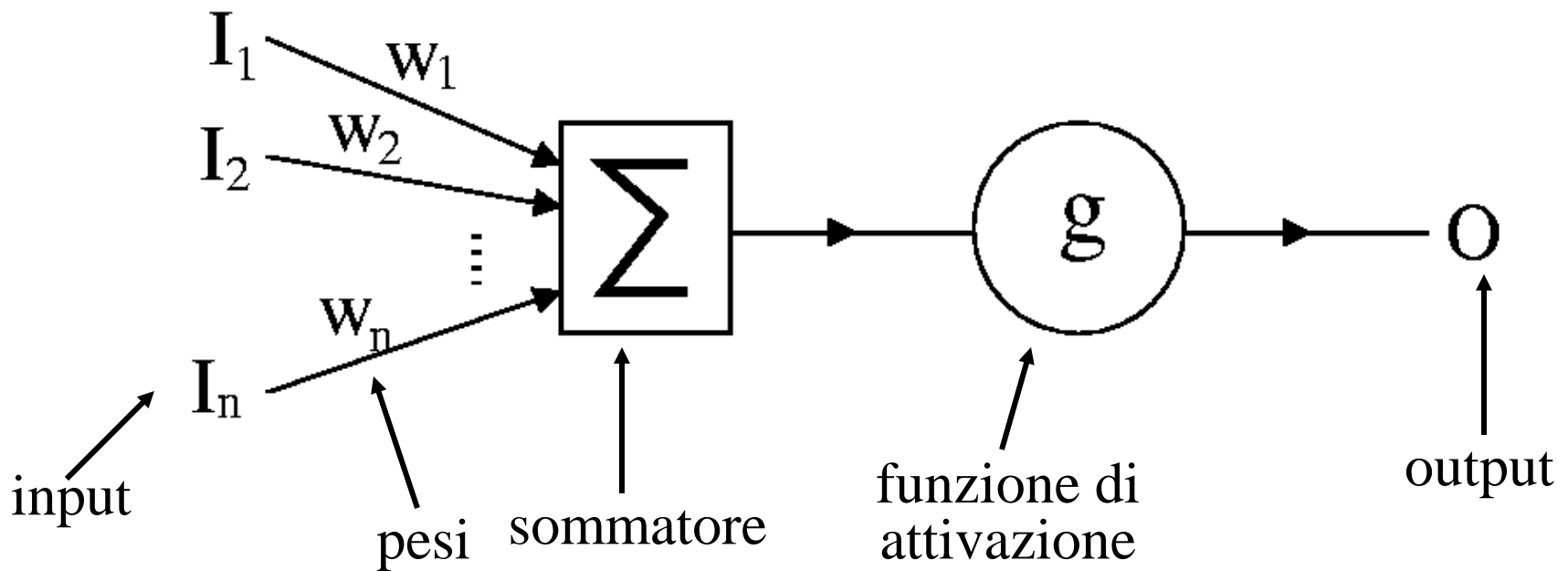
Algoritmo

- Dato un insieme di esempi X , dato un punto da classificare x_0 :
 - si calcola l'insieme U dei K punti di X più vicini a x_0 secondo una determinata metrica Σ (di solito la distanza euclidea);
 - si calcola la classe C più frequente all'interno dell'insieme U ;
 - x_0 verrà classificato come appartenente a C .
- Problema: scelta del parametro K e della metrica Σ .

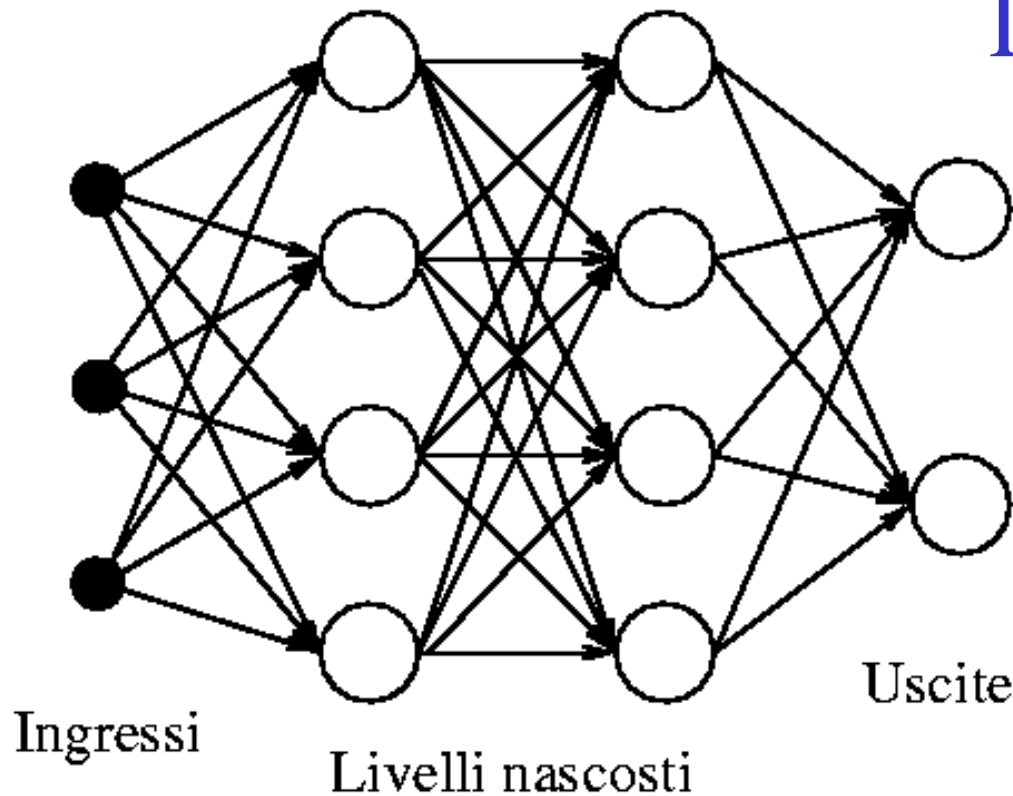
Reti neurali, un caso (quasi) a parte

- Sistema artificiale di elaborazione dell'informazione che emula il sistema nervoso animale.
- Caratteristiche del sistema nervoso animale:
 - robusto e resistente ai guasti;
 - flessibile, si adatta a situazioni nuove imparando;
 - lavora anche con informazione approssimata, incompleta o affetta da errore;
 - permette un calcolo altamente parallelo;
 - piccolo e compatto.

- Reti neurali: struttura complessa, composta da tante unità elementari di calcolo collegate tra loro in vario modo.
- Le unità elementari sono dette *neuroni*.
- I collegamenti sono detti *sinapsi*.



Diverse topologie



*Feed forward
neural networks*

Combinazioni di classificatori

- Problema poco studiato e solo recentemente ripreso.
- Assumono
 - classificatori diversi e con prestazioni diverse e non ottime;
 - training set diversi;
 - diversi classificatori addestrati con uguali training set e quindi con prestazioni diverse;
 - ugual classificatori addestrati differentemente (NN) e che risultano avere differenti prestazioni.
- Obiettivo di aumentare le prestazioni
- Approcci:
 - parallelo: si combinano i risultati dei singoli classificatori;
 - seriale: i risultati di uno sono input del successivo fino al risultato finale;
 - gerarchico: i classificatori sono strutturati ad albero e i risultati combinati adeguatamente.

In sintesi ...

- ... nell'uso di questi algoritmi di apprendimento bisogna tener conto di
 - complessità del modello (*model selection*)
 - fase di addestramento *online* o *batch*
 - disponibilità e tipo di dati di training, test e validazione
 - meccanismi di stop della fase di learning (pb di *overfitting*)
 - valutare l'uso di diversi modelli o dello stesso tipo di modello addestrato in modo diverso (insiemi di classificatori)
 - bilanciamento dei dati a disposizione
- Spesso, tutti questi modelli sono **adattati** allo specifico problema/applicazione da affrontare, sono spesso necessarie **metodi e modelli *ad hoc***

APPLICAZIONI

Il decollo della Pattern Recognition

- Fattori che hanno decretato il decollo definitivo della pattern recognition negli ultimi anni:
 - aumento della capacità computazionale dei calcolatori.
 - presenza di grosse quantità di dati
 - Internet
 - Social media (Facebook, Flickr, Pinterest)
 - nuovi sistemi di interazione uomo-macchina, uomo-robot, domotica.

Statistical Pattern Recognition (SPR)

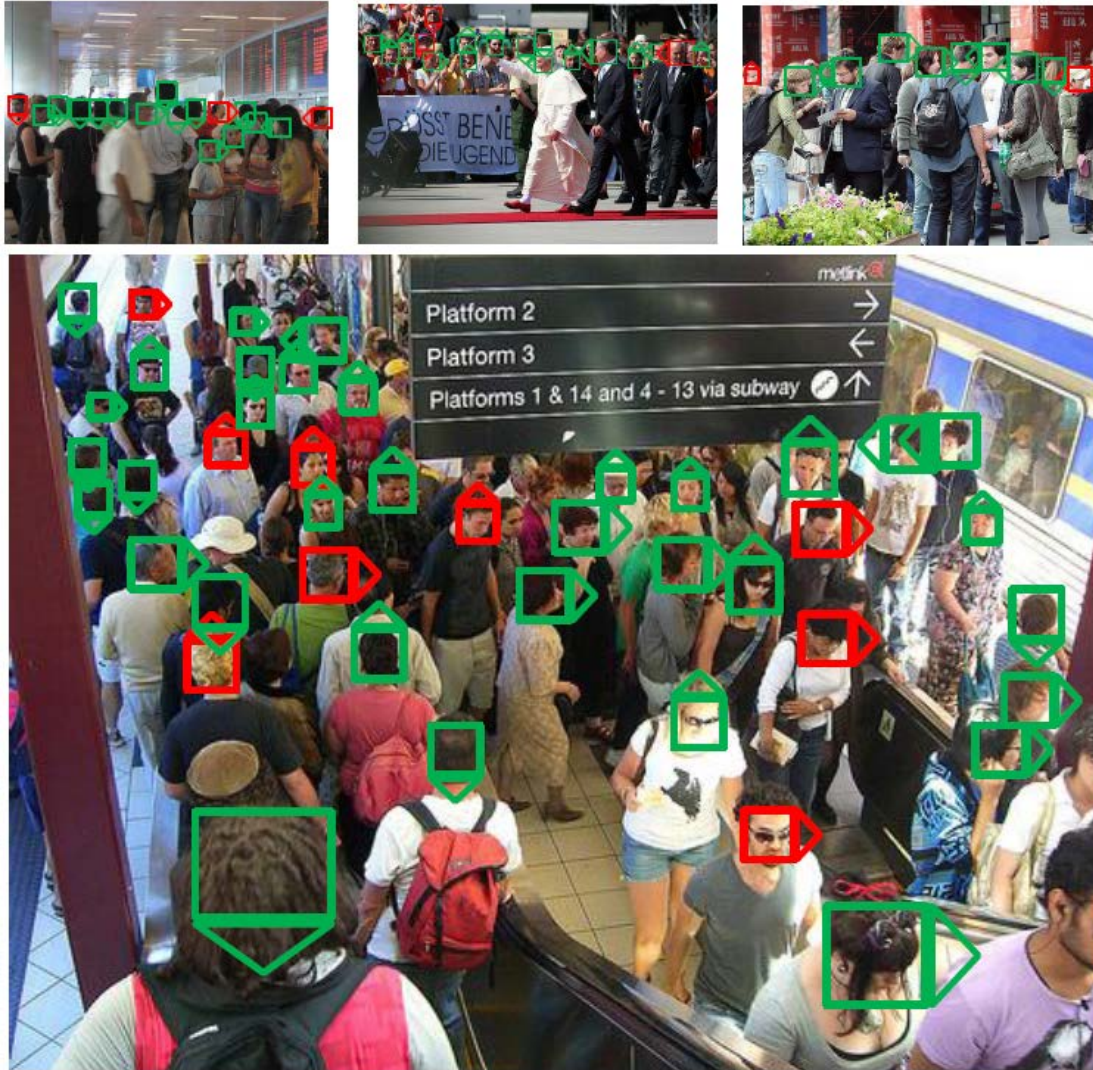


Costruire programmi che eseguano riconoscimento

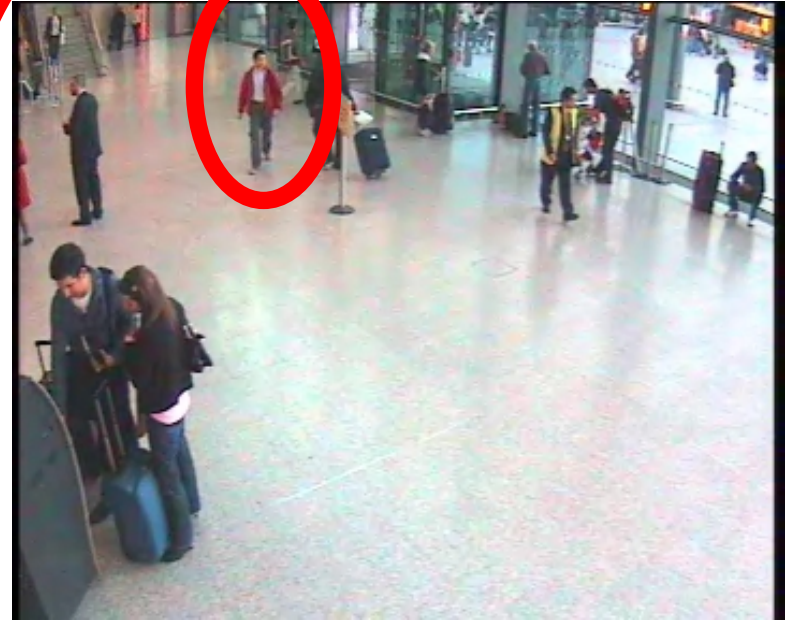
Video Analytics



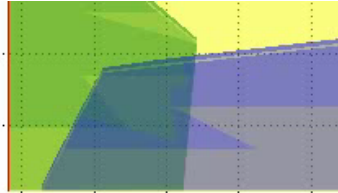
Video Analytics



Video Analytics



Social Signalling



Modeling Features John Pattern
 Classification Tracking Segmentation Analysis
 Shape TPORa Estimation
 Method Multiple
 Wavelet
 Models Scene
 Recognition
 RealTime

- *Captare il comportamento di*
- *Sorveglianza avanzata*



Social Signal Processing Network



Applicazioni classiche

- Riconoscimento del parlato:

- problematiche:

- tono
 - voce
 - velocità
 - stato d'animo

- applicazione: Siri, iOS 6



- Riconoscimento di caratteri scritti a mano:

- problematiche:

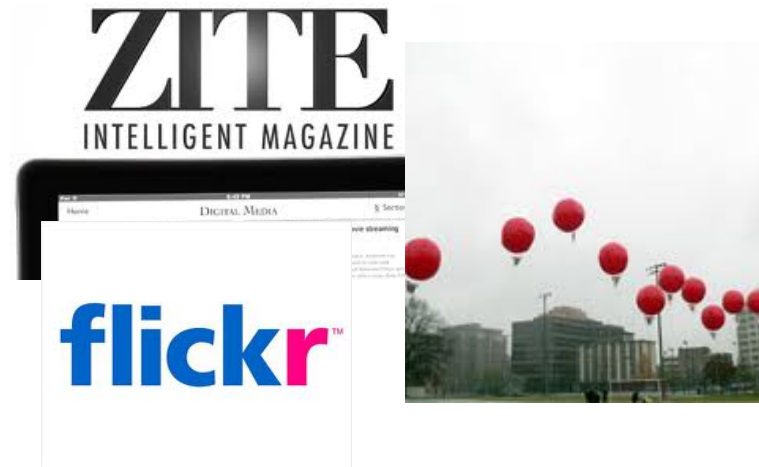
- grafia
 - stato d'animo

- applicazione: lettura automatica firme per supermercati, banche, etc



Data Mining

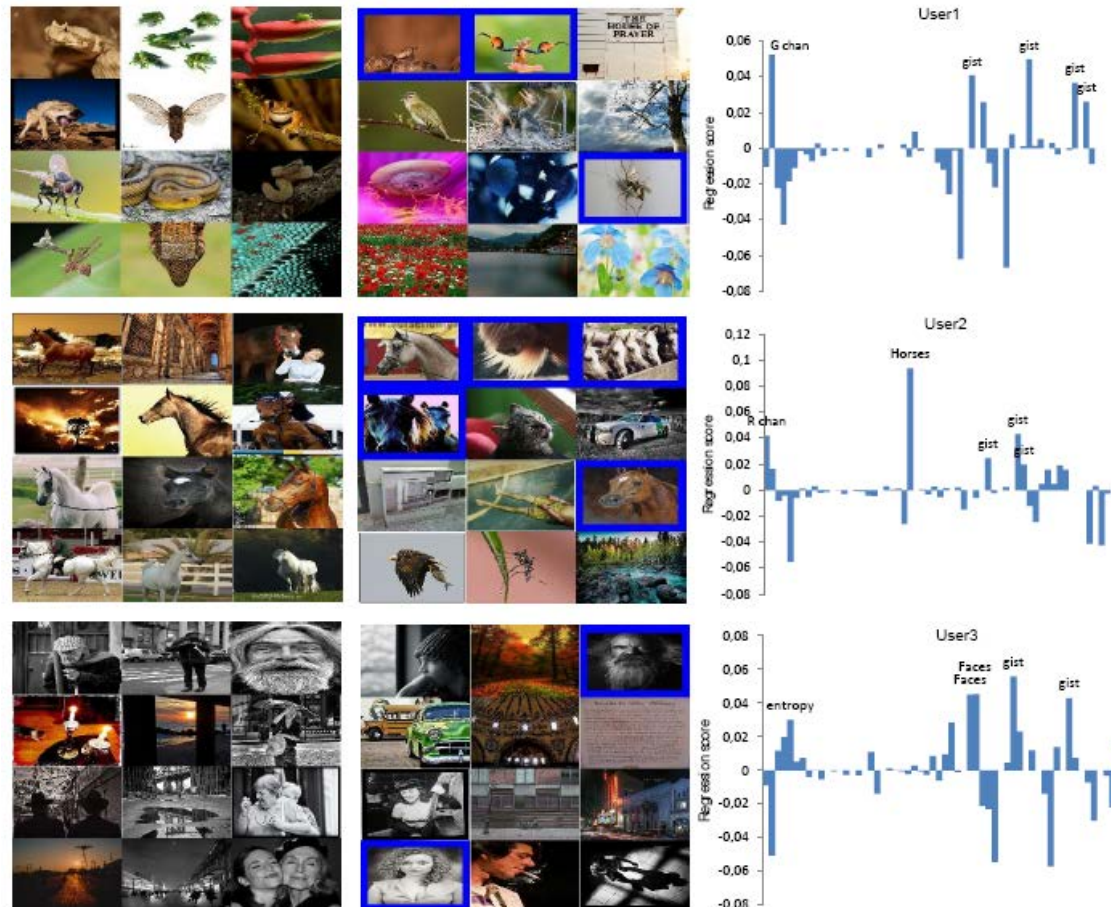
- Definizione:
 - estrazione di conoscenza da un insieme (tipicamente molto vasto) di dati multidimensionali.
- Scopi: predizione, classificazione, clustering, analisi delle associazioni, etc..
- I social media sono diventati gli obiettivi principali del data mining
 - Recommender systems
 - Crowdsourcing
 - Crowdsearching



<http://www.cs.ubc.ca/~nando/index.php>

http://en.wikipedia.org/wiki/DARPA_Network_Challenge

Data Mining - crowdsearching





Cristani et al. 2012, Tell Me What You Like and I'll Tell You What You Are: Discriminating Visual Preferences on Flickr Data ACM multimedia 2

Image retrieval by content

- Definizione:
 - *image retrieval*: trovare, in un database, immagini o sequenze di immagini rispondenti ad una determinata query;
 - *by content*: la ricerca avviene sulla base del contenuto dell'immagine, non più sulla base di un testo (annotato a mano su tutte le immagini del data set);
- Esempi di *query*:
 - “Trovami tutte le immagini simile ad un immagine data”.
 - “Trovami tutte le immagini che contengono un cavallo”.

Esempio

Main Image Class ->	Photographs	Graphs
Option 1	Click <i>Random</i>	Click <i>Random</i>
Option 2	Query image URL or ID <input type="text"/>	Query image URL or ID <input type="text"/>
Option 3	Start with 	Start with 

Ricerca di tutte le foto simili a questa

Risultato: c'è anche un punteggio sull'affidabilità del retrieval



6282 1.00 5



6290 0.92 4



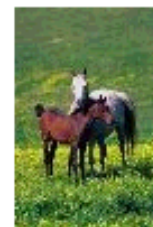
6265 0.92 5



6296 0.90 3



6276 0.90 4



6278 0.90 5



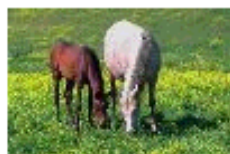
8785 0.90 4



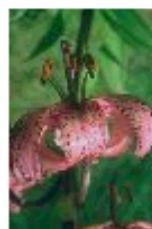
6277 0.90 4



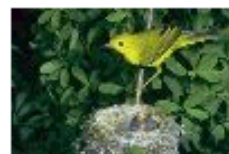
6283 0.90 6



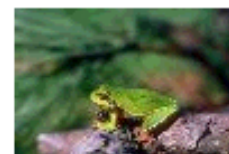
6288 0.89 5



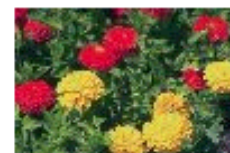
11050 0.89 4



4705 0.89 4



4938 0.89 6



44640 0.89 6



7123 0.89 6



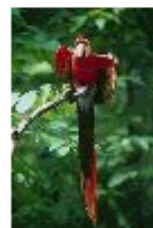
19512 0.89 5



44985 0.89 6



8710 0.89 4



52833 0.89 6



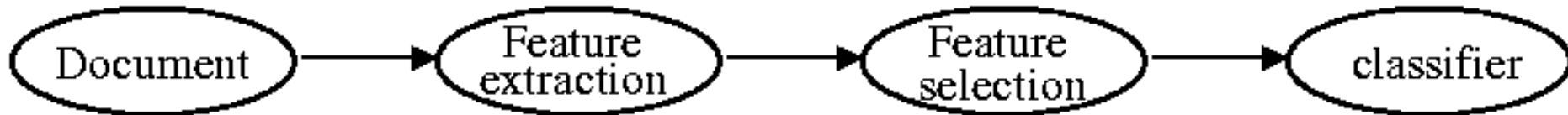
6294 0.89 5



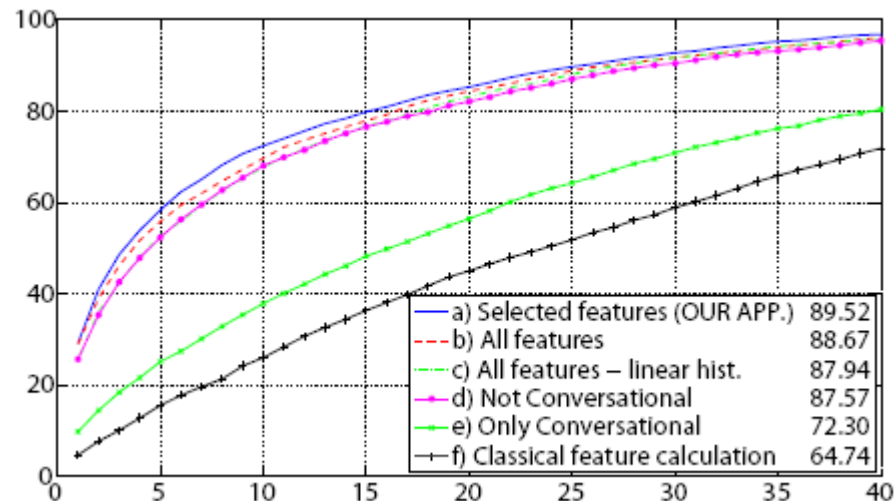
6226 0.89 4

Classificazione di documenti

- Definizione:
 - classificazione di documenti sulla base dell'argomento contenuto (sport, economia, ...).
- Feature:
 - tipologie di parole, frequenze assolute e relativa
- Clustering
- Applicazioni: ricerche in internet, data mining.



Classificazione avanzata di documenti



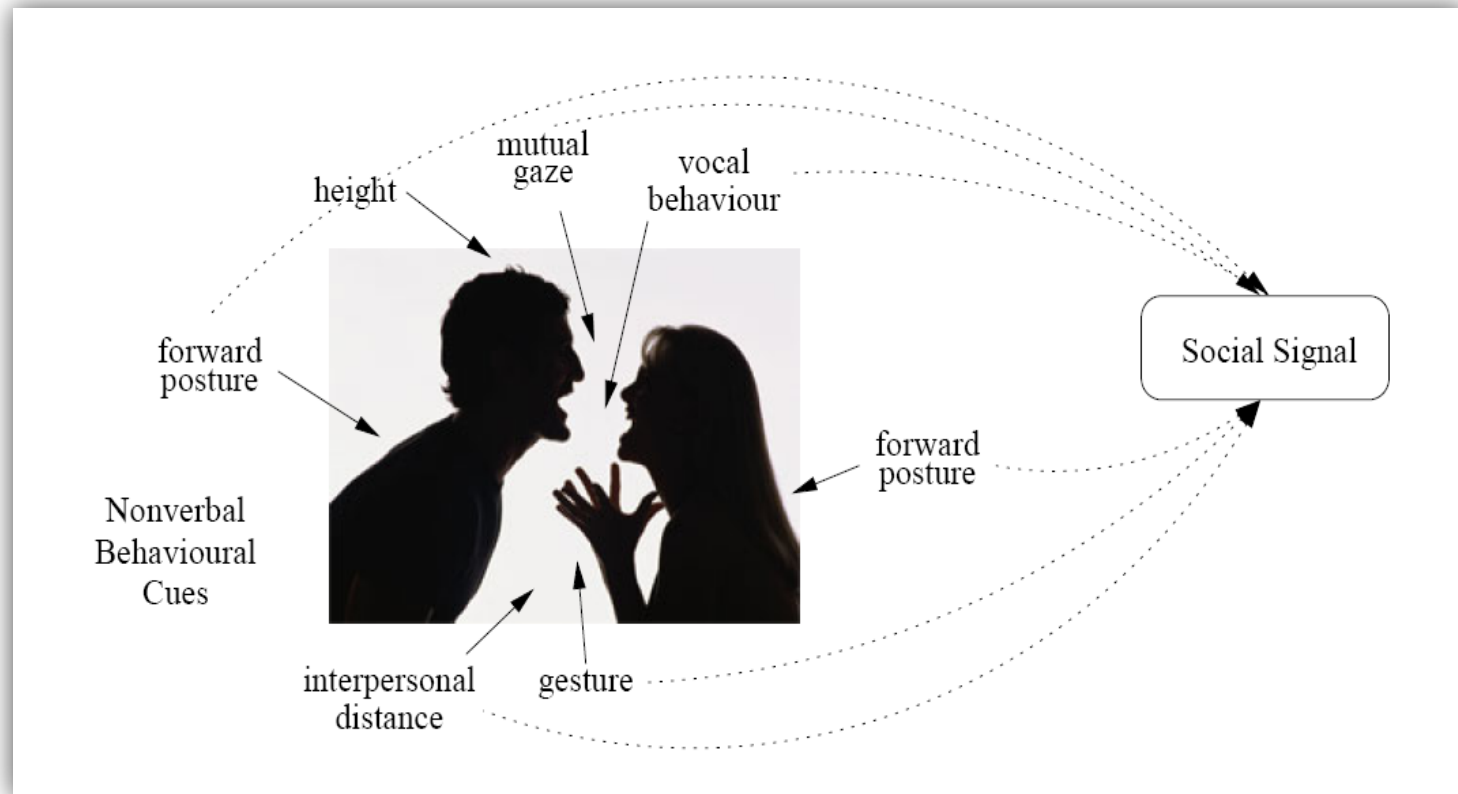
Riconoscimento di gesti

- Sistema che identifica gesti umani e li utilizza per portare informazione, oppure per il controllo di dispositivi.



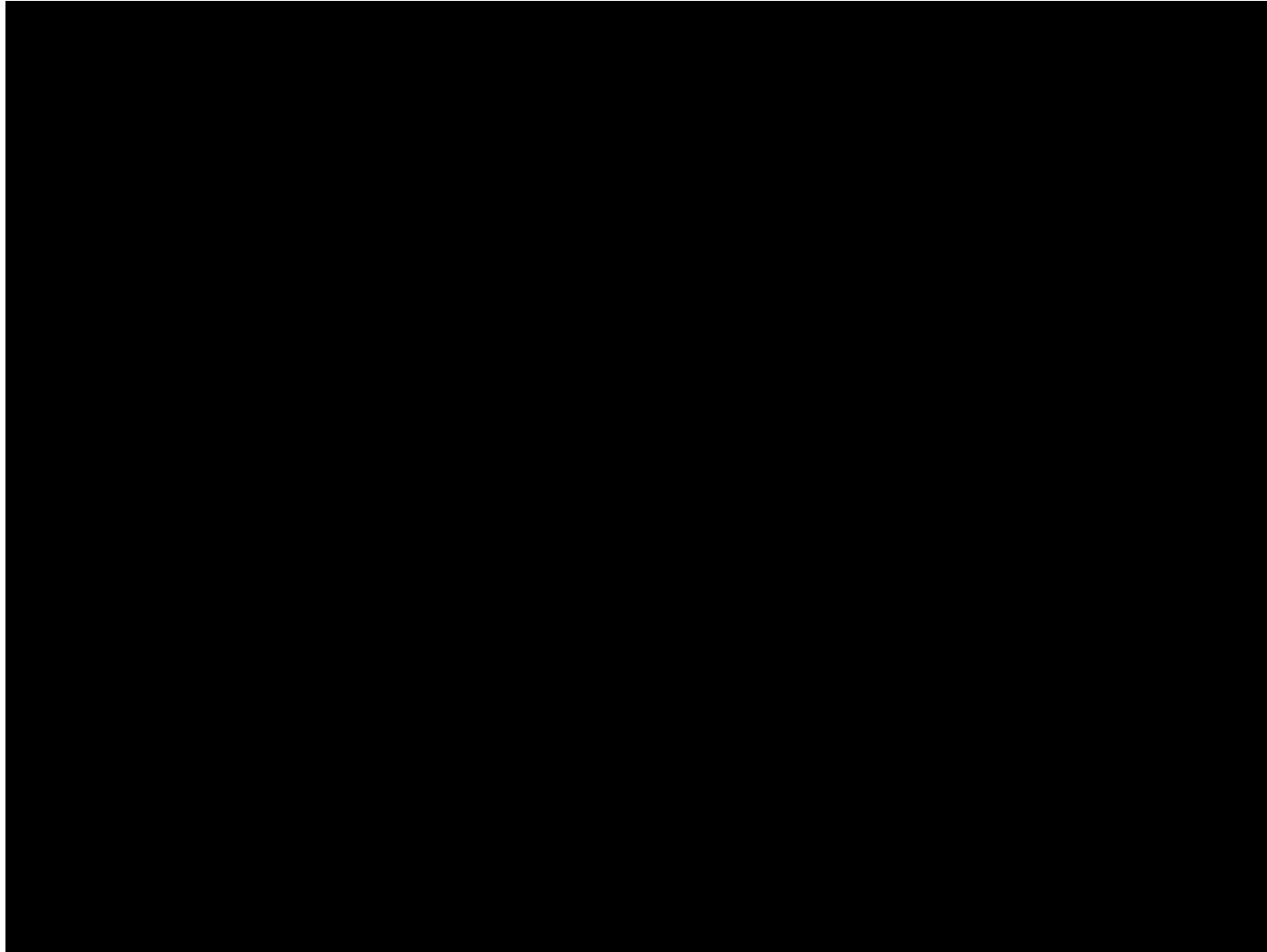
KINECT
for  XBOX 360

Social signaling



<http://sspnet.eu/>

Social signaling

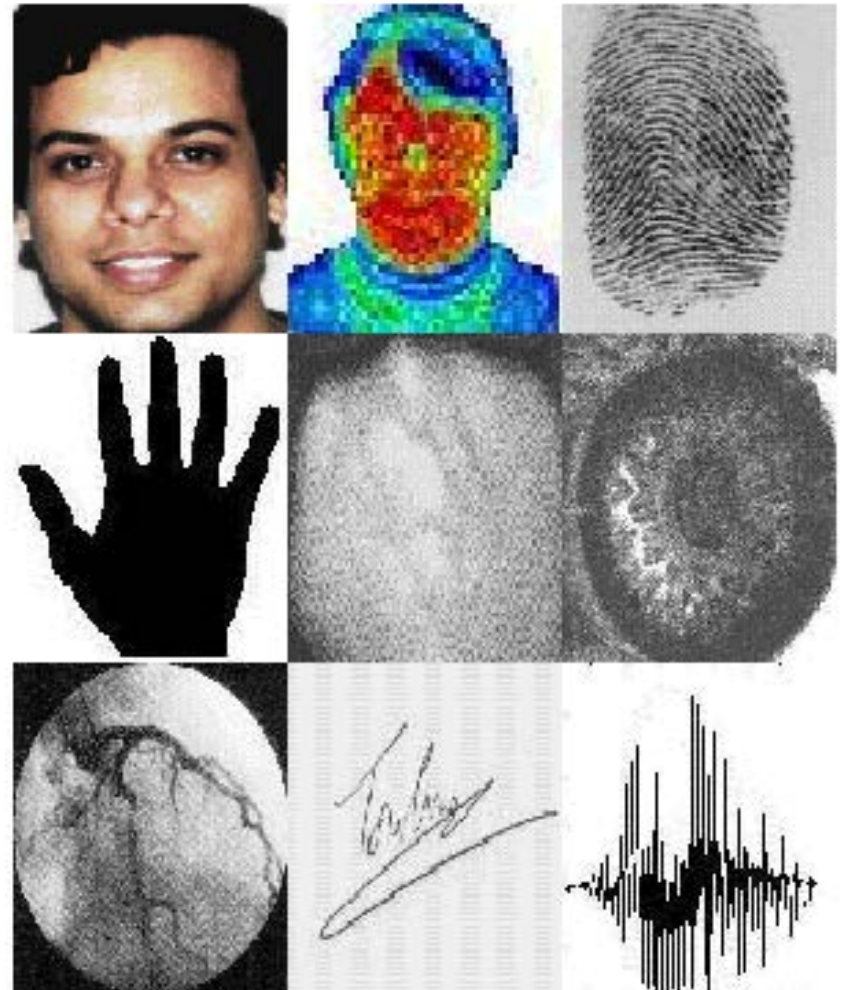


A. Pesarin, M. Cristani, V. Murino, A. Vinciarelli, Conversation Analysis at Work: Detection of Conflict in Competitive Discussions through Semi-Automatic Turn-Organization Analysis, Cognitive Processing, accepted on 12-09-2011

Biometria

- Definizione:
 - identificazione delle persone attraverso l'analisi delle loro caratteristiche fisiologiche e/o comportamentali.
- Caratteristiche del fattore biometrico:
 - universale (presente in ogni individuo);
 - unico (diverso in ogni individuo);
 - permanente (non rimovibile);
 - quantificabile (misurabile).

- Fattore biometrico:
 - faccia
 - termogramma facciale
 - impronte digitali
 - geometria della mano
 - firma
 - voce
 - iride
- Valutazione di un sistema di biometria:
 - performance;
 - sicurezza;
 - accettabilità.



<http://biometrics.cse.msu.edu/>

Stage?



NERVIANO MEDICAL SCIENCES



OMRON

