

## SOM (feature mapping)



Iuri Frosio  
frosio@dsi.unimi.it  
N. Alberto Borghese  
borghese@dsi.unimi.it



A.A. 2011-2012

1/49



## Sommario



- Clustering a feature mapping
- SOM

A.A. 2011-2012

2/49



## Clustering hard



Inizializzazione  $C_j$

Classificazione  $X_i$

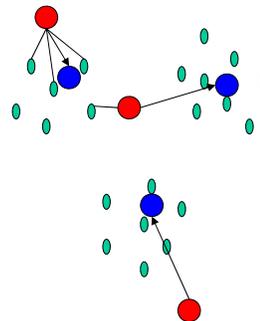
Aggiornamento  $C_j$

I prototipi  $C_j$  si sono spostati significativamente?

NO

Termine addestramento

SI



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

A.A. 2011-2012

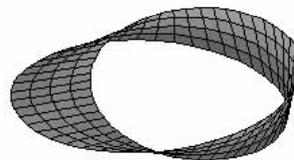
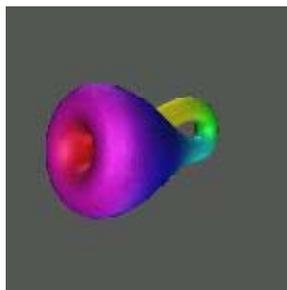
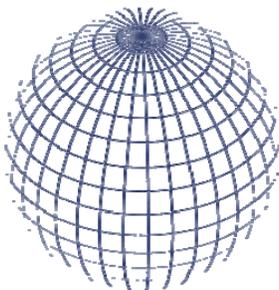
3/49



## Topology



- **Topology** (from the [Greek](#) τόπος, “place”, and λόγος, “study”) is a major area of [mathematics](#) concerned with properties that are preserved under [continuous](#) deformations of objects, such as deformations that involve stretching, but no tearing or gluing. It emerged through the development of concepts from [geometry](#) and [set theory](#), such as space, dimension, and transformation (Wikipedia)
- **Geometric topology** is the study of [manifolds](#) and maps between them, particularly [embeddings](#) of one manifold into another. (Wikipedia).



A.A. 2011-2012

4/49



## Clustering / Classificazione



- **Clustering** => Siano dati  $N$  dati, un algoritmo di clustering divide i dati in  $M$  classi diverse (senza necessità di alcuna supervisione). Non supervisionato.
- **Classificazione** => Sia  $X$  un dato, la classificazione consta nell'associare il dato ad una classe tra le  $M$  classi disponibili. Apprendimento supervisionato su un insieme di esempi.



## Addestramento supervisionato / non supervisionato.



- **Addestramento supervisionato** (in presenza di supervisore):
  - i dati di addestramento sono già associati alle loro classi di appartenenza. Apprendimento che generalizzi bene a nuovi dati.
- **Addestramento non supervisionato** (non è necessaria la presenza del supervisore):
  - *L'algoritmo* di addestramento effettua in maniera automatica una partizione dei dati di addestramento.

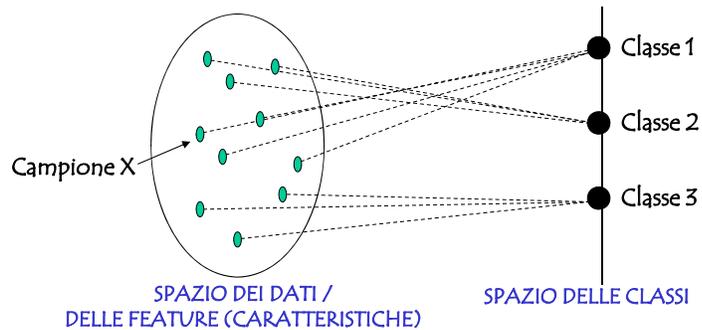
IL SUPERVISORE PUO' EFFETTUARE CONTROLLI A POSTERIORI PER VERIFICARE CHE LA PARTIZIONE OTTENUTA SIA CORRETTA. Occorre identificare il numero di partizioni.....



## Classificazione



Un'interpretazione geometrica:  
*Mappatura dello spazio dei dati (o feature) nello spazio delle classi.*



A.A. 2011-2012

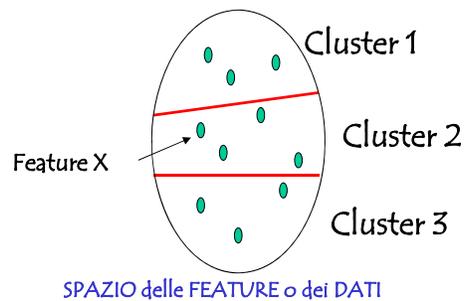
7/49



## Clustering



Interpretazione geometrica: *Partizione dello spazio dei dati in base alle loro features*



A.A. 2011-2012

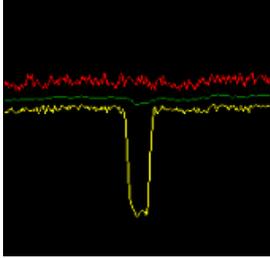
8/49



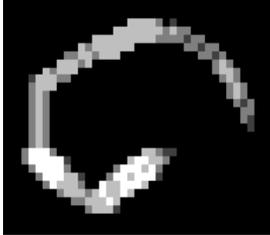
## Features



Macchie  
dense

Fili

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

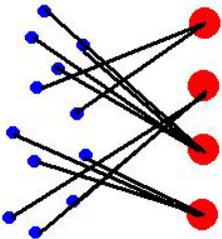
A.A. 2011-2012

9/49



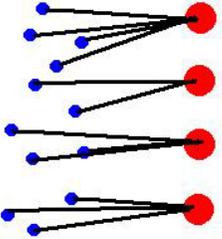
## Clustering vs. feature mapping





↑

Clustering



↑

Feature Mapping

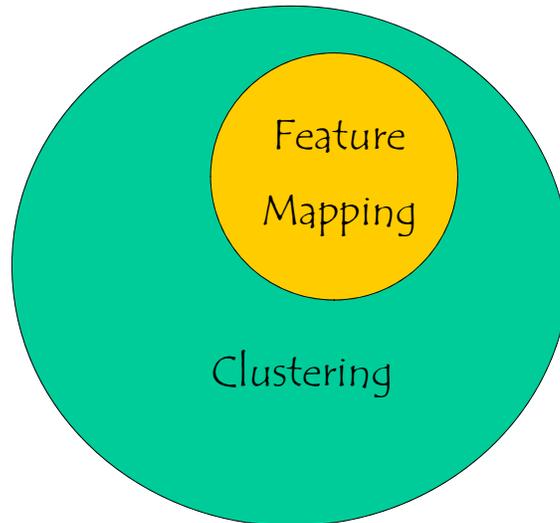
A differenza di quanto accade con il clustering, nel feature mapping vengono preservate le relazioni tologiche tra i dati.

A.A. 2011-2012

10/49



## Clustering e feature mapping



A.A. 2011-2012

11/49



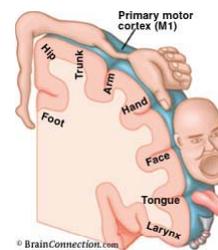
## Feature mapping



Con il Feature Mapping si dà importanza alla posizione dei prototipi (xxx-topia, Homunculus motorio);

A uscite contigue corrispondono configurazioni d'ingresso contigue;

Il clustering opera una trasformazione tra lo spazio degli ingressi e lo spazio delle uscite (prototipi) che preserva le relazioni di vicinanza tra i vari elementi.



A.A. 2011-2012

12/49



## Self Organizing Maps (SOM)



A.A. 2011-2012

13/49



## Tipo di apprendimento



**COMPETITIVE LEARNING.** *Apprendimento competitivo. Dato un certo input, le unità **compettono** tra loro per “aggiudicarsi” l’input.*

*Questo meccanismo può essere hard. Nel caso estremo: “winner-take-all”, “spara” un solo neurone per volta (grandmother cell). Oppure può essere soft, le unità raggiungono un grado diverso di “vincita”.*

*(In molte reti neurali, es. perceptrone, viene invece fornita un’uscita desiderata della rete per ogni ingresso di esempio...)*

A.A. 2011-2012

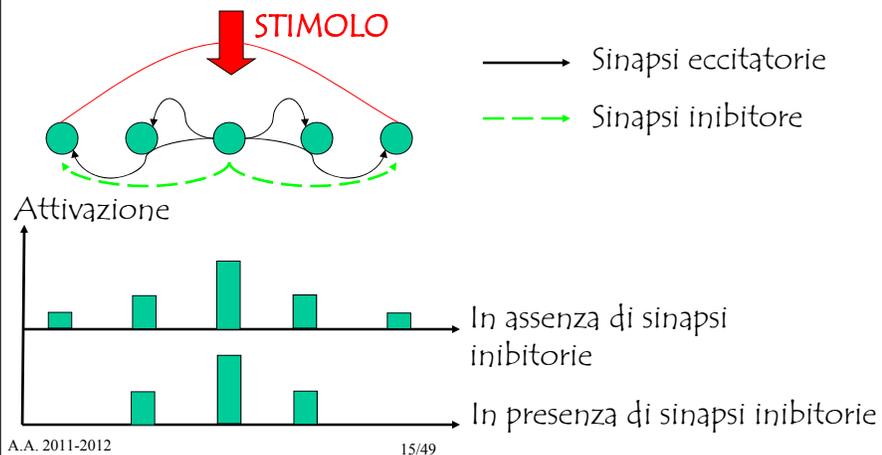
14/49



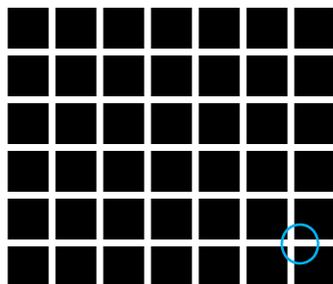
## L'ispirazione biologica : Il campo recettivo



- Sinapsi eccitatorie verso i neuroni vicini;
- sinapsi inibitorie verso i neuroni lontani.



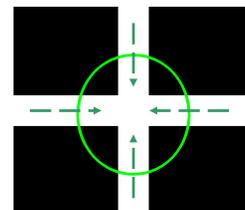
## Il campo recettivo: effetto Hermann



Pallini neri agli incroci delle linee bianche.

Il neurone centrale viene "spento" dai neuroni vicini attivi → generazione della zona scura.

- - - - - Inibitoria



A.A. 2011-2012

16/49



## Sommario



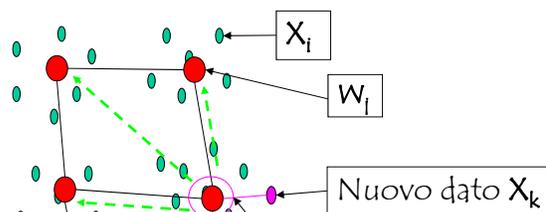
- Clustering a feature mapping
- SOM

A.A. 2011-2012

17/49



## SOM: competitive learning in versione "hard"



Connessione inibitoria  
"spegne gli altri neuroni"

Associazione di  $X_k$  alla classe  $W_l$ ; le sinapsi inibitorie spengono tutti gli altri neuroni (*winner takes all*). Il solo neurone che "spara",  $W_l$ , individua la classe di  $X_k$ .

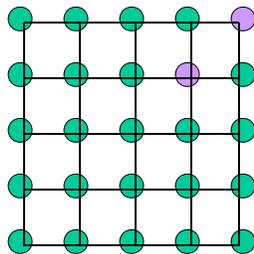
A.A. 2011-2012 Mapping  $R^N \rightarrow R^M$ , con  $M \leq N$  (da spazio delle caratteristiche a spazio delle classi).  
18/49



## SOM: organizzazione topologica



I neuroni della SOM sono ordinati topologicamente nello spazio dei neuroni (es. griglia ordinata in  $R^2$ ). In tale spazio viene definita la distanza tra neuroni.



La distanza tra i due neuroni (nello spazio dei neuroni) è:

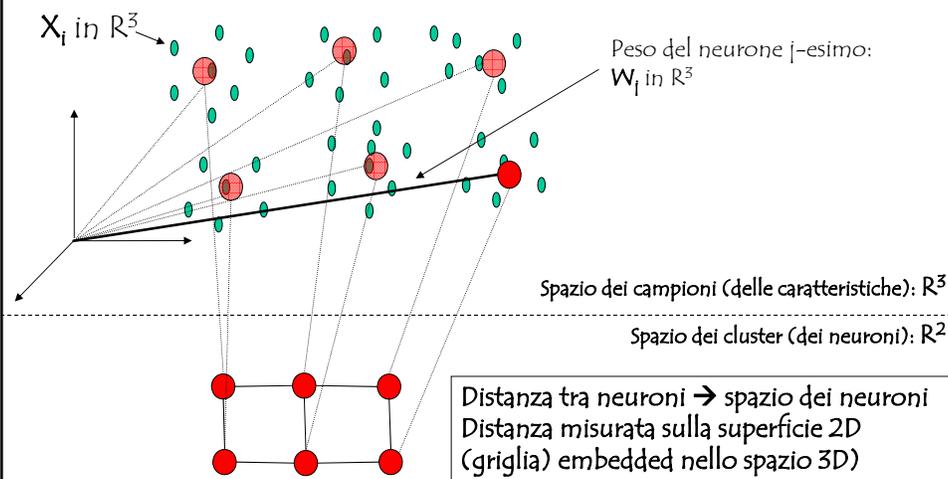
$$\sqrt{(\Delta x^2 + \Delta y^2)} = 1.4142 \quad \text{[Metrica Euclidea]}$$

$$|\Delta x| + |\Delta y| = 2 \quad \text{[Manhattan]}$$

...



## SOM: pesi dei neuroni





## SOM: addestramento



- Siano  $\mathbf{X}_1, \dots, \mathbf{X}_D$  i dati di addestramento (per semplicità, definiti in  $\mathbb{R}^3$ );
- siano  $\mathbf{W}_1, \dots, \mathbf{W}_K$  i *prototipi* di  $K$  classi, definiti anch'essi in  $\mathbb{R}^3$ ; ogni *prototipo* identifica il peso di un neurone della SOM ( $\mathbf{W}_j$ );
- lo schema di associazione adottato sia il seguente: “ $\mathbf{X}_i$  appartiene a  $\mathbf{W}_j$  se e solo se  $\mathbf{W}_j$  è il *prototipo* (*peso del neurone*) più vicino a  $\mathbf{X}_i$ , nello spazio dei campioni (delle caratteristiche,  $\mathbb{R}^3$ )”;
- l’algoritmo di addestramento permette di determinare i pesi dei neuroni (le posizioni dei *prototipi*)  $\mathbf{W}_j$  mediante successive approssimazioni;
- L’algoritmo di addestramento tiene conto della topologia dei neuroni nello spazio dei neuroni (*feature mapping*).

A.A. 2011-2012

21/49



## SOM: output



- All’interazione  $k$ -esima, si presenti alla rete il dato  $\mathbf{X}_i$ ;
- Unità vincente (associazione):  
$$j^* \text{ t.c. } \|\mathbf{W}_{j^*} - \mathbf{X}_i\| = \min_j \|\mathbf{W}_j - \mathbf{X}_i\|$$
- Uscita:
  - $u_{j^*} = 1$ , se  $j = j^*$
  - $u_j = 0$  se  $j \neq j^*$

UNITA' VINCENTE

A.A. 2011-2012

22/49



## Competitive learning



- All'interazione k- esima, si presenti alla rete il dato  $\mathbf{X}_i$ ;
  - **Aggiornamento di tutti i prototipi  $\mathbf{W}_j$ ;**
  - Generalized competitive Learning Rule:
    - $\Delta \mathbf{W}_j = \eta_k \Lambda_k(i,j) (\mathbf{X}_i - \mathbf{W}_j)$  ← AGGIORNAMENTO PESI DEI NEURONI
- $\Lambda_k(i,j)$  è una funzione “campo recettivo”
- $\Lambda_k(i,j) = \exp(-\|f(\mathbf{X}_i) - f(\mathbf{W}_j)\|^2 / 2\sigma_k^2)$  ← FUNZIONE DI VICINATO



## Soft-clustering



$$\Delta \mathbf{W}_j = \eta_k \Lambda_k(i,j) (\mathbf{X}_i - \mathbf{W}_j)$$

$\Lambda_k(i,j)$  è l'elemento chiave. I “Campi recettivi” dei diversi neuroni sono parzialmente sovrapposti.

In “Competitive clustering”  $\Lambda_k(i,j)$  è una Gaussiana nello spazio dei dati e dei prototipi.

In “Neural-gas”  $\Lambda_k(i,j)$  è una ranking function nello spazio dei dati e dei prototipi.

In “Fuzzy c-means”  $\Lambda_k(i,j)$  è realizzata mediante membership function nello spazio dei dati e dei prototipi.



## SOM e competitive learning



- All'interazione k- esima, si presenti alla rete il dato  $\mathbf{X}_i$ ;

- Unità vincente (associazione):  
 $j^*$  t.c.  $\|\mathbf{W}_{j^*} - \mathbf{X}_i\| = \min_j \|\mathbf{W}_j - \mathbf{X}_i\|$

UNITA' VINCENTE

- Competitive Learning Rule (SOM, Kohonen '81):
  - $\Delta \mathbf{W}_j = \eta_k \Lambda_k(i,j) (\mathbf{X}_i - \mathbf{W}_j)$  ← AGGIORNAMENTO PESI DEI NEURONI
  - $\Lambda_k(i,j) = \exp(-\|\mathbf{W}_{j^*} - \mathbf{W}_j\|^2 / 2\sigma_k^2)$  ← FUNZIONE DI VICINATO  
 Calcolata nello spazio dei prototipi

A.A. 2011-2012

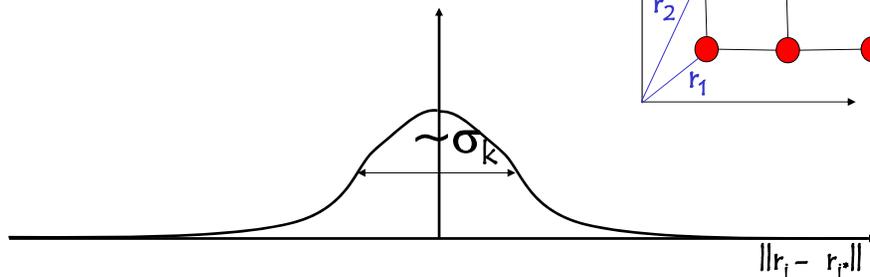
25/49



## Funzione di vicinato



$$\Lambda_k(i, j^*) = \exp(-\|r_i - r_{j^*}\|^2 / 2\sigma_k^2)$$



Neurone lontano dal neurone vincente:  $\Lambda_k(i, j^*) \rightarrow 0$ ,  $\Delta w_i \rightarrow 0$   
 $\sigma_k \rightarrow 0$ , l'esponenziale va rapidamente a 0, solo il neurone vincente,  $j^*$ , viene aggiornato.

$\sigma_k \rightarrow \infty$  l'esponenziale rimane costantemente = 1. aggiornare tutti i neuroni di una quantità pari a:  $\Delta \mathbf{W}_j = \eta_k (\mathbf{X}_i - \mathbf{W}_j)$

A.A. 2011-2012

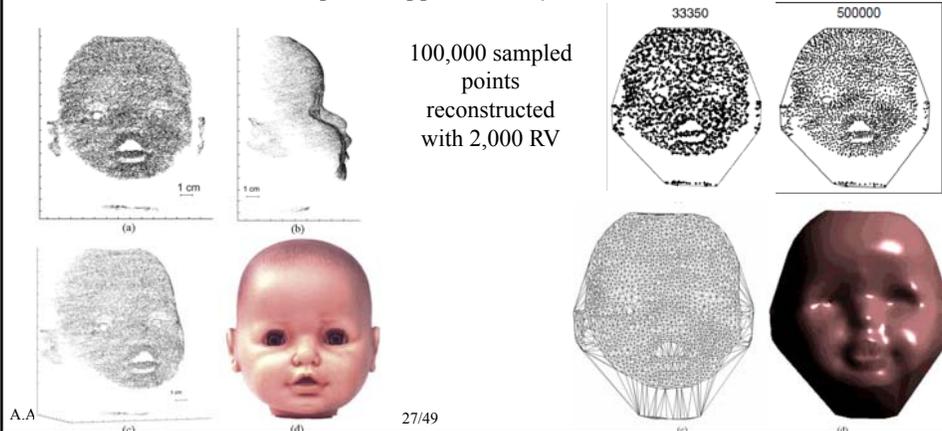
26/49



## Competitive learning ("First search then converge")



- 1) **ORDERING PHASE:**  $\eta, \sigma$  grandi; ogni neurone può spostarsi molto verso l'ingresso  $X_i$ ; il neurone trascina con sé i vicini; in tale fase la rete si dispiega nello spazio  $R^N$  "spargendo" i suoi neuroni.
- 2) **TUNING PHASE:**  $\eta, \sigma$  piccoli; ogni neurone si muove da solo; è una fase di raffinamento in cui vengono raggiunti con precisione i centri dei cluster.



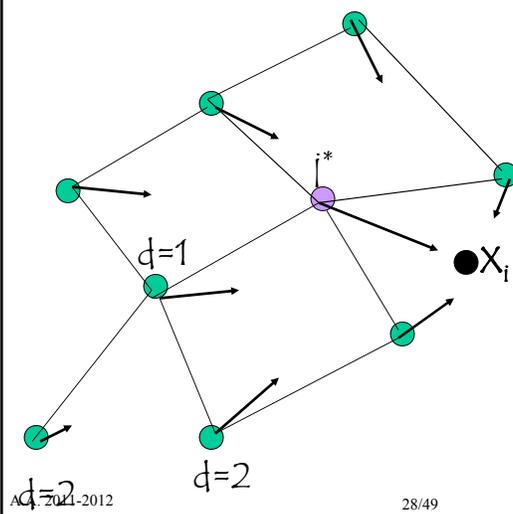
## SOM: addestramento



In definitiva:

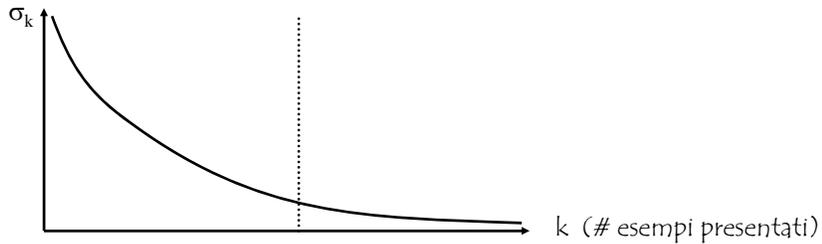
- Il neurone vincente si sposta verso  $X_i$ , trascinando i vicini in proporzione alla loro distanza nello spazio delle feature.

- L'ordinamento dei pesi dei neuroni nello spazio dei dati è simile all'ordinamento dei neuroni nello spazio dei neuroni.





## Funzione di vicinato nel tempo



$$\Delta w_i = \eta_k \Lambda_k(j, i^*) (X_i - w_i) \rightarrow w' = w + \eta_k \Lambda_k X_i - \eta_k \Lambda_k w$$

$$\Lambda_k(j, i^*) = \exp(-\|r_i - r_{i^*}\|^2 / 2\sigma_k^2)$$

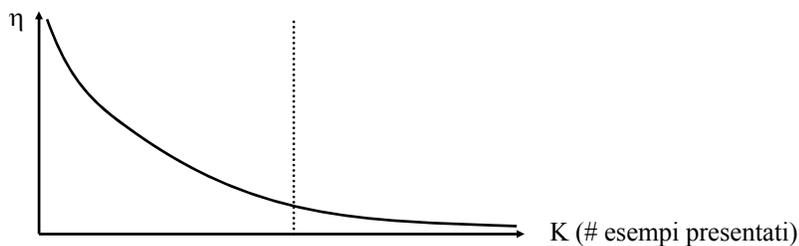
Procedendo nell'addestramento della rete, un neurone perde la capacità di spostare i suoi vicini.

A.A. 2011-2012

29/49



## Learning rate nel tempo



$$\Delta w_j = \eta_k \Lambda_k(j, j^*) (X_i - w_j) \rightarrow w' = w + \eta_k \Lambda_k X_i - \eta_k \Lambda_k w_j$$

Procedendo nell'addestramento della rete, i pesi dei neuroni perdono la possibilità di muoversi => rete più stabile.

A.A. 2011-2012

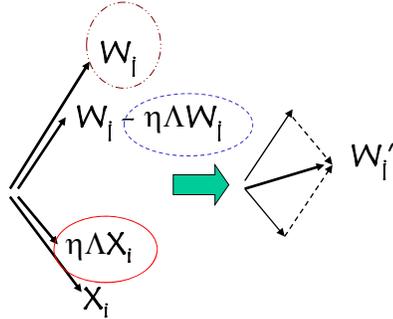
30/49



## Competitive Learning come forze contrapposte



$$\Delta W_i = \eta_k \Lambda_k(i,j) (X_i - W_i) \rightarrow W_i' = W_i + \eta_k \Lambda_k(i,j) X_i - \eta_k \Lambda_k(i,j) W_i$$



Il peso  $W_i$  del neurone vincente  $j$  si sposta verso l'ingresso presentato  $X_i$ .

$-\eta W_i$  evita che il peso  $W_i$  cresca a dismisura.

A.A. 2011-2012

31/49



## SOM: addestramento



- L'addestramento avviene presentando alla rete i vettori (dati)  $X_i \in \mathbb{R}^N$  per un numero di epoche  $E$ ;
- Per ogni esempio presentato  $X_i$  vengono aggiornati i pesi dei neuroni della rete;
- Durante l'addestramento il learning rate  $\eta$  e la neighborhood distance  $\sigma$  decrescono;
- Se presentiamo alla rete un nuovo esempio  $X_i$  alla fine dell'addestramento, la rete lo associa al neurone vincente (clustering);

In più:

- Categorie simili sono rappresentate da neuroni vicini (feature mapping).

A.A. 2011-2012

32/49



## SOM: problemi



- E' necessario scegliere  $\eta$ ,  $\sigma$ , numero di epoche, durata della ordering phase  $\rightarrow$  metodi empirici(!);
- Scelta della topologia e del numero di neuroni corretti;
- I dati di addestramento devono presentare una certa ridondanza;
- Unità "morte" (dead-units);

A.A. 2011-2012

33/49

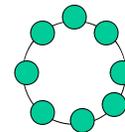


## SOM per ordinamento



Spazio dei dati  $X_i$  (e dei pesi  $w$ ):  $\mathbb{R}^3$

Topologia della SOM : circolare



Parametri di addestramento : # neuroni,  $\eta(t)$ ,  $\sigma(t)$ , ...

[Movie here](#) (with RealPlayer)

A.A. 2011-2012

34/49

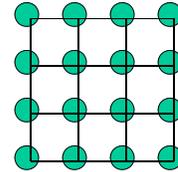


## SOM per ricostruzione 3D



Spazio dei dati  $\mathbf{q}$  (e dei pesi  $\mathbf{w}$ ) :  $\mathbb{R}^3$

Topologia della SOM : griglia 2D



Parametri di addestramento : # neuroni = 10x10, 0.5  $\rightarrow$  0.1 lin, 10  $\rightarrow$  1 lin, ...

[Movie here](#)

A.A. 2011-2012

35/49



## SOM per ricostruzione 3d



Problemi:

- Oscillazioni della rete all'inizio dell'addestramento;  
*Sol.: Scelta accurata di  $\eta(t)$ ,  $\sigma(t)$*
- Raggiungimento dei confini della superficie aperta;  
*Sol.: Boundary First Method +  $\eta$ ,  $\sigma$  modificati*
- Numero insufficiente di neuroni;  
*Sol.: Parametrizzazione della points cloud*

A.A. 2011-2012

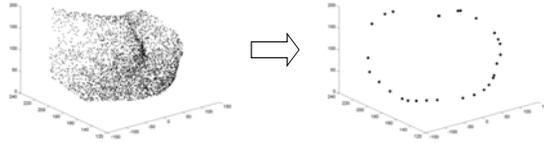
36/49



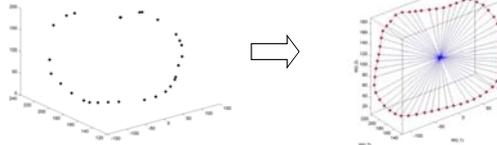
## Boundary First Method



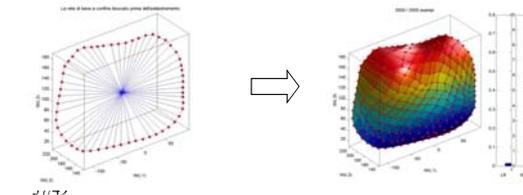
1) Individuazione dei punti di confine della superficie



2) Posizionamento dei neuroni di bordo



3) Addestramento della SOM (neuroni di bordo bloccati),  $\eta, \sigma$  maggiori ai lati maggiori ai lati



A.A. 2011-2012



## SOM per ricostruzione 3D



- [Movie here](#)
- [Movie here](#)

A.A. 2011-2012

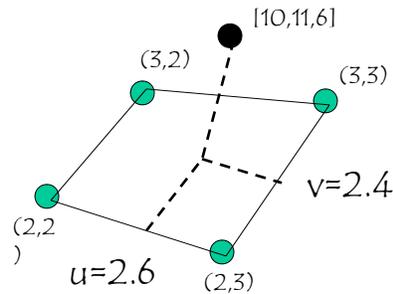
38/49



## Parametrizzazione Points Cloud



- Ad ogni punto 3D  $[x,y,z]$  vengono assegnate le coordinate 2D  $[u,v]$  corrispondenti nello spazio 2D della SOM tramite una proiezione (parametrizzazione)



A.A. 2011-2012

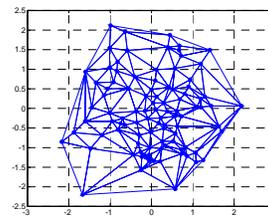
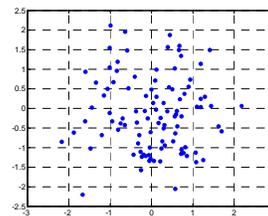
39/49



## Triangolazione punti in $[u,v]$ (Delaunay)



- Dato un set di punti nello spazio 2D, è possibile trovare un set di triangoli tale per cui (triangolazione di Delaunay):
  - ogni punto si trova al vertice di uno o più triangoli
  - nessun punto si trova all'interno di un triangolo
- Utile in computer graphics.
- In modo più rigoroso:  
a Delaunay triangulation for a set  $P$  of points in the plane is a triangulation  $DT(P)$  such that no point in  $P$  is inside the circumcircle of any triangle in  $DT(P)$ . Delaunay triangulations **maximize the minimum angle of all the angles of the triangles** in the triangulation; they tend to **avoid "sliver" triangles**.



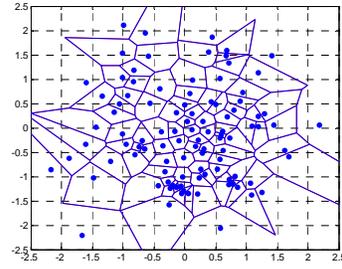
A.A. 2011-2012

40/49



## Delaunay e Voronoi

- Duale della triangolazione di Delaunay  $\rightarrow$  Tessellazione di Voronoi.
- Suddivisione “regolare” dello spazio.

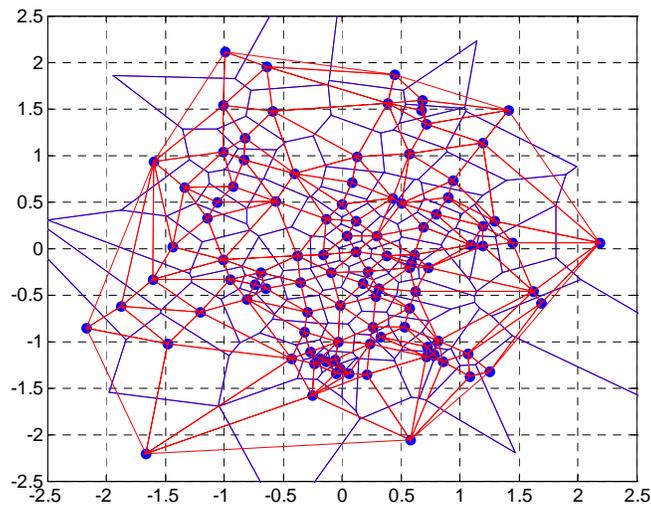


A.A. 2011-2012

41/49



## Dealunay & Voronoi



A.A. 2011-20



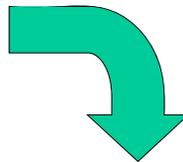
## Parametrizzazione Points Cloud



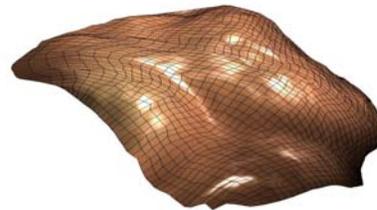
I punti 2D  $[u,v]$  possono essere triangolati (ricostruzione a mesh di triangoli). La mesh viene poi filtrata/interpolata.



Ricostruzione a mesh di triangoli



Filtraggio o interpolazione



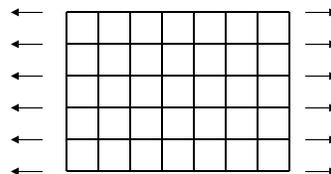
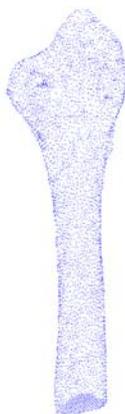
Ricostruzione Finale

A.A. 2011-2012

43/49



## Tibia Left



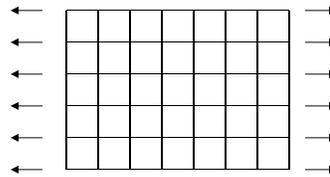
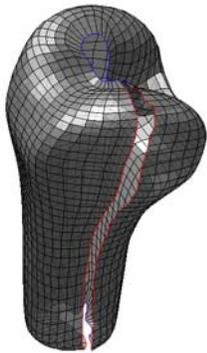
Topologia: cilindro

A.A. 2011-2012

44/49



## Femoral Bone Left



Topologia: cilindro

A.A. 2011-2012

45/49



## Varianti SOM



- SELF CREATING MAP
  - Aggiunta di un neurone:
    - Vicino al neurone vincente con  $f$  maggiore;
    - Vicino al neurone con curvatura massima;
- SOM SUPERVISIONATE (per la classificazione);
- ...

A.A. 2011-2012

46/49



## Software available



Clustering topologico dei caratteri Braille

SW available



## Bibliografia



- J. Hertz, A. Krogh, R. G. Palmer, “Introduction to the theory of neural computation”, Santa Fe Institute Editorial Board, 1990
  - Feature mapping & SOM: cap. 9.4



## Sommario



- Clustering a feature mapping
- SOM