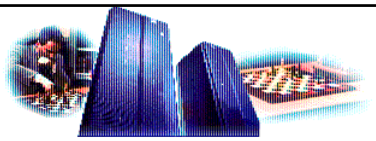
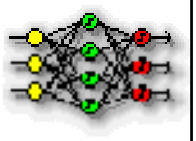




Sistemi Intelligenti
Learning and Clustering

Alberto Borghese and Iuri Frosio

Università degli Studi di Milano
Laboratorio di Sistemi Intelligenti Applicati (AIS-Lab)
Dipartimento di Scienze dell'Informazione
borghese@dsi.unimi.it

Riassunto

- I tipi di apprendimento
- Il clustering
- K means
- Quad-tree decomposition

I vari tipi di apprendimento

$x(t+1) = f[x(t), a(t)]$ Ambiente
$a(t) = \mathbf{g}[x(t)]$ Agente

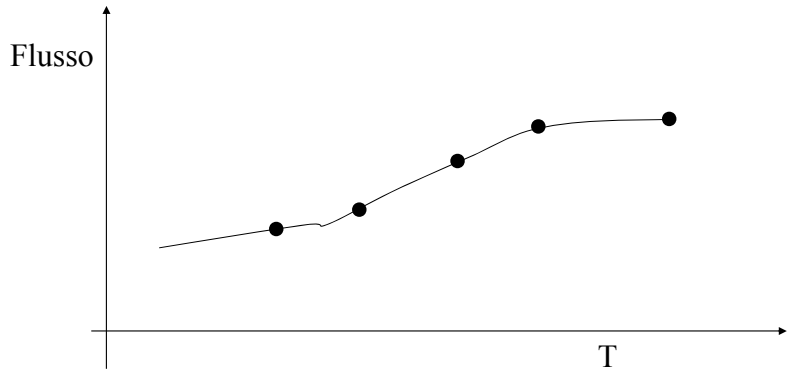
Supervisionato (learning with a teacher). Viene specificato per ogni pattern di input, il pattern desiderato in output.

Semi-Supervisionato. Viene specificato solamente per **alcuni** pattern di input, il pattern desiderato in output.

Non-supervisionato (learning without a teacher). Estrazione di similitudine statistiche tra pattern di input. Clustering. Mappe neurali.

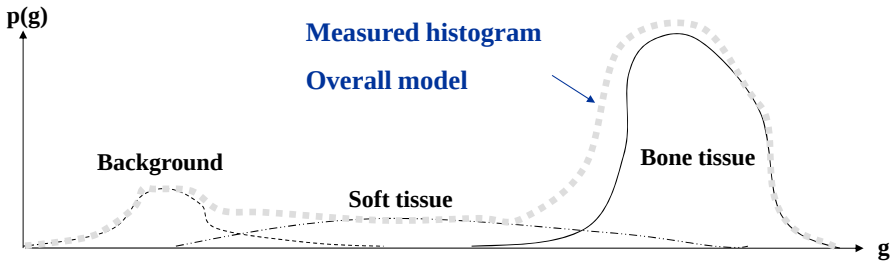
Apprendimento con rinforzo (reinforcement learning, learning with a critic). L'ambiente fornisce un'informazione puntuale, di tipo qualitativo, ad esempio success or fail.

Apprendimento supervisionato: regressione = predictive learning



Controllo della portata di un condizionatore in funzione della temperatura. "Imparo" una funzione continua a partire da alcuni campioni: devo imparare ad **interpolare**.

I modelli parametrici



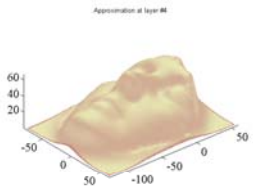
$$p(g)=\sum_{j=1}^{M} P(j)\cdot p(g\mid j)=\sum_{j=1}^{M} w_j \cdot p_j(g)$$

La probabilità di avere un livello di grigio g è la somma pesata delle tre probabilità di avere background, $p_1(g)$, tessuto molle, $p_2(g)$ o tessuto osseo, $p_3(g)$.

I modelli semi-parametrici

- L'approssimazione è ottenuta mediante funzioni "generiche", dette di **base,** soluzione molto utilizzata nelle NN e in Machine learning.
- (Il concetto di Base in matematica è definito mediante certe proprietà di approssimazione che qui non consideriamo, consideriamo solo l'idea intuitiva).

$$z(p(x,y))=\sum_{i} w_i G(p,p_i;\sigma)$$



Combinazione lineare di funzioni di base

Da calcolare

Funzione di base (fissate)

Modelli lineari e non lineari

Classificazione alternativa dei modelli. Vengono utilizzate classi molto diversi di algoritmi per stimare i parametri di questi due tipi di modelli.

$$z(p(x,y)) = \sum_i w_i G(p, p_i; \sigma)$$

$$z(p(x,y)) = \sum_i f_i\left(G(p, p_i; \sigma)\right)$$

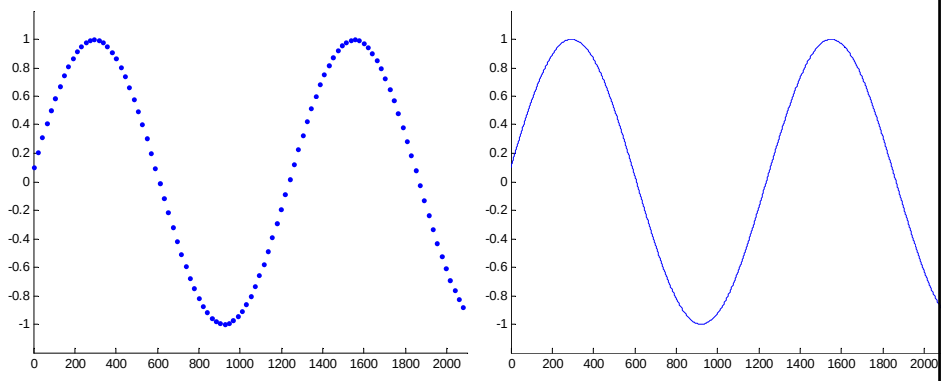
f(.) = w_i è funzione lineare

G( ) è data a-priori, i suoi parametri nondevono essere determinati

f(.) è funzione non lineare

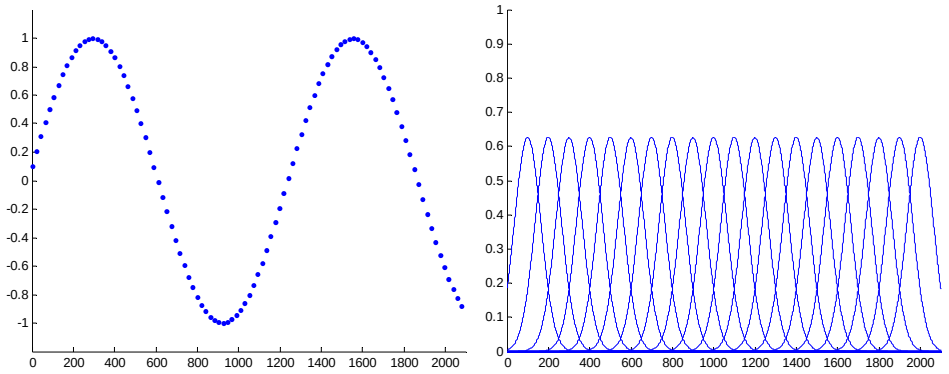
e.g. $f(.) = e^{G(.)}$

Funzionamento di un modello parametrico (non-lineare)



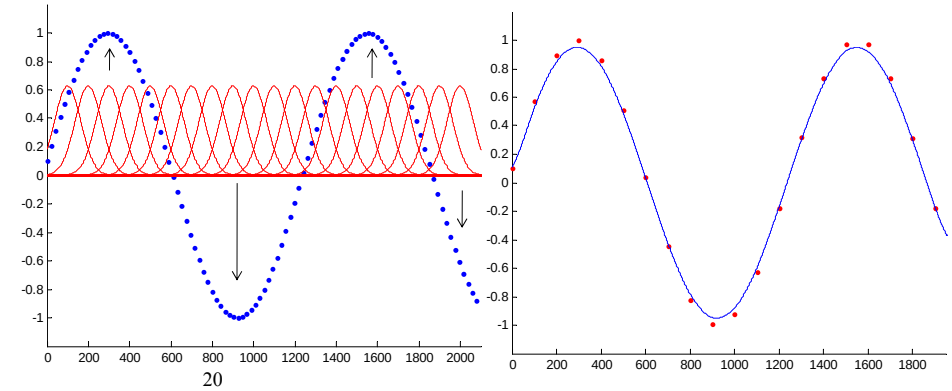
I punti vengono fittati perfettamente da una sinusoide: $y = \mathbf{A} \sin(\omega x + \phi)$. Devo determinare i parametri della sinusoide (non lineare), i cui valori ottimali sono: $\omega = 1/200$, $\phi = 0.1$.

Approssimazione mediante un modello semi-parametrico (lineare)



Vogliamo fittare i punti con l'insieme di Gaussiane riportate sulla dx. In questo caso hanno tutte $\sigma = 90$. Come le utilizzo?

Funzionamento di un modello semi-parametrico (lineare)

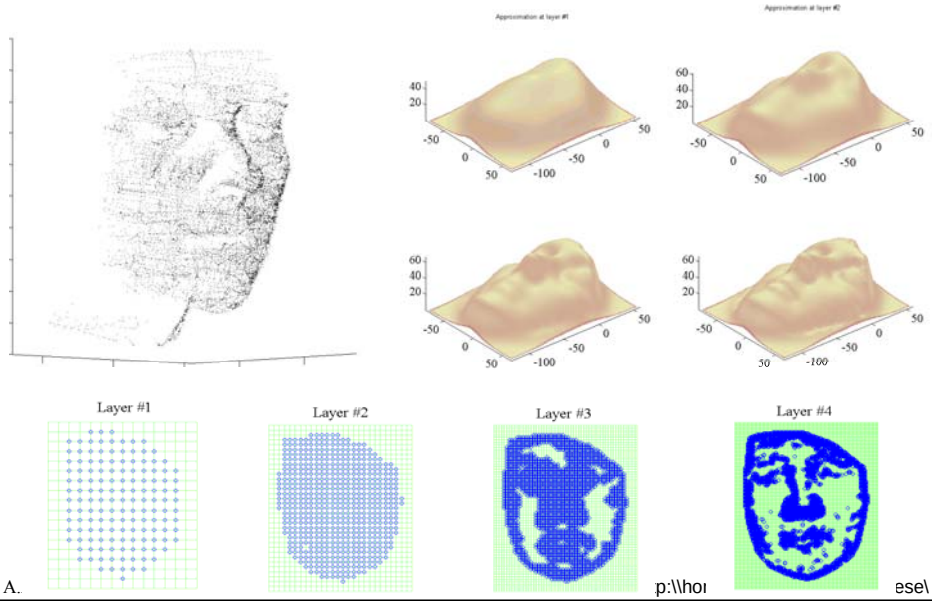


$$y(x) = \sum_{i=1}^{20} w_i G(x - x_{o_i}; 90)$$

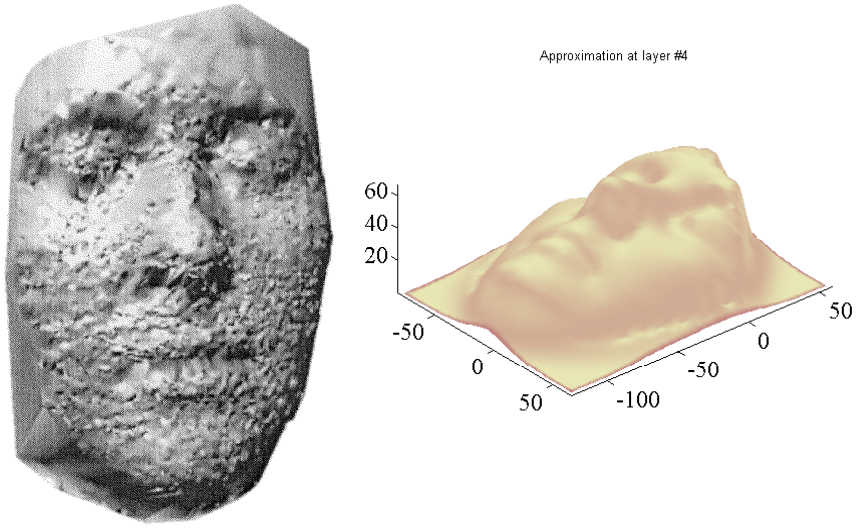
Devo definire, i $\{w_i\}$

I σ sono tutti uguali ed uguali a 90, le Gaussiane sono equispaziate.
Le Gaussiane sono note tutte a priori, devono essere definiti i pesi.

Applicazione della regressione



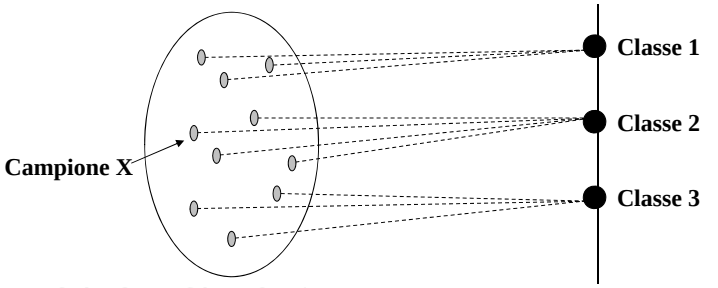
Quante unità semi-parametriche? (rischio di overfitting)



Classificazione

Un'interpretazione geometrica:

Mappatura dello spazio dei campioni nello spazio delle classi.



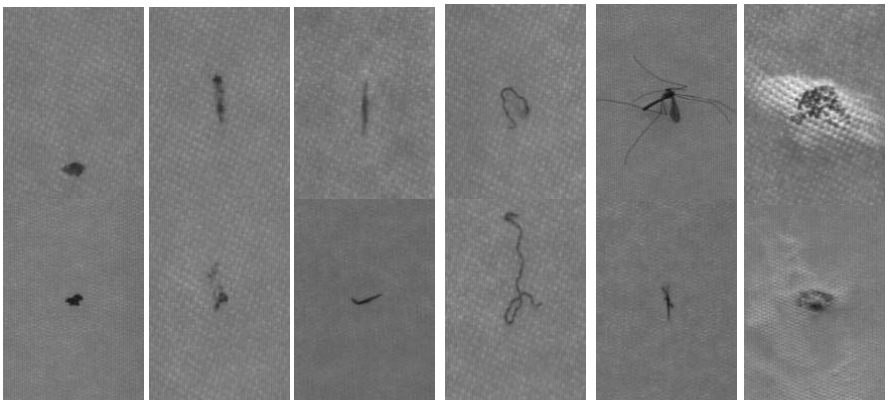
SPAZIO DEI CAMPIONI / DELLE FEATURES (CARATTERISTICHE)

SPAZIO DELLE CLASSI

Apprendimento Supervisionato: Classificazione

Task di classificazione
Uscita intera (etichetta o label della classe)

CLASSIFICAZIONE: Riconoscimento difetti in linee di produzione
(progetto finanziato da Electronic Systems: 2006-2007)



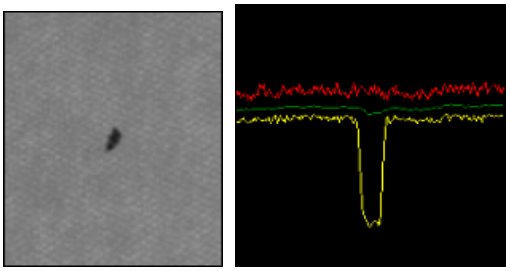
regolari | irregolari | allungati | fili | insetti | macchie su denso

Difetti – Classificazione real-time e apprendimento mediante **booosting.**
Committee (linear combinaion) of weak (binary) classifiers.

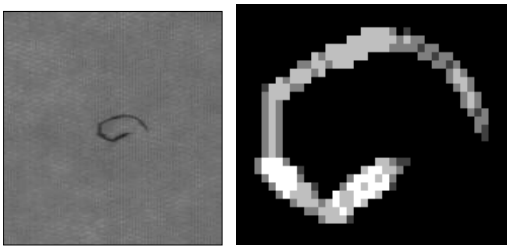
Features

- *Località.*
- *Significatività.*
- *Rinoscibilità.*

Macchie dense



Fili

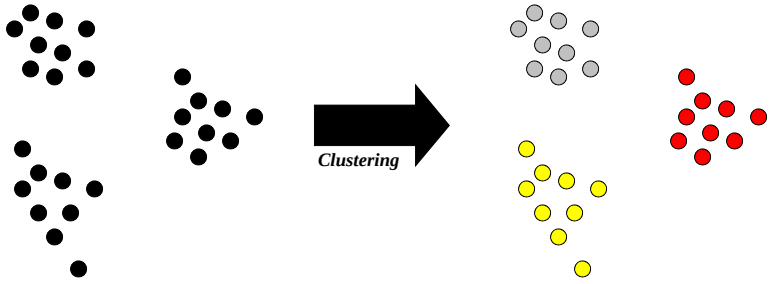


Riassunto

- I tipi di apprendimento
- Il clustering
- K means
- Quad-tree decomposition

Clustering

- Clustering: raggruppamento degli "oggetti" in classi omogenee tra loro.
  - Raggruppamento per colore
  - Raggruppamento per forme
  - Raggruppamento per tipi
  - …..



Novel name: **data mining**

Esempio di clustering

Ricerca immagini su WEB.

Clustering -> Indicizzazione

Il clustering per…

… Confermare ipotesi sui dati (es. “E’ possibile identificare tre diversi tipi di clima in Italia: mediterraneo, continentale, alpino…”);

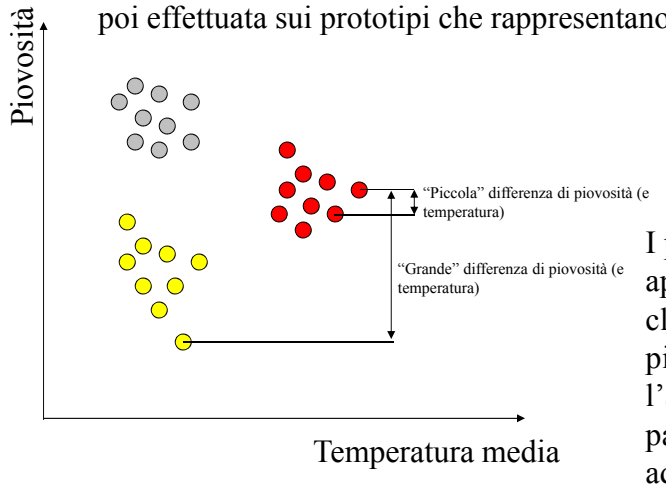
… Esplorare lo spazio dei dati (es. “Quanti tipi diversi di clima sono presenti in Italia?”);

… Semplificare l’interpretazione dei dati (“Il clima di ogni città d’Italia è approssimativamente mediterraneo, continentale o alpino.”).

…“Ragionare” sui dati o elaborare i dati in modo stereotipato.

Clustering

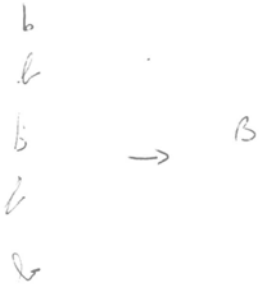
Processo attraverso il quale i dati (pattern, vettori) vengono organizzati in cluster, basata sulla similarità. L'elaborazione verrà poi effettuata sui prototipi che rappresentano ciascun cluster.



I pattern appartenenti ad un cluster valido sono più simili l'uno con l'altro rispetto ai pattern appartenenti ad un cluster differente.

Clustering: definizioni

- **Pattern**: un singolo dato $\mathbf{p} = [p_1, p_2, \ldots p_n]$. Il dato appartiene quindi ad uno spazio multi-dimensionale, solitamente eterogeneo.
- **Feature**: le caratteristiche dei dati significative per il clustering, possono costituire anch'esso un vettore, il vettore delle feature: $x_1, x_2, \ldots x_d$. Questo vettore costituisce l'input agli algoritmi di clustering.

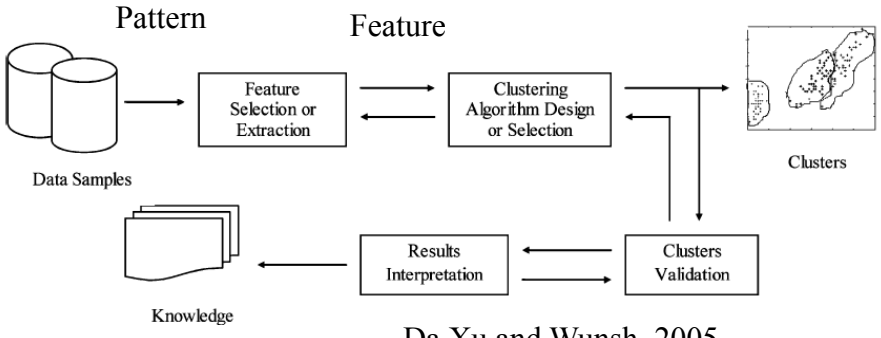


Inclinazione, occhielli, lunghezza, linee orizzontali, archi di cerchio ...

Clustering: definizioni

- **n**: dimensione dello spazio dei pattern;
- **d**: dimnesione dello spazio delle feature;
- **Cluster**: in generale, insieme che raggruppa dati simili tra loro, valutati in base alle feature;
- **Funzione di similarità o distanza**: una metrica (o quasi metrica) nello spazio delle feature, usata per quantificare la similarità tra due pattern.
- **Algoritmo**: scelta di come effettuare il clustering (motore di clustering).

Analisi mediante clustering



Da Xu and Wunsh, 2005

I cluster ottenuti sono significativi?
Il clustering ha operato con successo?

NB i cammini all'indietro consentono di fare la sintonizzazione dei diversi passi.

Il clustering

Per una buona review: Xu and Wunsch, IEEE Transactions on Neural Networks, vol. 16, no. 3, 2005.

Il clustering non è di per sé un problema ben posto. Ci sono diversi gradi di libertà da fissare su come effettuare un clustering.

Rappresentazione dei pattern;

Calcolo delle feature;

Definizione di una misura di prossimità dei pattern attraverso le feature;

Tipo di algoritmo di clustering (gerachico o partizionale)

Validazione dell'output (se necessario) -> Testing.

Problema a cui non risponderemo: **quanti cluster**? Soluzione teorica (criterio di Akaike), soluzione empirica (growing networks di Fritzke).

Rappresentazione dei pattern

- Feature selection: identificazione delle feature più significative per la descrizione dei pattern.

  Esempio: descrizione del clima e della città di Roma.
  Roma: [17°; 500mm; 1.500.000 ab.]

- Come scegliere i pattern?
  - Vicinanza ai bordi di ciascun cluster (Support Vector Machines)
  - Tutti i pattern

- Come valutare le feature?
  - Analisi statistica del potere discriminante: correlazione tra feature e loro significatività.

Feature & feature

- Feature extraction: trasformazione delle feature per creare nuove, significative feature;
- Elaborazione di primo livello, per ottenere informazioni caratteristiche del fenomeno che, ad esempio, siano inverianti.

Esempio: descrizione di oggetti circolari.

Posso misurare l'area e il perimetro, ma il loro rapporto è più significativo.

Esempio: descrizione del clima.

Milano: [13°; 900mm; 265 giorni sole; 100 giorni pioggia]

oppure

Milano: [13°; 900mm / 100 giorni pioggia; 265 giorni sole]

Similarità tra feature

- Definizione di una **misura di distanza tra due features**;

Esempio:
Distanza euclidea…
dist (Roma, Milano) = …
dist ([17°; 500mm], [13°; 900mm]) = …
= … Distanza euclidea? = …
$= ((17-13)^2+(500-900)^2)^{1/2} = 400.02 \sim 400$

Ha senso?

Normalizzazione feature

Att.ne!

dist (Roma, Milano) = …
dist ([17°; 500mm], [13°; 900mm]) = …
= … Distanza euclidea? = …
= $((17-13)^2+(500-900)^2)^{½} = 400.02 \sim 400$

La distanza tra le due città in termini di gradi è insignificante nel nostro calcolo… **E' necessario trovare una metrica corretta per la rappresentazione dei dati. Ad esempio, normalizzare le feature**!

$T_{Max}= 20°$ $T_{Min}= 5°$ → $T_{Norm} = (T - T_{Min})/(T_{Max} - T_{Min})$
$P_{Max}= 1000mm$ $P_{Min}= 0mm$ → $P_{Norm} = (P - P_{Min})/(P_{Max} - P_{Min})$
$Roma_{Norm} = [0.8\ 0.5]$
$Milano_{Norm} = [0.53\ 0.9]$
$dist(Roma_{Norm}, Milano_{Norm}) = ((0.8-0.53)^2+(0.5-0.9)^2)^{½} = 0.4826$

Altre funzioni di distanza

- Distanza euclidea:
  $dist(x,y)=[\Sigma_{k=1..d}(x_k-y_k)^2]^{1/2}$
- Minkowski:
  $dist(x,y)=[\Sigma_{k=1..d}(x_k-y_k)^p]^{1/p}$
- Mahalanobis:
  $dist(x,y)= (x_k-y_k)S^{-1}(x_k-y_k)$, con S matrice di covarianza.
- Context dependent:
  $dist(x,y)= f(x, y, context)$

Tassonomia (sintetica) degli algoritmi di clustering

- Algoritmi gerarchici (agglomerativi, divisivi), e.g. **Hirarchical clustering**.
- Algoritmi partizionali, hard: **K-means, quad-tree decomposition**.
- Algoritmi partizionali, soft: fuzzy c-mean, neural-gas, enhanced vector quantization, **mappe di Kohonen**.
- Algoritmi statistici: **mixture models**.

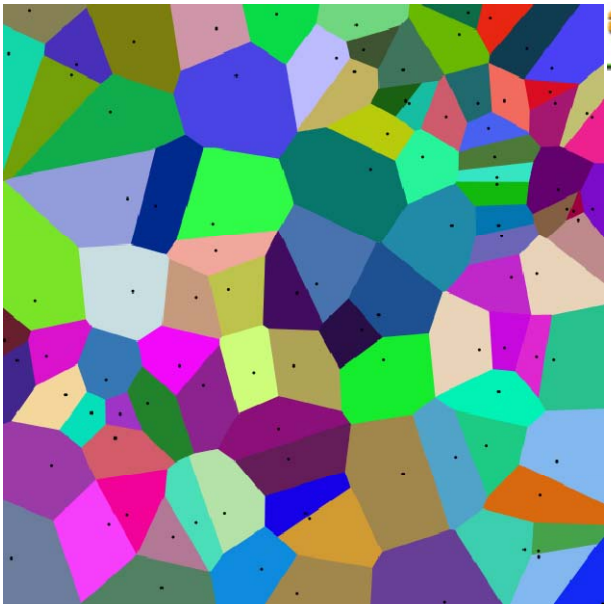
Riassunto

- I tipi di apprendimento
- Il clustering
- K means
- Quad-tree decomposition

K-means (partitional): framework

- Siano $\mathbf{X_1}$, …, $\mathbf{X_D}$ i dati di addestramento, features (per semplicità, definiti in R^2);
- Siano $\mathbf{C_1}$, …, $\mathbf{C_K}$ i *prototipi* di K classi, definiti anch'essi in R^2; ogni *prototipo* identifica il baricentro della classe corrispondente;
- Lo schema di classificazione adottato sia il seguente: "$\mathbf{X_i}$ appartiene a $\mathbf{C_j}$ se e solo se $\mathbf{C_j}$ è il *prototipo* più vicino a $\mathbf{X_i}$ (distanza euclidea)**";**
- L'algoritmo di addestramento permette di determinare le posizioni dei *prototipi* $\mathbf{C_j}$ mediante successive approssimazioni.

Risultato del clustering è un diagramma di Voronoj

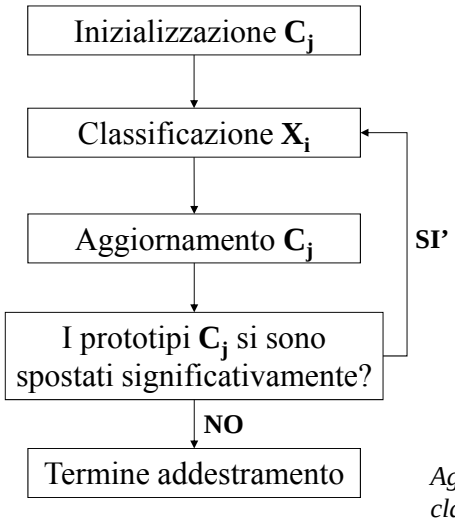


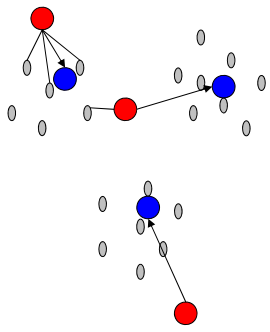
I poligoni azzurri rappresentano i diversi cluster ottenuti. Ogni punto marcato all'interno del cluster (cluster center) è rappresentativo di tutti i punti del cluster

Algoritmo K-means

L'obiettivo che l'algoritmo si prepone è di minimizzare la varianza totale intra-cluster. Ogni cluster viene identificato mediante un centroide o punto medio. L'algoritmo segue una procedura iterativa. Inizialmente crea K partizioni e assegna ad ogni partizione i punti d'ingresso o casualmente o usando alcune informazioni euristiche. Quindi calcola il centroide di ogni gruppo. Costruisce quindi una nuova partizione associando ogni punto d'ingresso al cluster il cui centroide è più vicino ad esso. Quindi vengono ricalcolati i centroidi per i nuovi cluster e così via, finché l'algoritmo non converge (Wikipedia).

K-means: addestramento





Aggiornamento C_j*: baricentro degli* X_i *classificati da* C_j.

Algoritmo K-means::formalizzazione

- Dati N pattern in ingresso $\{x_j\}$ e C_k prototipi che vogliamo diventino i centri dei cluster, x_j e $C_k \in R^N$. Ciascun cluster identifica una regione nello spazio, P_k.
- Valgono le seguenti proprietà:

$$\bigcup_{k=1}^{K} P_k = Q \supseteq R^D$$ I cluster coprono lo spazio delle feature

$$\bigcap_{k=1}^{K} P_k = \varnothing$$ I cluster sono disgiunti.

- $x_j \in C_k$ Se: $\left(\left|x_j - C_k\right|\right)^2 \leq \left(\left|x_j - C_l\right|\right)^2 \quad l \neq k$

- La funzione obbiettivo viene definita come: $\sum_{i=1}^{K}\sum_{j=1}^{N}\left(\left|x_j - C_k\right|\right)^2$

Algoritmo K-means::dettaglio dei passi

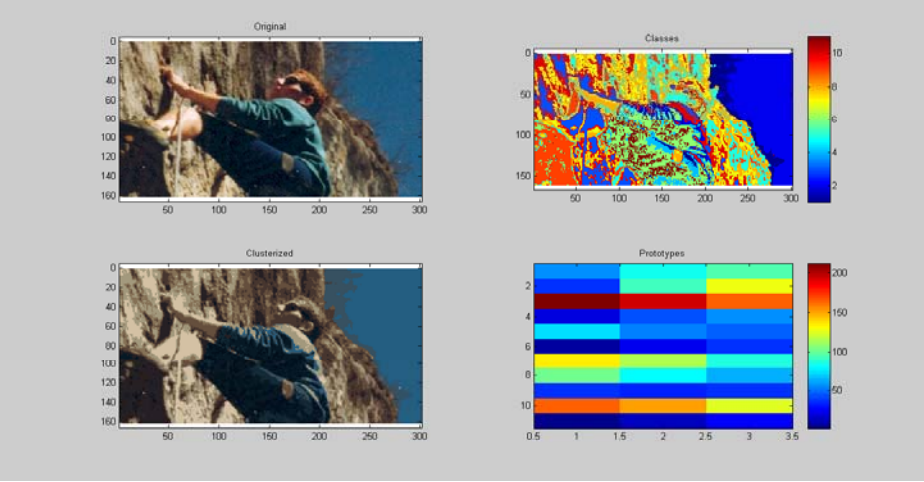
- Inizializzazione.
  - Posiziono in modo arbitrario o guidato i K centri dei cluster.
- Iterazioni
  - Assegno ciascun pattern al cluster il cui centro è più vicino, formando così un certo numero di cluster ($\leq$ K).
  - Calcolo la posizione dei cluster, C_k, come baricentro dei pattern assegnati ad ogni cluster, sposto quindi la posizione dei centri dei cluster.
- Condizione di uscita
  - I centri dei cluster non si spostano più.

K-means::limiti

- Partitional, hard, deterministic;
- Veloce, semplice da implementare;
- Trova un minimo locale della funzione $f=\Sigma_j\Sigma_i[dist(x_i,prot_j)] / N_j$;
- Il risultato dipende dall'inizializzazione!
- Possono essere usati altri metodi (es. GA) per inizializzare K-means… es. GA per la minimizzazione di f, effettuano una ricerca globale, ma sono lenti!

Sw in Matlab available

K-Means per immagine RGB



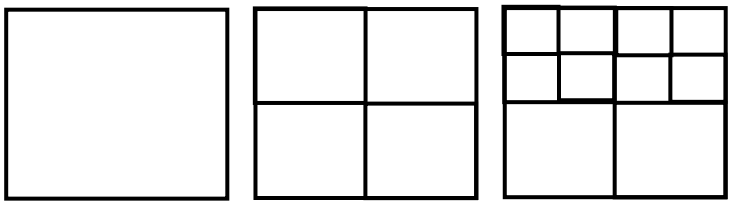
Da 255 colori a 33 colori

Riassunto

- I tipi di apprendimento
- Il clustering
- K means
- Quad-tree decomposition

Algoritmi gerarchici: QTD

- Quad Tree Decomposition;
- Suddivisione gerarchica dello spazio delle feature, mediante splitting dei cluster;
- Criterio di splitting (~distanza tra cluster).



Algoritmi gerarchici: QTD

- Clusterizzazione immagini RGB, 512x512;
- Pattern: pixel (x,y);
- Feature: canali R, G, B.
- Distanza tra due pattern (non euclidea):
  dist (p1, p2) =
  dist ([R1 G1 B1], [R2 G2 B2]) =
  max (|R1-R2|, |G1-G2|, |B1-B2|).

Algoritmi gerarchici: QTD

p1 = [0 100 250]
p2 = [50 100 200]
p3 = [255 150 50]

dist (p1, p2) = dist ([R1 G1 B1], [R2 G2 B2]) =
max (|R1-R2|, |G1-G2|, |B1-B2|) = max([50 0 50]) = 50.

dist (p2, p3) = 205.

dist (p3, p1) = 255.

Algoritmi gerarchici: QTD

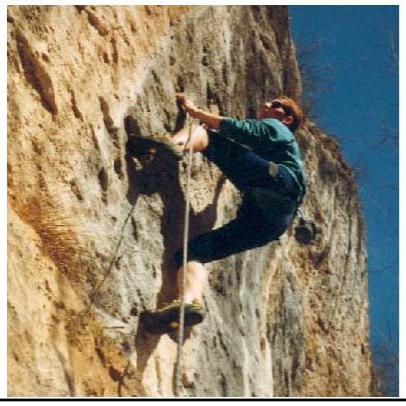
Criterio di splitting: se due pixel all'interno dello stesso cluster distano più di una determinata soglia, il cluster viene diviso in 4 cluster.

Esempio applicazione: segmentazione immagini, compressione immagini, analisi locale frequenze immagini…

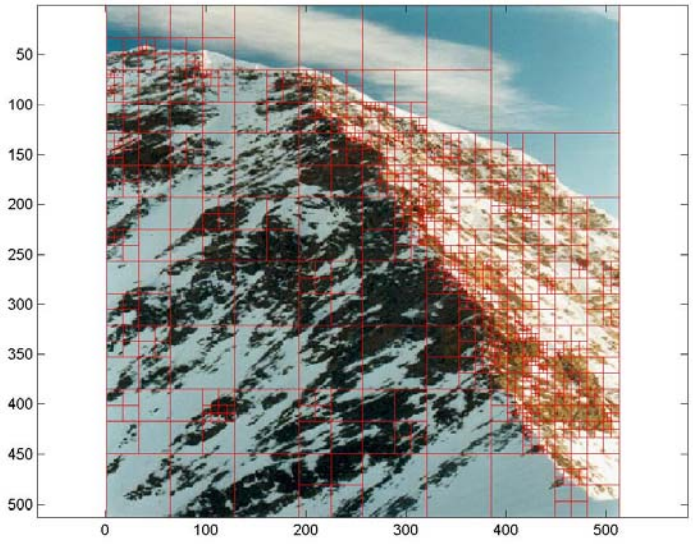
QTD: Risultati

Original

Clusterized



QTD: Risultati

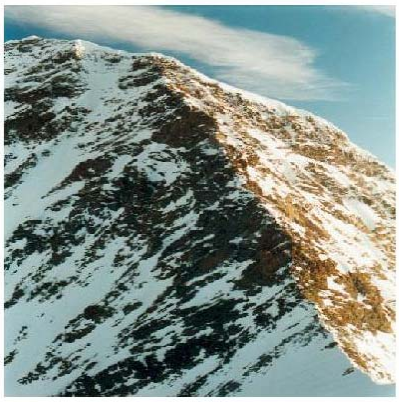


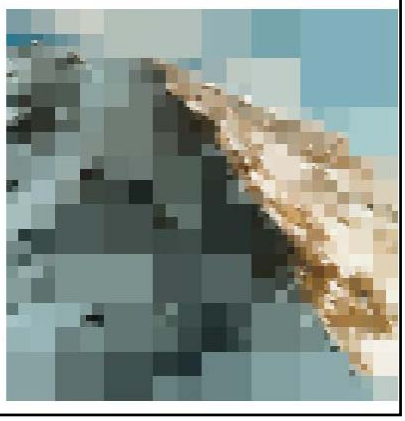
A.A

QTD: Risultati

Original

Clusterized





Apprendimento non-supervisionato: Clustering

Descrizione numerica dell'oggetto:
altezza, colore, forma, posizione, ….

SPAZIO DEI CAMPIONI / DELLE FEATURES

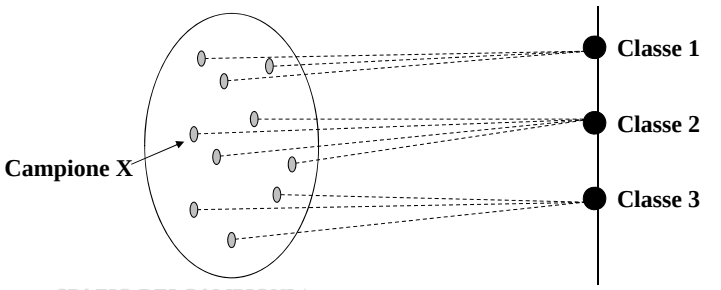
Classificatore

Classi omogene dell'oggetto:
(classe A, classe B, …)

SPAZIO DELLE CLASSI

Classificazione e clustering

Un'interpretazione geometrica:
Mappatura dello spazio dei campioni nello spazio delle classi.



SPAZIO DEI CAMPIONI / DELLE FEATURES (CARATTERISTICHE)

SPAZIO DELLE CLASSI

Che differenza c'è rispetto al clustering?

Cos'è un concetto?

Riassunto

- I tipi di apprendimento
- Il clustering
- K means
- Quad-tree decomposition