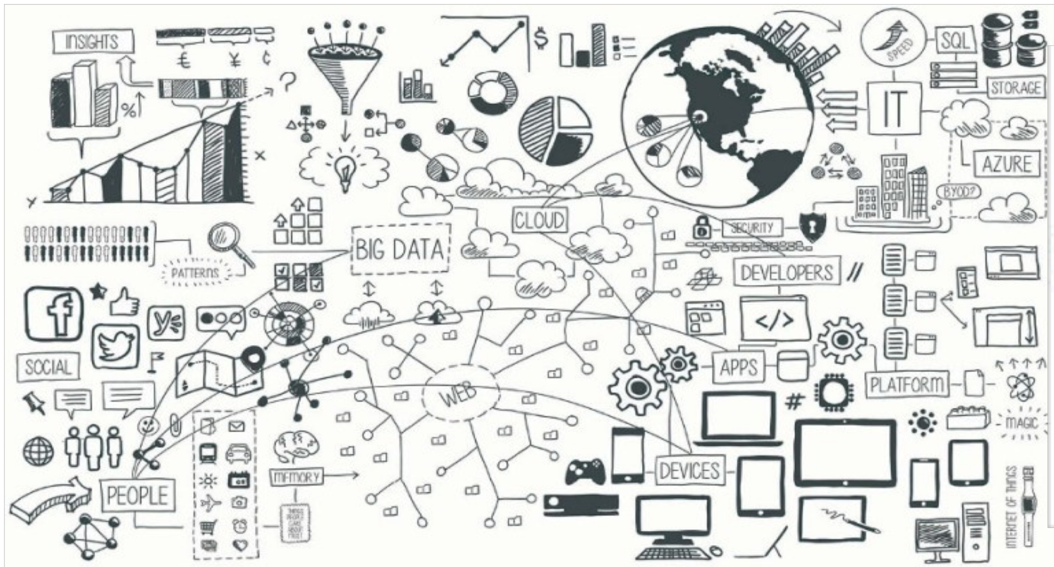
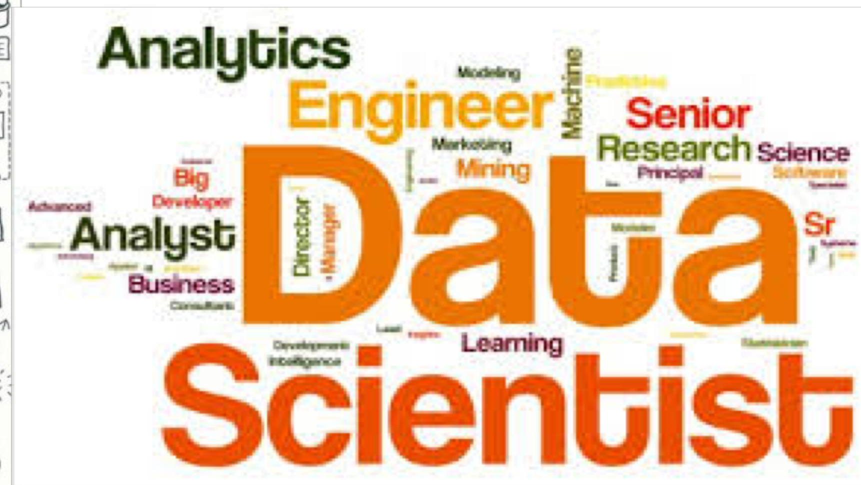
Machine Learning II (Neural Networks)

COMPETITIVE AND SELF-ORGANIZING NETWORKS



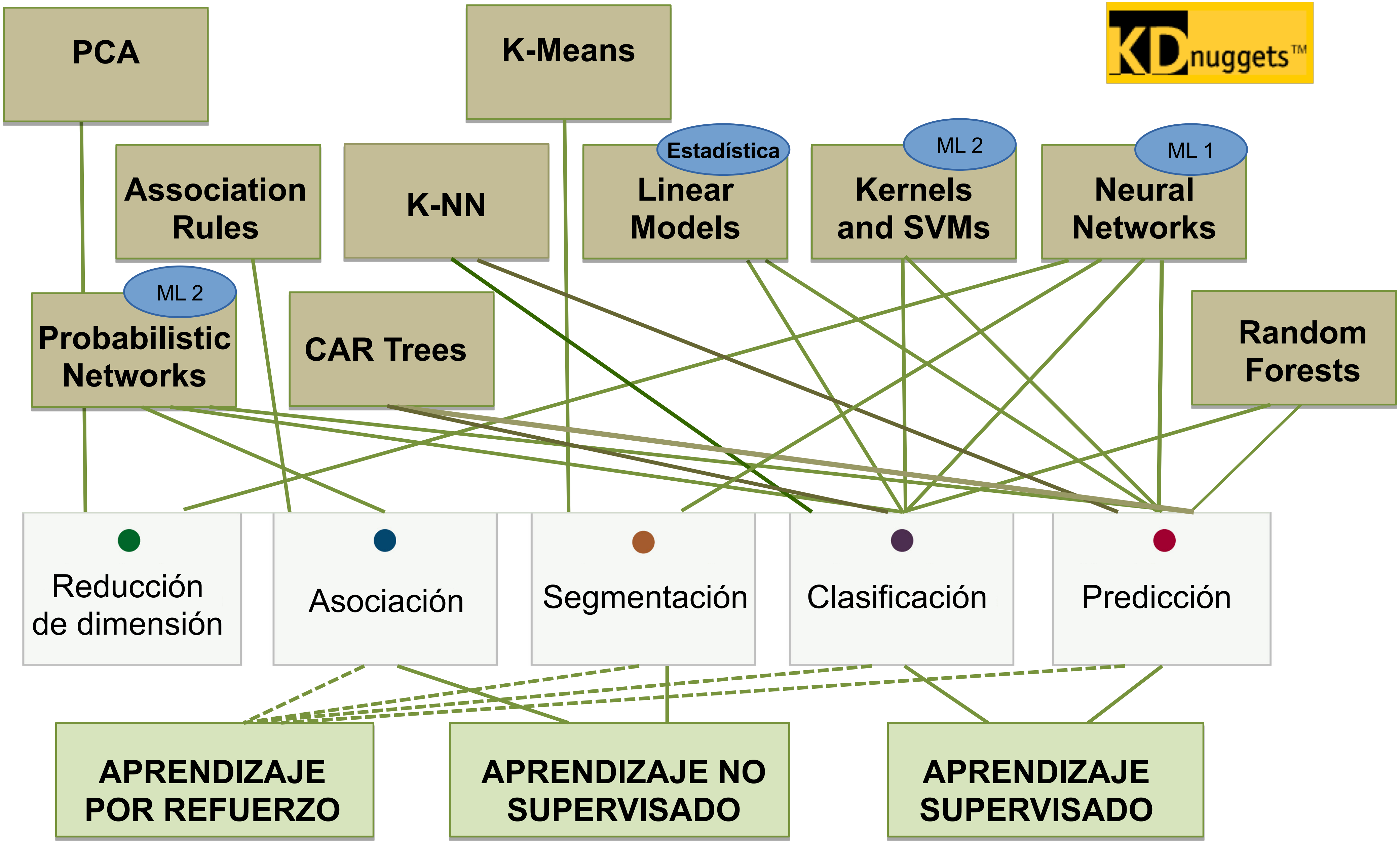


José Manuel Gutiérrez

Grupo de Meteorología

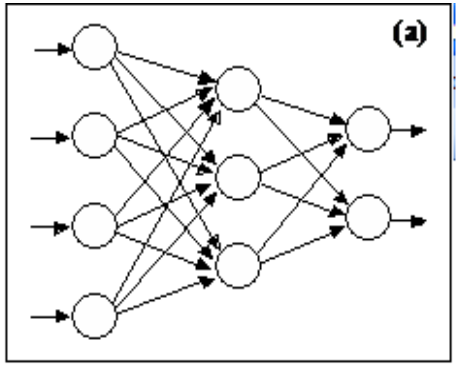
Univ. de Cantabria – CSIC
MACC / IFCA

https://www.kdnuggets.com/2017/10/top-10-machine-learning-algorithms-beginners.html



Supervised Problems. Input-Output pairs are provided:
(x1,y1), (x2,y2), ..., (xn,yn) and the network learns y = f(x+ε).

Multilayer* *Networks or Feedforward Nets.
Several layers connected (input+hidden+output)



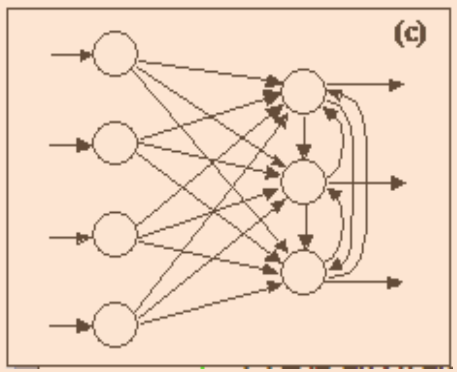
Pattern Recognition
OCR, images
Interpolation and fitting

Prediction: Input => Output

Learning: Backpropagation

Unsupervised Problems. Only input data is provided:
x1, x2, ..., xn and the network self-organizes it to provide a clustering.

Competitive* *Networks
Multilayer networks with lateral connections (competitive) in the last layer.



Segmentation, visualization

Feature extraction.

Prediction: Input => Clusters

Learning: Ad hoc
Winner-takes-all

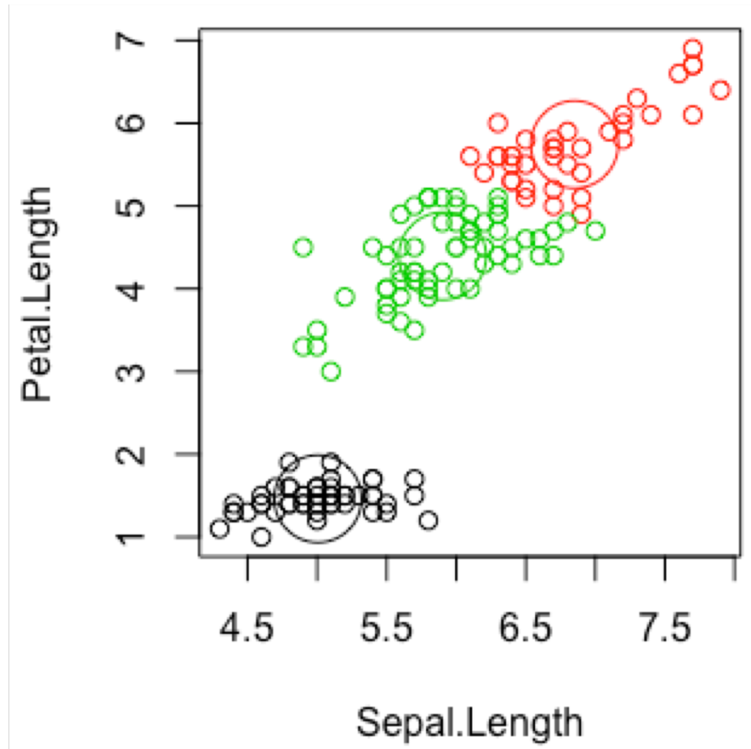
Problemas habituales

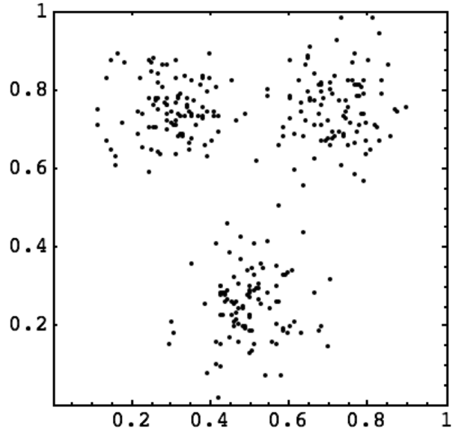
Reducción de dimensión | Asociación | **Segmentación** | Clasificación | Predicción

APRENDIZAJE POR REFUERZO | **APRENDIZAJE NO SUPERVISADO** | **APRENDIZAJE SUPERVISADO**

Centroid-based clustering:
Non-overlapping (e.g. K-Means)

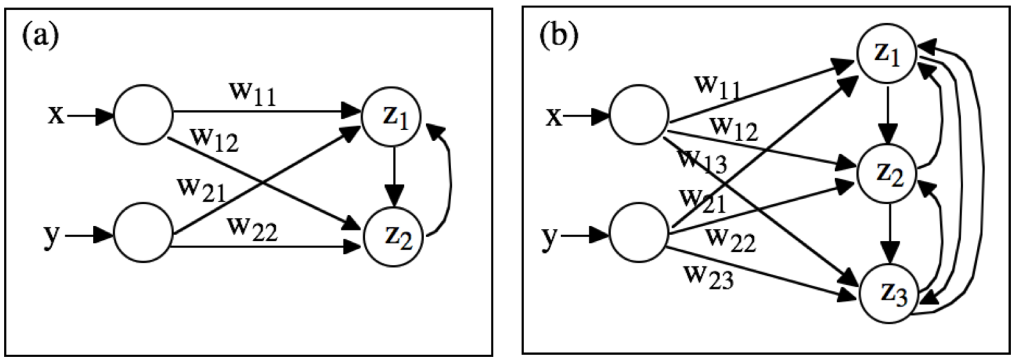
```
library(stats)
? kmeans
km<-kmeans(iris[,-5],3,nstart=1)
summary(km)
## Point center of two attributes
plot(iris[,c(1,3)],col=km$cluster)
points(km$centers[,c(1,3)],col=1:3,cex=5)
confusionMatrix(as.numeric(iris[,5]),
                     km$cluster)
```

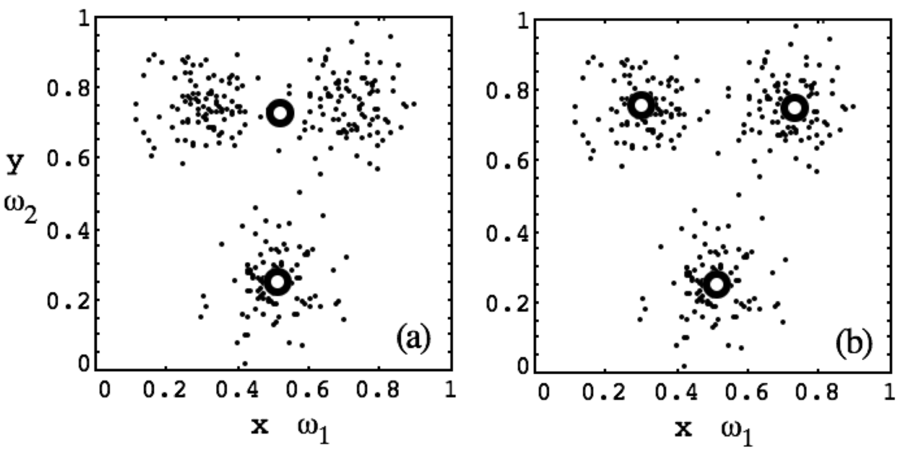


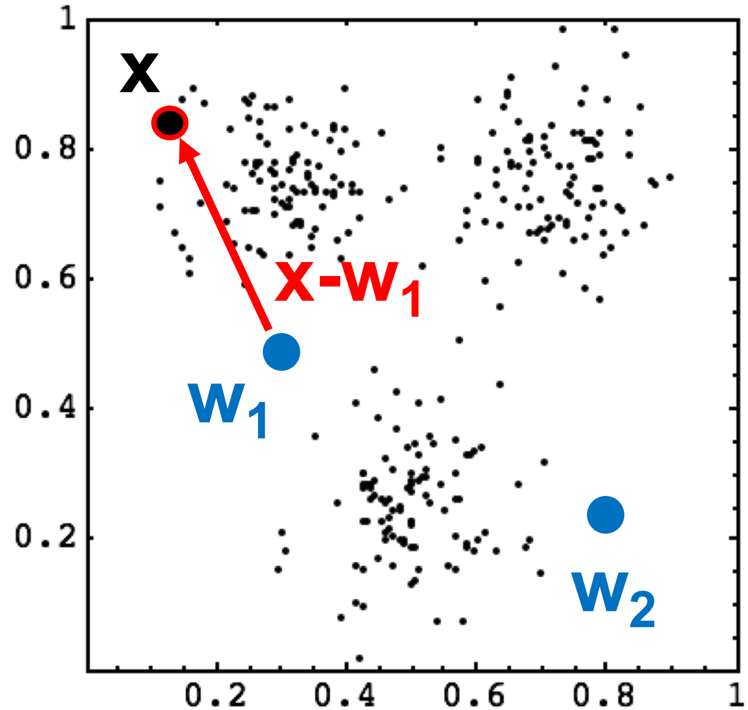


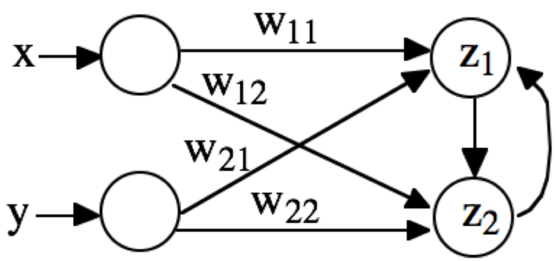
El algoritmo de aprendizaje (ad hoc) se basa en especializar los vectores peso para que se aproximen al centroide de una clase.

Winner-takes-all: sólo una de las neuronas de salida se activará para un patrón de entrada dado.









El algoritmo de aprendizaje (ad hoc) se basa en especializar los vectores peso para que se aproximen al centroide de una clase.

Winner-takes-all: sólo una de las neuronas de salida se activará para un patrón de entrada dado.

Algorithm:

1. (Set-up.) Select a number of outputs (clusters) and establish connections with inputs using random unit weights.

2. For each input pattern **x**, the node with the highest output is the winner:

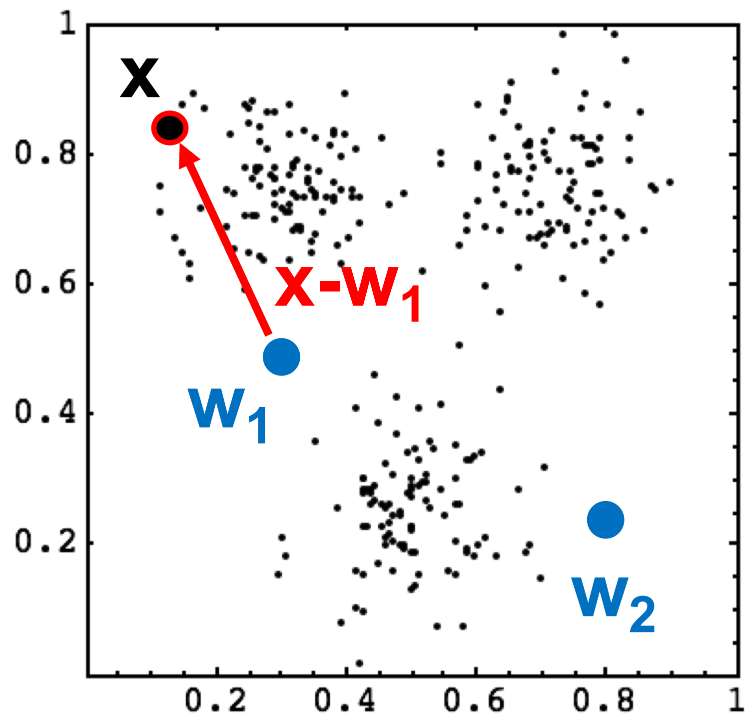
   $S(\mathbf{w},\mathbf{x}) = \mathbf{w}\,\mathbf{x}.$ Max. cosine similarity
   Alternativa: Euclidean distance

3. the winner $\mathbf{o}_i$ updates its weights:

   $$w_i\,(t+1) = w_i\,(t) + \eta\,(t)\,(x - w_i\,(t))$$

4. Repeat step 2 a number of epochs *n*.

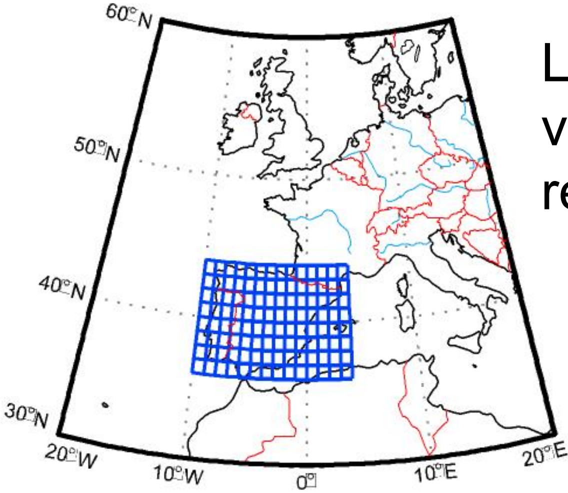
Each node (weight) converges to the center of the cluster.



```r
activation <- function(z) { 1/(1 + exp(-z)) }  #identidad (lineal)

backprop <- function(a,b, epochs = 100, eta = 0.1) {
   ### Inicializar matrices y listas
   RMSE <- matrix(data=NA, nrow=epochs, ncol=ncol(b));
   a <- cbind(a,rep(1,nrow(a)))
   neurons <- c(ncol(a),ncol(b))
   W <- matrix(data = runif(prod(neurons), min = -1, max = 1),
          nrow = neurons[2], ncol = neurons[1])
   # Hacer los pesos unitarios
   for (j in 1:epochs) {
      ### Propagar hacia delante
      bout <- t(activation(W %*% t(a)))  # calcular ganadoras
      ### Actualizar pesos
      aux <- (b - bout) * (bout * (1 - bout))
      Wdelta <- t(aux) %*% a
      W <- W + eta * Wdelta     #eta ➔ decrecer con epochs
      ### Error output
      RMSE[j,]=sqrt(mean((bout - b)**2))
   }
   ### Return values
   return(list("pred" = bout, "W" = W, "RMSE"=RMSE))
}
```

Programar

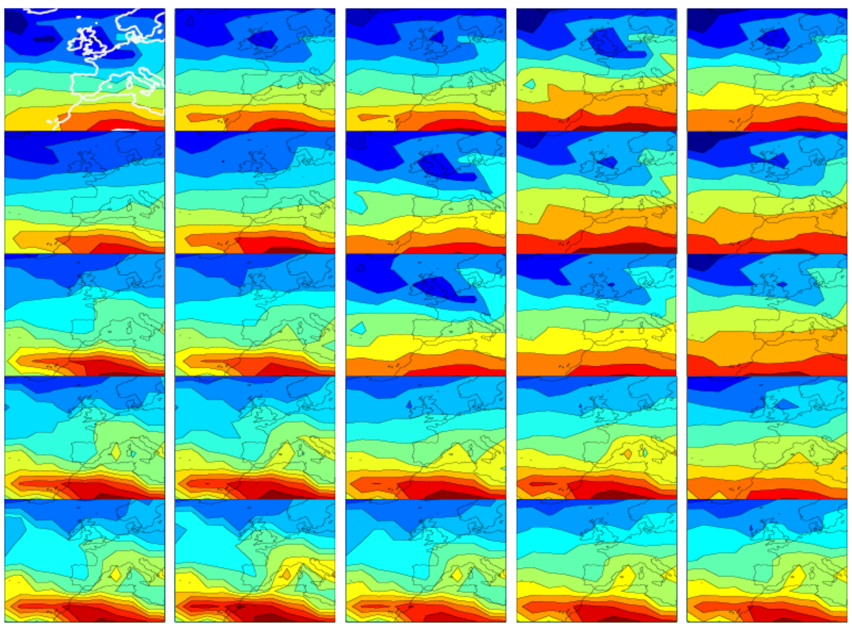
T. Kohonen

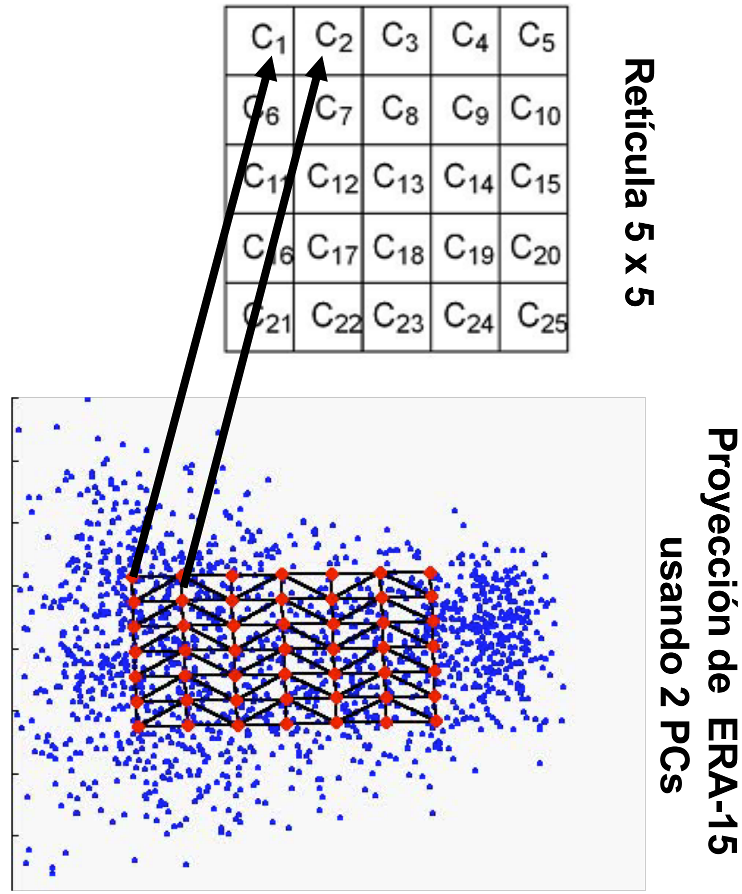


Las redes auto-organizativas SOM son una variante de red competitiva que incluyen una restricción topológica de vecindad en una retícula:

Patrón atmosférico de circulación:

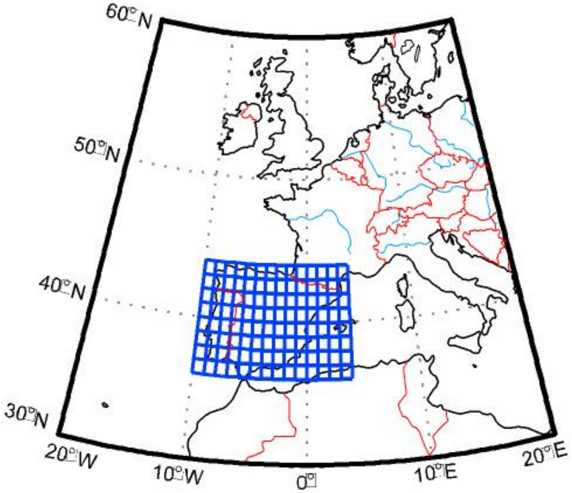
P=(**T**(1000mb), **T**(850mb),...,**Z**,**H**...)

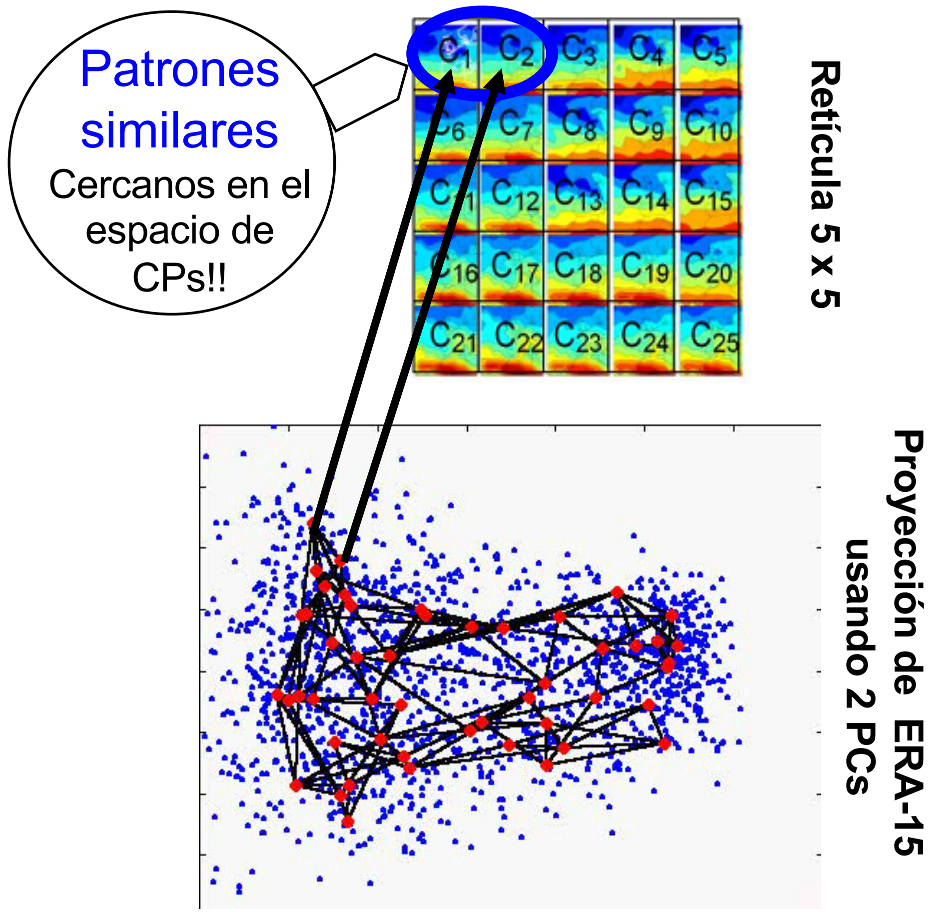




Los patrones se agrupan en una retícula 2D.

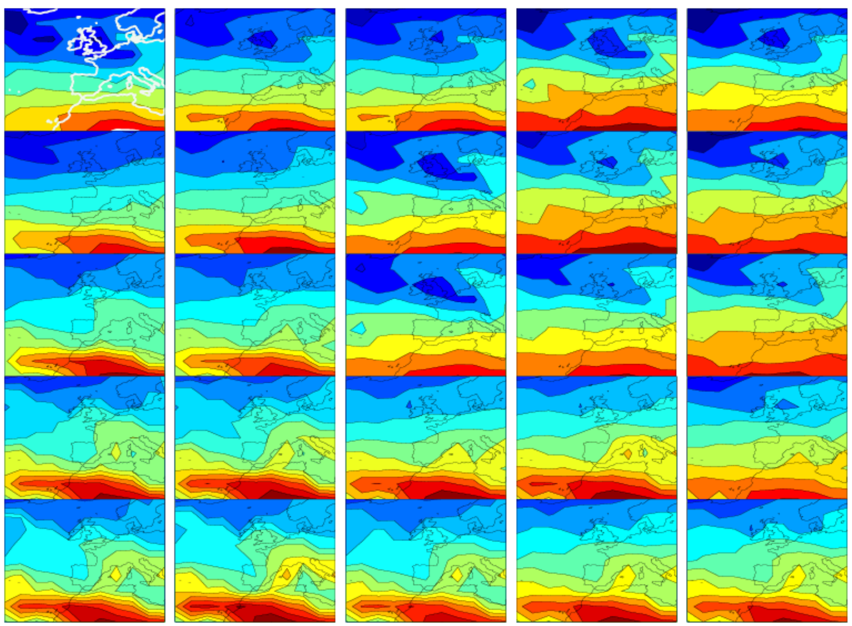
Las redes auto-organizativas SOM son una variante de red competitiva que incluyen una restricción topológica de vecindad en una retícula:





Patrón atmosférico de circulación:

P=(**T**(1000mb), **T**(850mb),...,**Z**,**H**...)

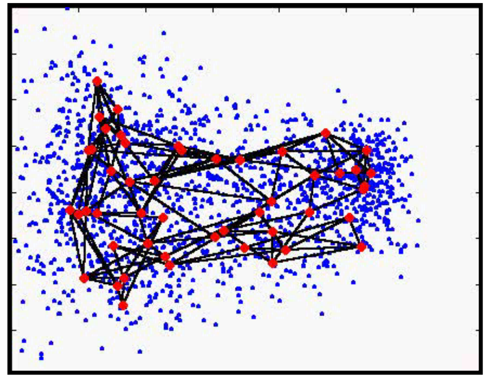


Los patrones se agrupan en una retícula 2D.

SOM 5 x 5

Una SOM está formada por un número arbitrario de (prototipos) centros $c_1,\ldots,c_m$, dispuestos en una retícula 2D.

Real space



Feature space

C_1	C_2	C_3	C_4	C_5
C_6	C_7	C_8	C_9	C_{10}
C_{11}	C_{12}	C_{13}	C_{14}	C_{15}
C_{16}	C_{17}	C_{18}	C_{19}	C_{20}
C_{21}	C_{22}	C_{23}	C_{24}	C_{25}

Cada prototipo $\mathbf{w}_i=(w_{i1}, \ldots, w_{in})$

n es la dimensión del espacio original.

El entrenamiento se realiza en ciclos **$(k=1,\ldots,n)$**:

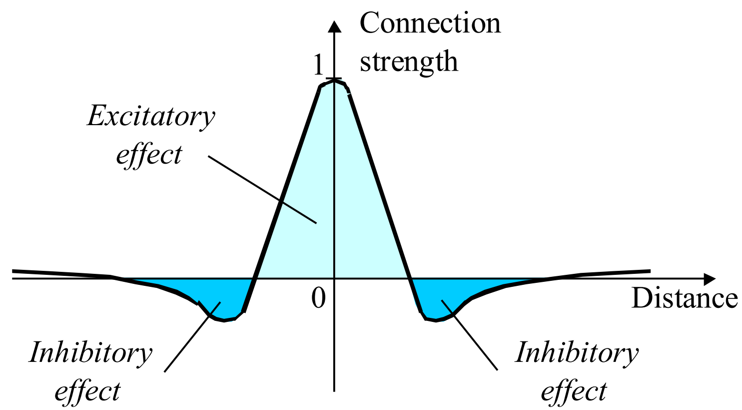
1) Calcular el prototipo "ganador" (más cercano) $\mathbf{w}_{i(k)}$ para cada patrón x_k:

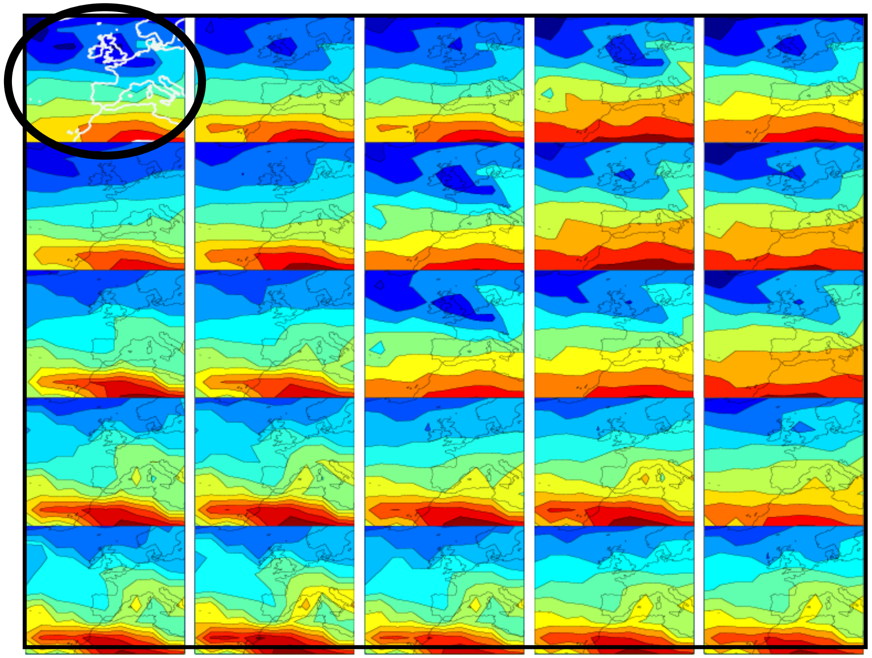
$$\| x_k - w_{i(k)} \| = \min_i\{\| x_k - w_i \|, i=1,\ldots,m\}.$$

2) El prototipo ganador y sus vecinos se mueven hacia el vector:

$$w_i(t+1)=w_i(t)+a(t)\,(\mathbf{x}_k - w_i(t))\,h(\|w_i(t) - w_{i(k)}(t)\|),$$

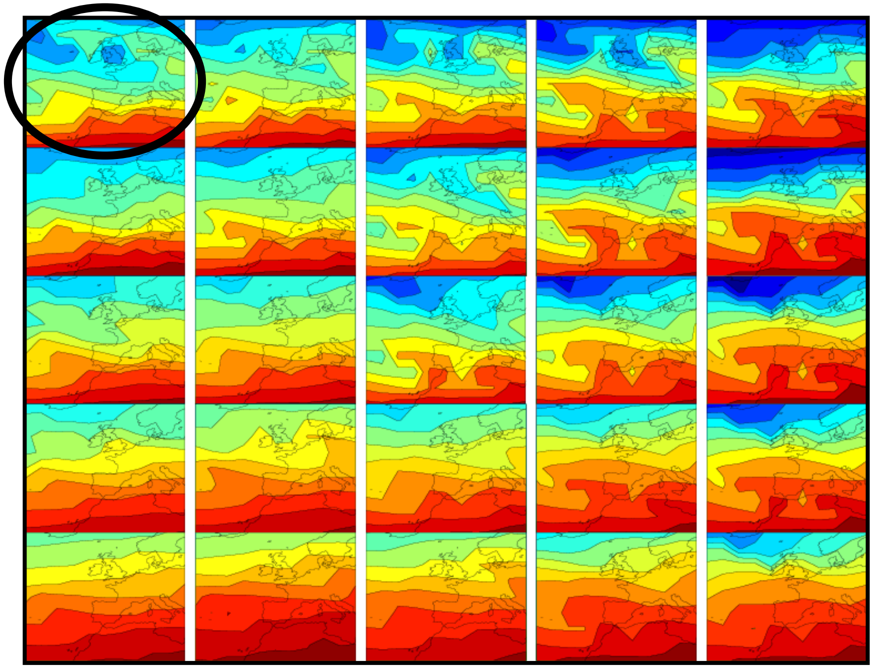
a(t) tasa de aprendizaje **(decrece linealmente);**
h(x) núcleo de vecindad **(la varianza decrece linealmente)**

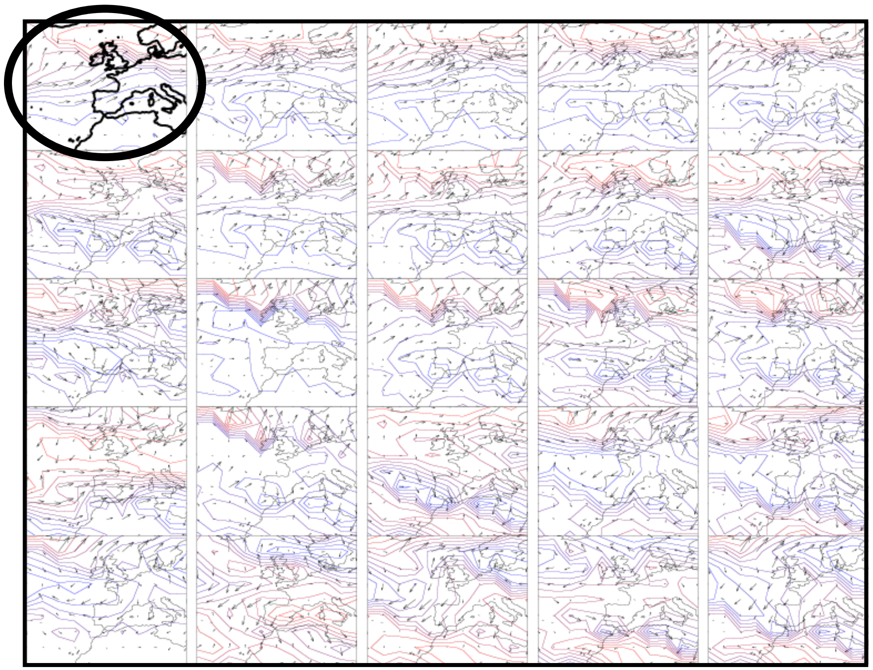


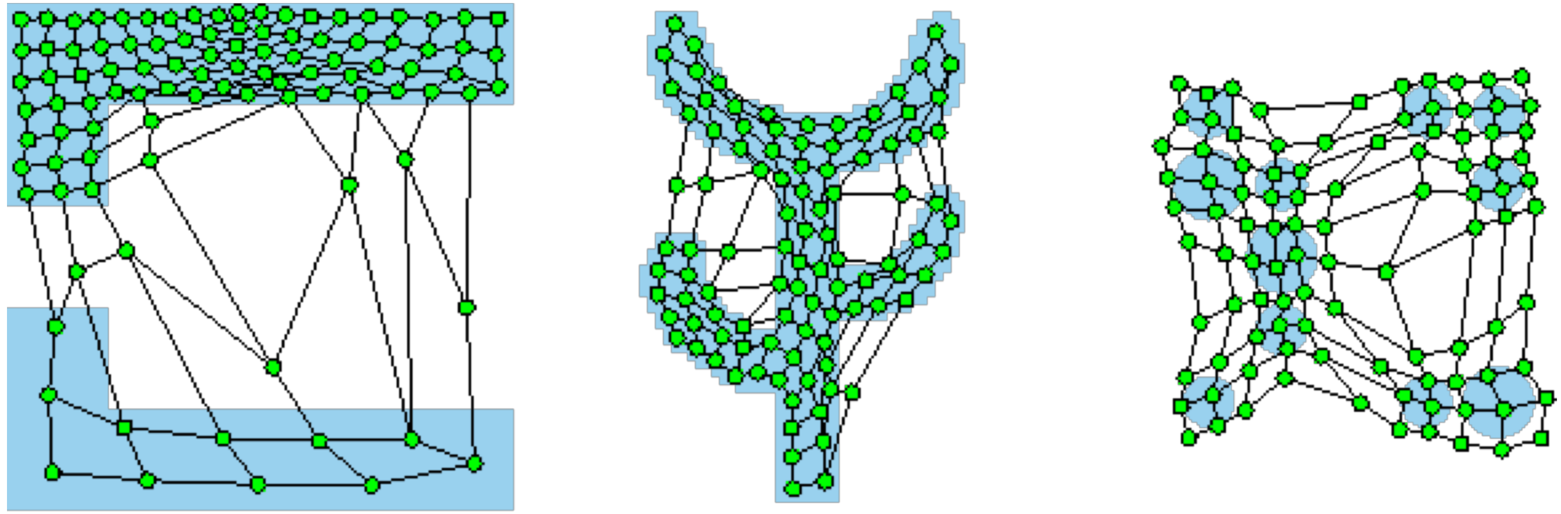


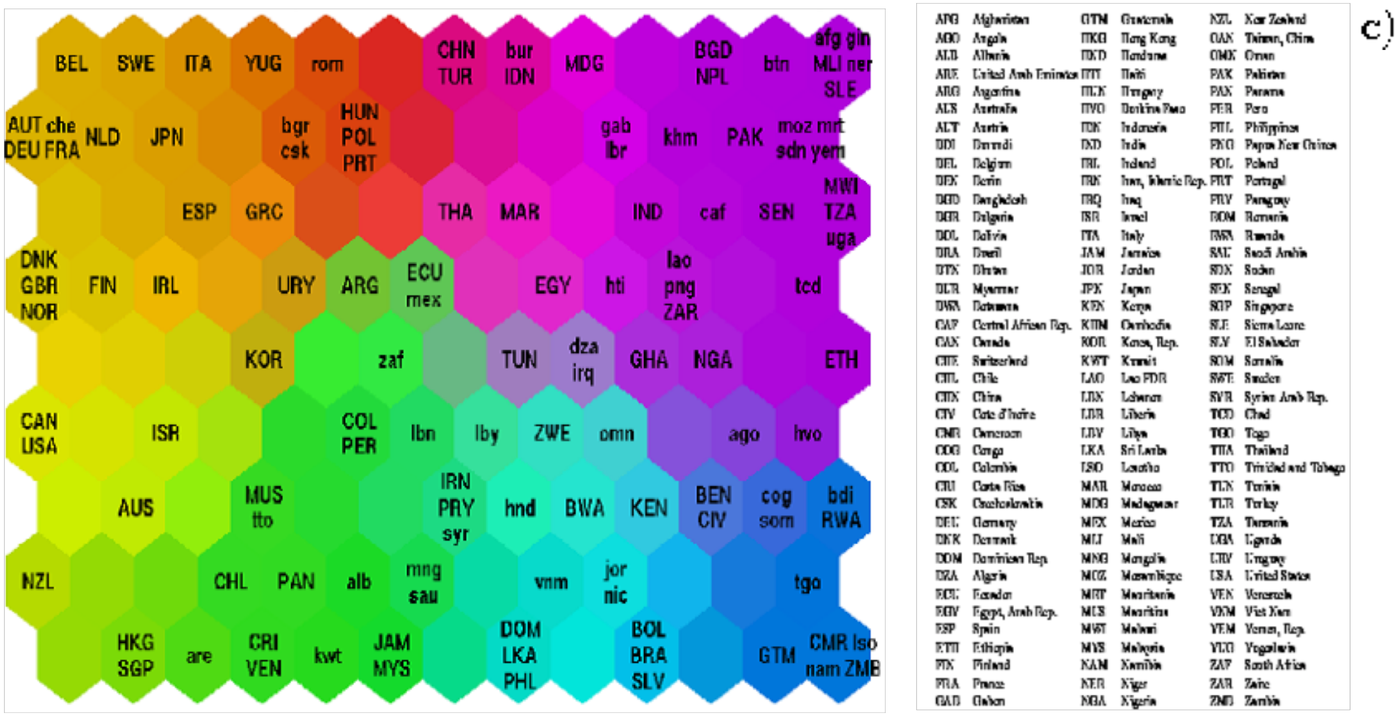
Prototypes for a trained SOM.
Close units in the lattice are associated with similar atmospheric patterns.

- **T 1000mb**
- **T 500mb**
- **Z, U, V 500mb**









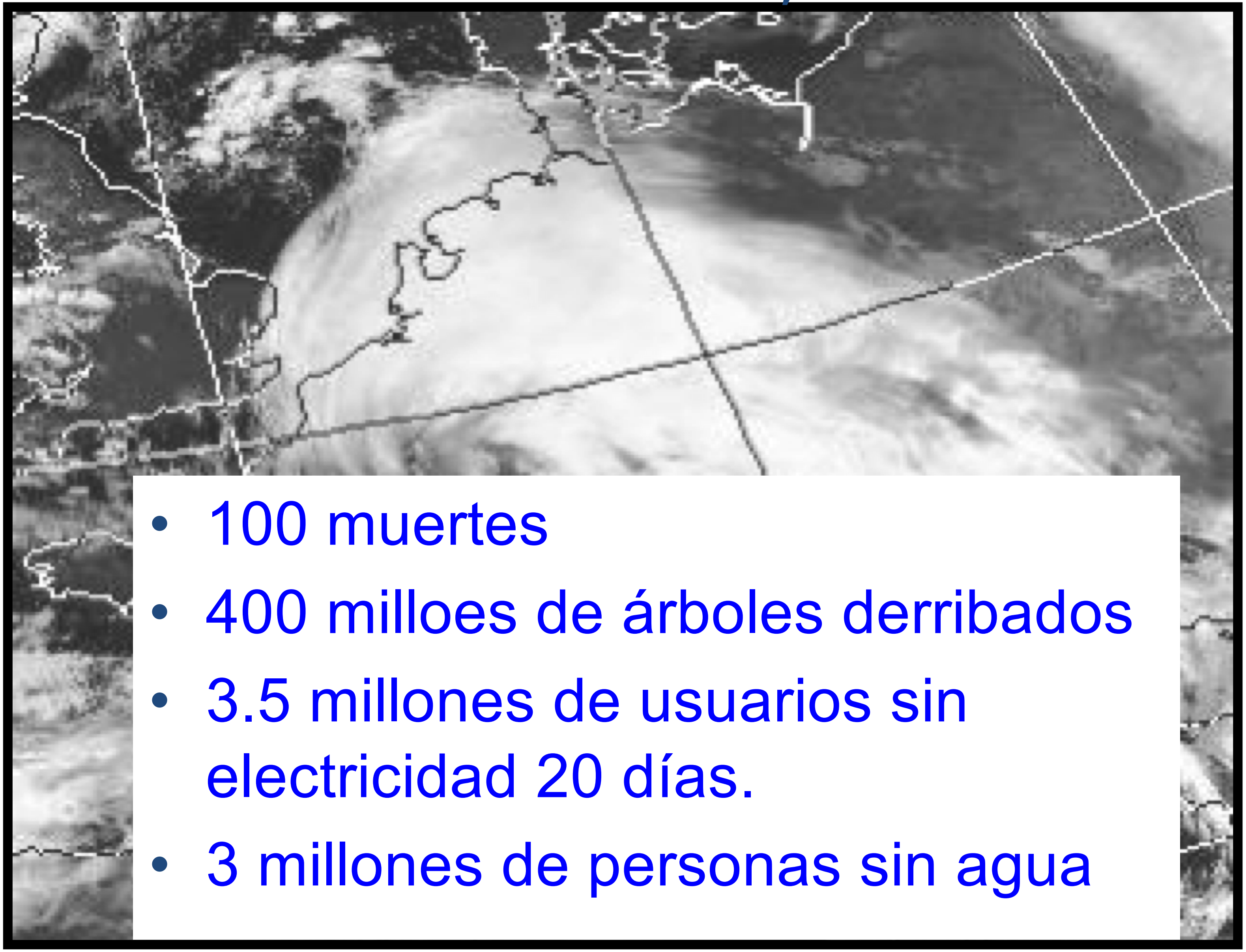
Pobreza
(34 indicadores)

http://blog.schochastics.net/post/dimensionality-reduction-methods/

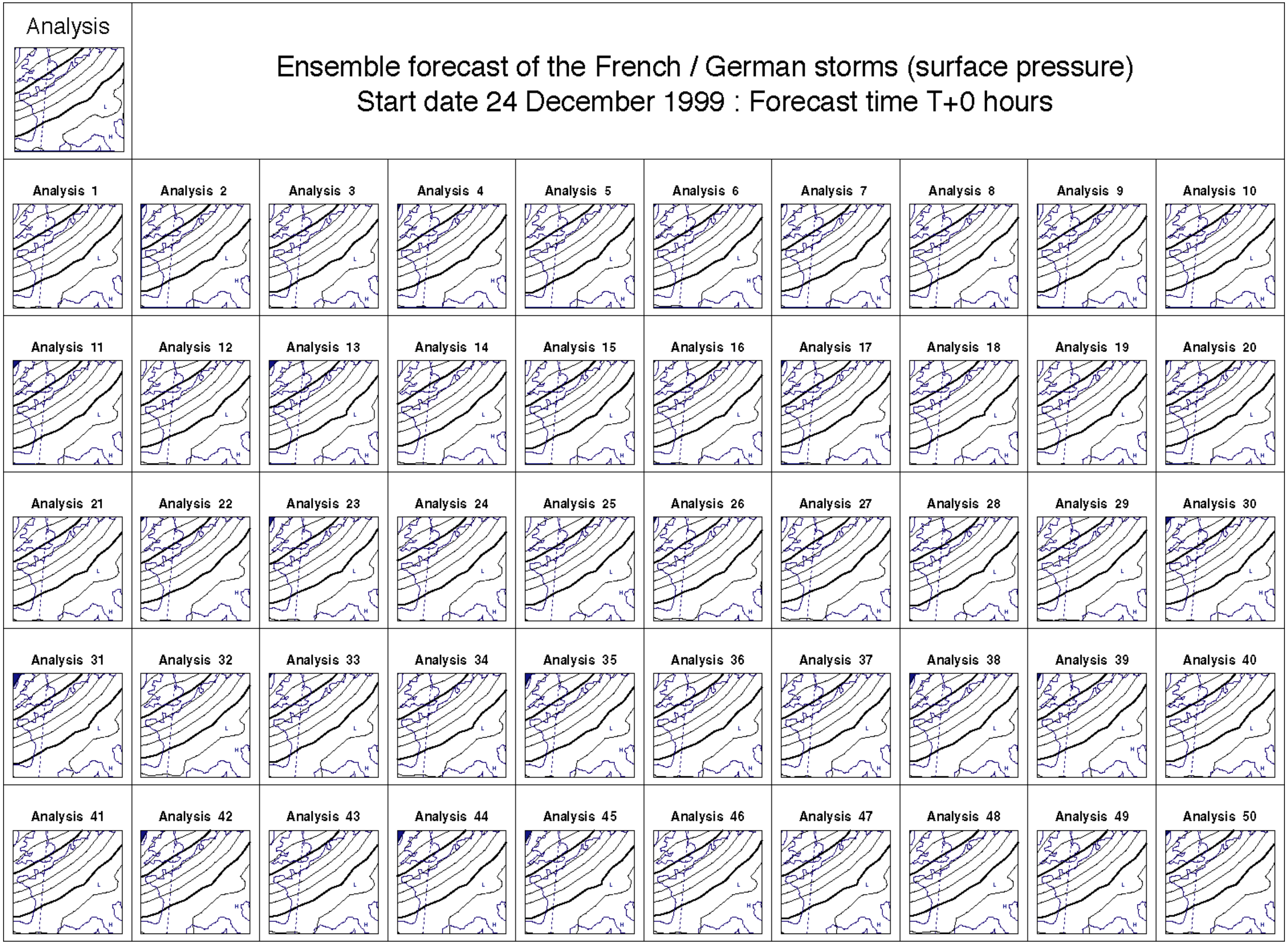
K-Means vs. SOM

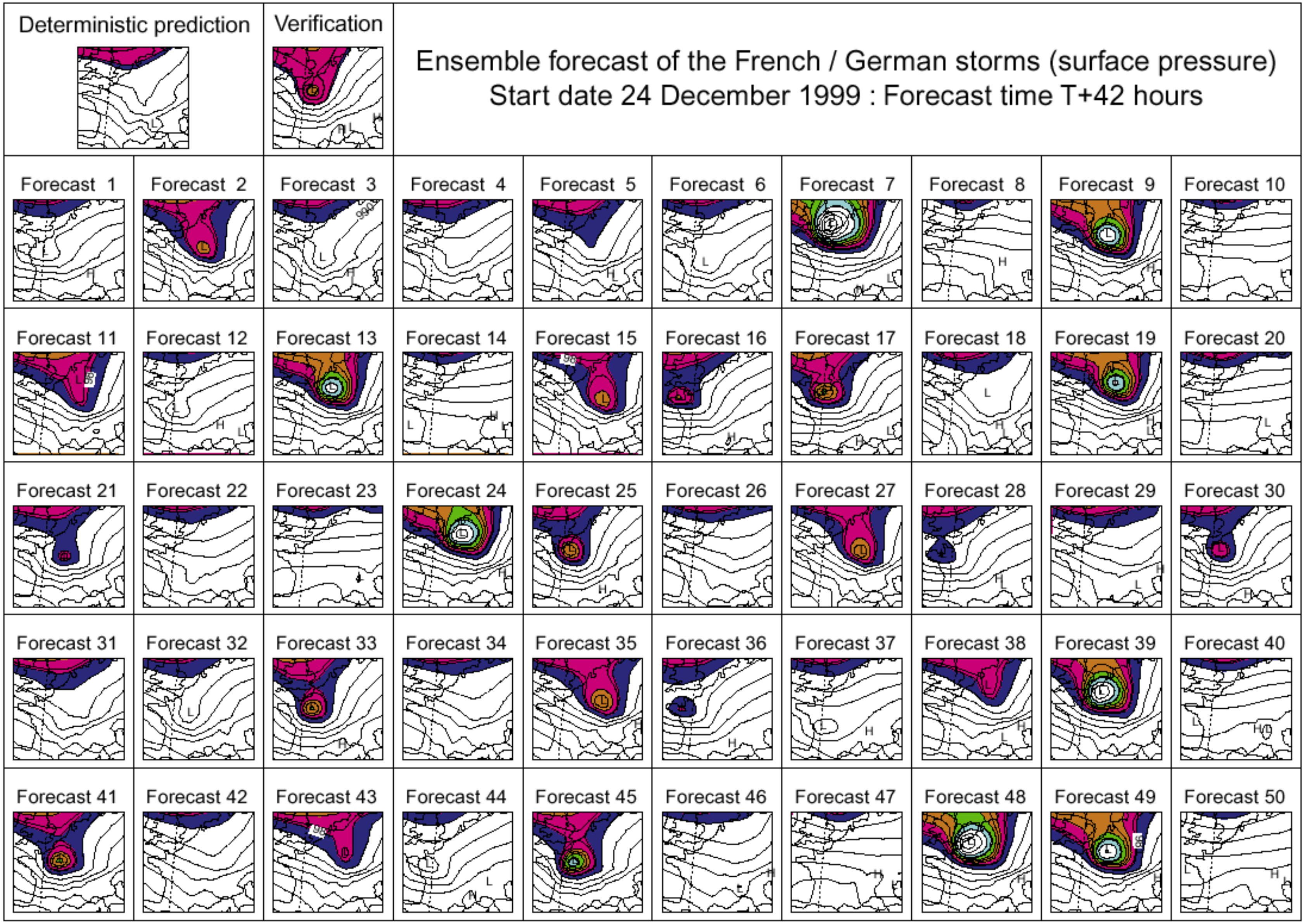
```
library(stats)
? kmeans
km<-kmeans(iris[,-5],3,nstart=1)
summary(km)
## Point center of two attributes
plot(iris[,c(1,3)],col=km$cluster)
points(km$centers[,c(1,3)],col=1:3,cex=5)
confusionMatrix(as.numeric(iris[,5]),km$cluster)

library(RSNNS)
? som
inputs <- normalizeData(iris[,1:4], "norm")
model <- som(inputs, mapX=16, mapY=16, maxit=500,
              calculateActMaps=TRUE, targets=iris[,5])

plotActMap(model$labeledMap,col=rev(topo.colors(12)))

par(mfrow=c(2,2))
for(i in 1:ncol(inputs))
      plotActMap(model$componentMaps[[i]],col=rev(topo.colors(12)))
```

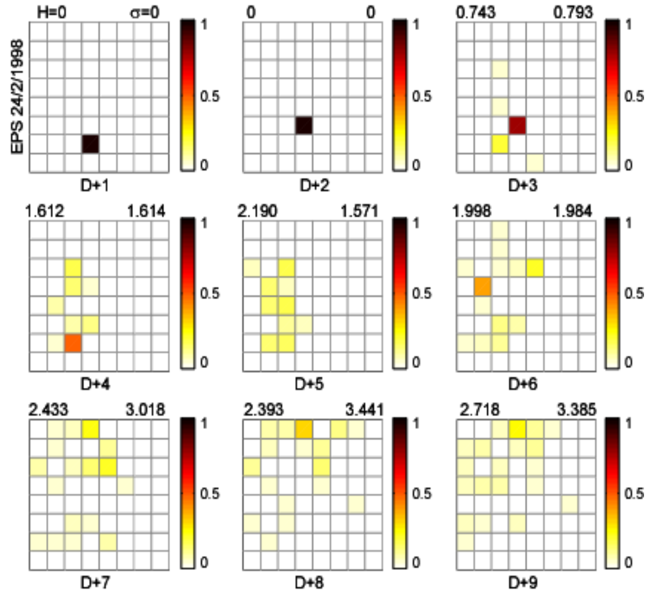
Lothar: 08Z, 26 Dec. 1999



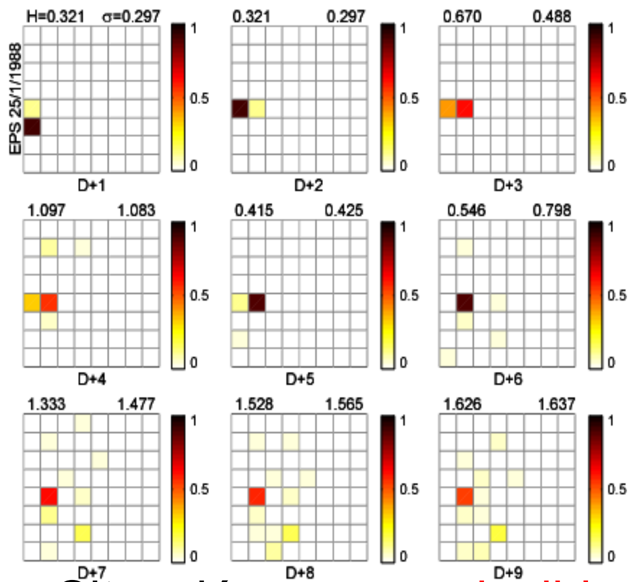
- 100 muertes
- 400 milloes de árboles derribados
- 3.5 millones de usuarios sin electricidad 20 días.
- 3 millones de personas sin agua


Analysis
Ensemble forecast of the French / German storms (surface pressure)
Start date 24 December 1999 : Forecast time T+0 hours
Analysis 1
Analysis 2
Analysis 3
Analysis 4
Analysis 5
Analysis 6
Analysis 7
Analysis 8
Analysis 9
Analysis 10
Analysis 11
Analysis 12
Analysis 13
Analysis 14
Analysis 15
Analysis 16
Analysis 17
Analysis 18
Analysis 19
Analysis 20
Analysis 21
Analysis 22
Analysis 23
Analysis 24
Analysis 25
Analysis 26
Analysis 27
Analysis 28
Analysis 29
Analysis 30
Analysis 31
Analysis 32
Analysis 33
Analysis 34
Analysis 35
Analysis 36
Analysis 37
Analysis 38
Analysis 39
Analysis 40
Analysis 41
Analysis 42
Analysis 43
Analysis 44
Analysis 45
Analysis 46
Analysis 47
Analysis 48
Analysis 49
Analysis 50

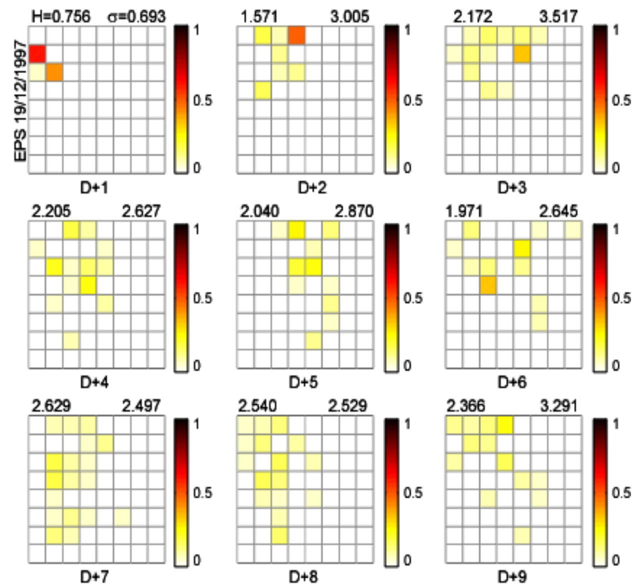

Deterministic prediction
Verification
Ensemble forecast of the French / German storms (surface pressure)
Start date 24 December 1999 : Forecast time T+42 hours
Forecast 1
Forecast 2
Forecast 3
Forecast 4
Forecast 5
Forecast 6
Forecast 7
Forecast 8
Forecast 9
Forecast 10
Forecast 11
Forecast 12
Forecast 13
Forecast 14
Forecast 15
Forecast 16
Forecast 17
Forecast 18
Forecast 19
Forecast 20
Forecast 21
Forecast 22
Forecast 23
Forecast 24
Forecast 25
Forecast 26
Forecast 27
Forecast 28
Forecast 29
Forecast 30
Forecast 31
Forecast 32
Forecast 33
Forecast 34
Forecast 35
Forecast 36
Forecast 37
Forecast 38
Forecast 39
Forecast 40
Forecast 41
Forecast 42
Forecast 43
Forecast 44
Forecast 45
Forecast 46
Forecast 47
Forecast 48
Forecast 49
Forecast 50



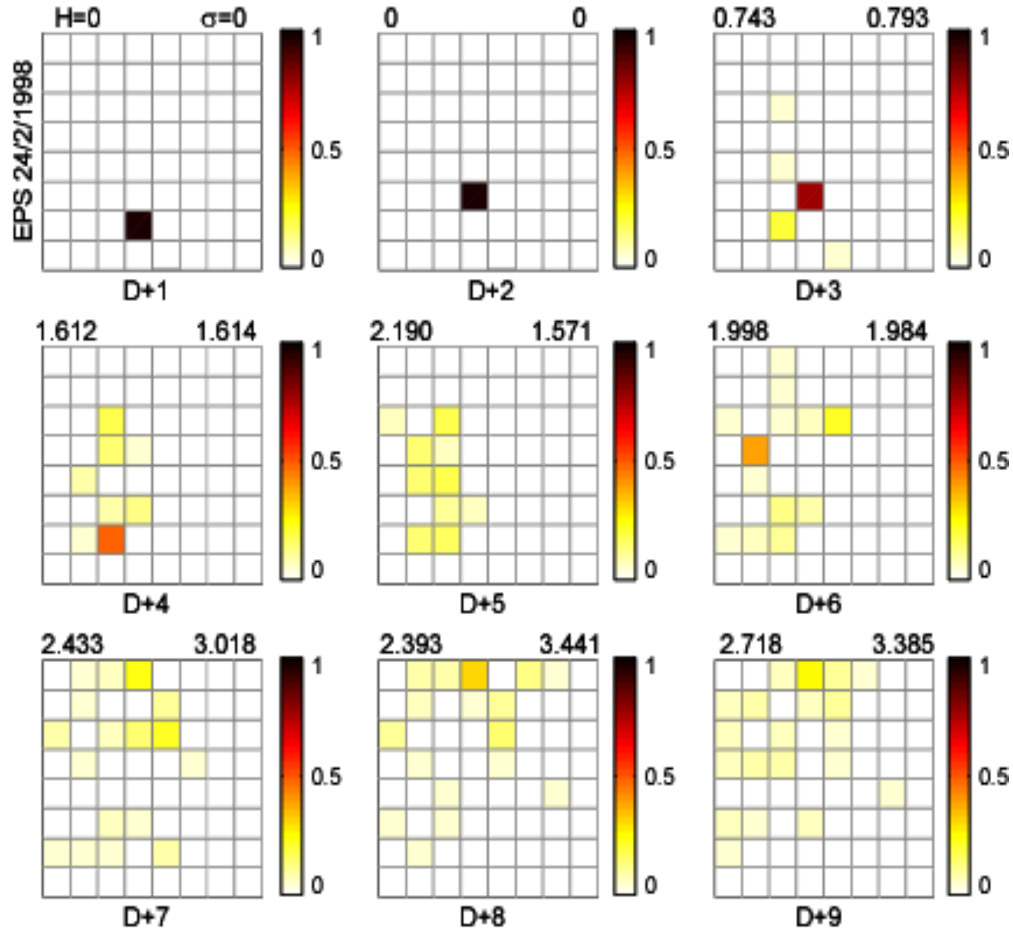
Las redes auto-organizativas nos permiten estudiar la dispersión del EPS de forma gráfica a través de una retícula que representa distintos grupos de patrones atmosféricos (Eckert and Cattani, 1996).



Situación muy predecible



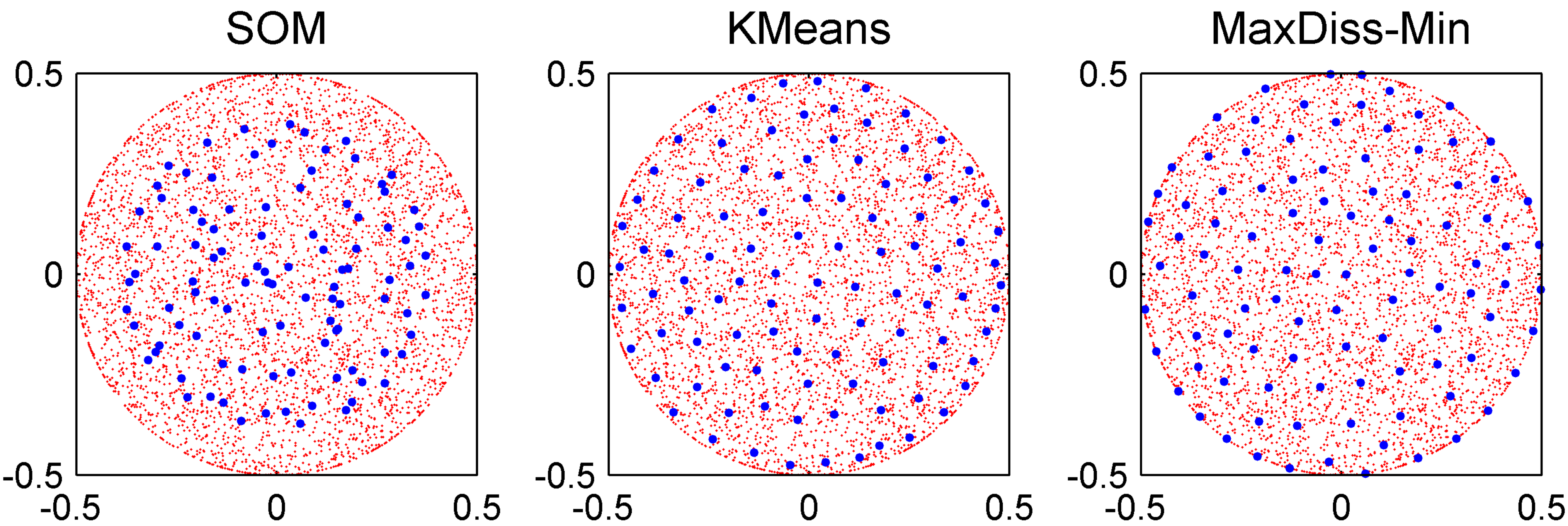
Situación poco predecible



$$H(P) = -\sum_{i} p_i \, log_b p_i = \sum_{i} p_i \, log_b \frac{1}{p_i} = E\left[log_b \frac{1}{P(x)}\right]$$

$$H(P|Q) = \sum_{q_i \neq 0} p_i \, log \frac{p_i}{q_i}$$

COMPETITIVE:

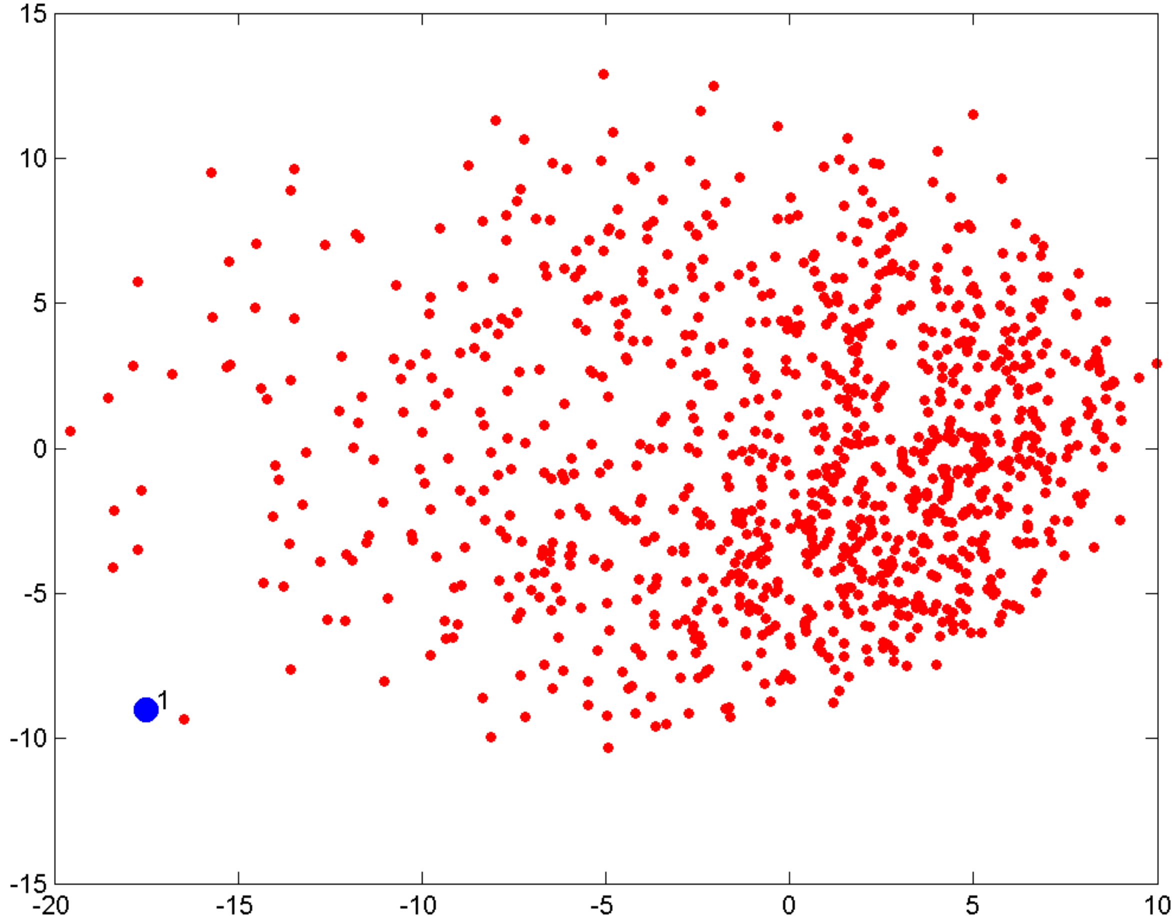

SOM
KMeans
MaxDiss-Min
0.5
0
-0.5
-0.5
0
0.5

MaxDiss – Maximum dissimilarity

Muestra de datos bidimensionales $x_i \quad i = 1, \ldots, N$

Objetivo: Subconjunto formado por 16 datos

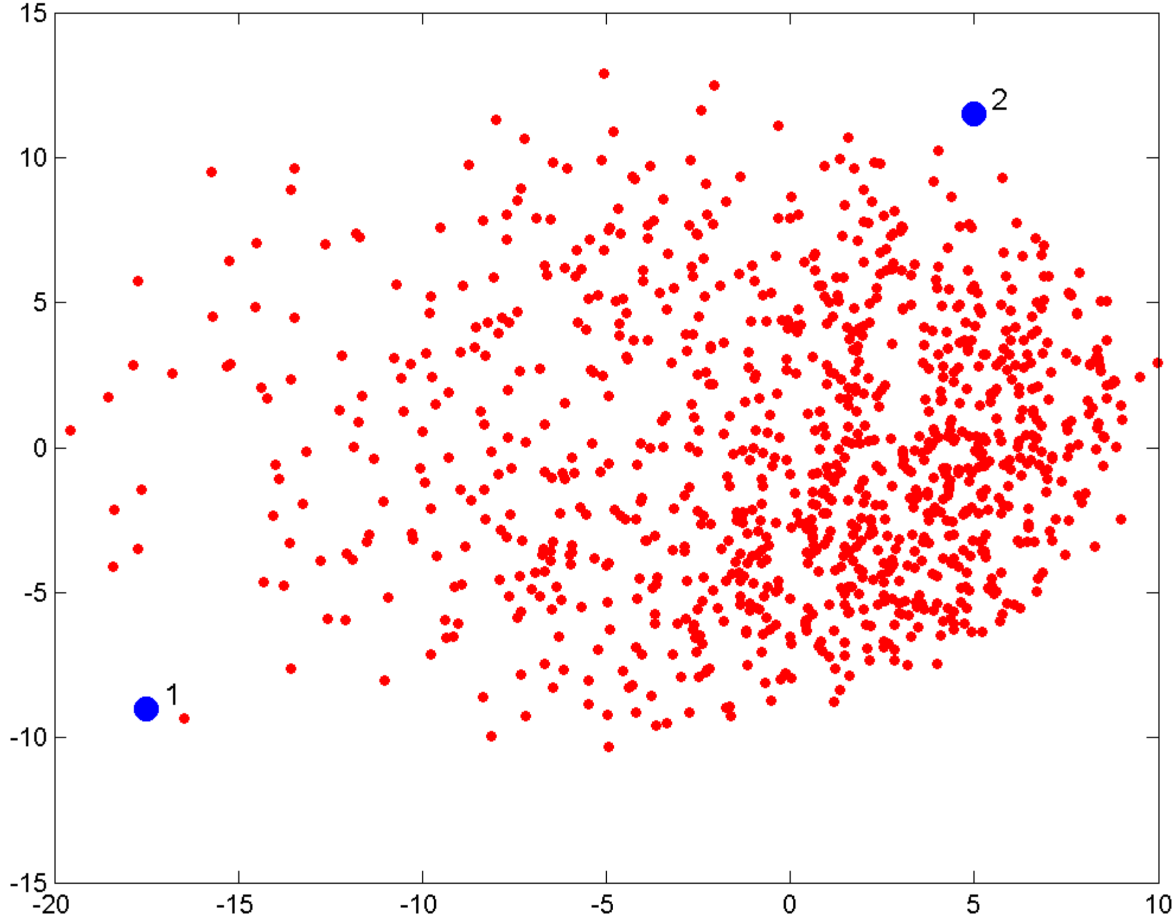
Inicialización: dato más diferente $\longrightarrow D_i = \sum_{j=1}^{N-1} \left\| x_i - x_j \right\|; i = 1, \ldots, N \longrightarrow \max \left\{ D_i; i = 1, \ldots, N \right\}$



MaxDiss

Subconjunto: $\{v_1\}$

Nuevo dato del Subconjunto $\{v_2\}$: $\max\left\{d_{i,subconjunto} = \|x_i - v_1\|; i = 1,\ldots,N-1\right\}$



MaxDiss

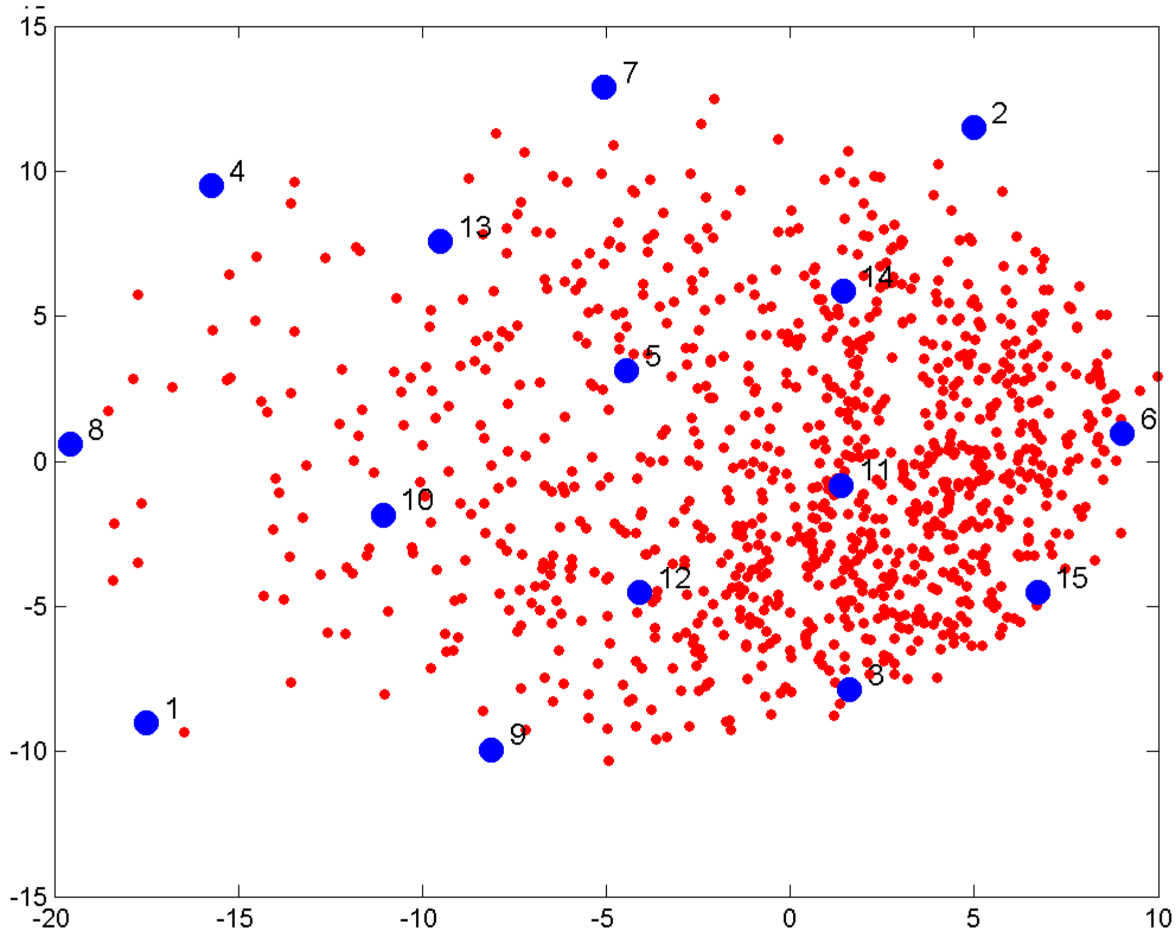
Subconjunto: $\{v_1, v_2\}$

Nuevo datos del subconjunto $\{v_3\}$:

$$d_{ij} = \left\| x_i - v_j \right\|; i = 1,\ldots,N-2; j = 1,\ldots,2$$

$$d_{i,subconjunto} = \min\left\{\left\| x_i - v_j \right\|; j = 1,\ldots,2\right\}; i = 1,\ldots,N-2$$

$$\max\left\{d_{i,subconjunto}; i = 1,\ldots,N-2\right\}$$



MaxDiss

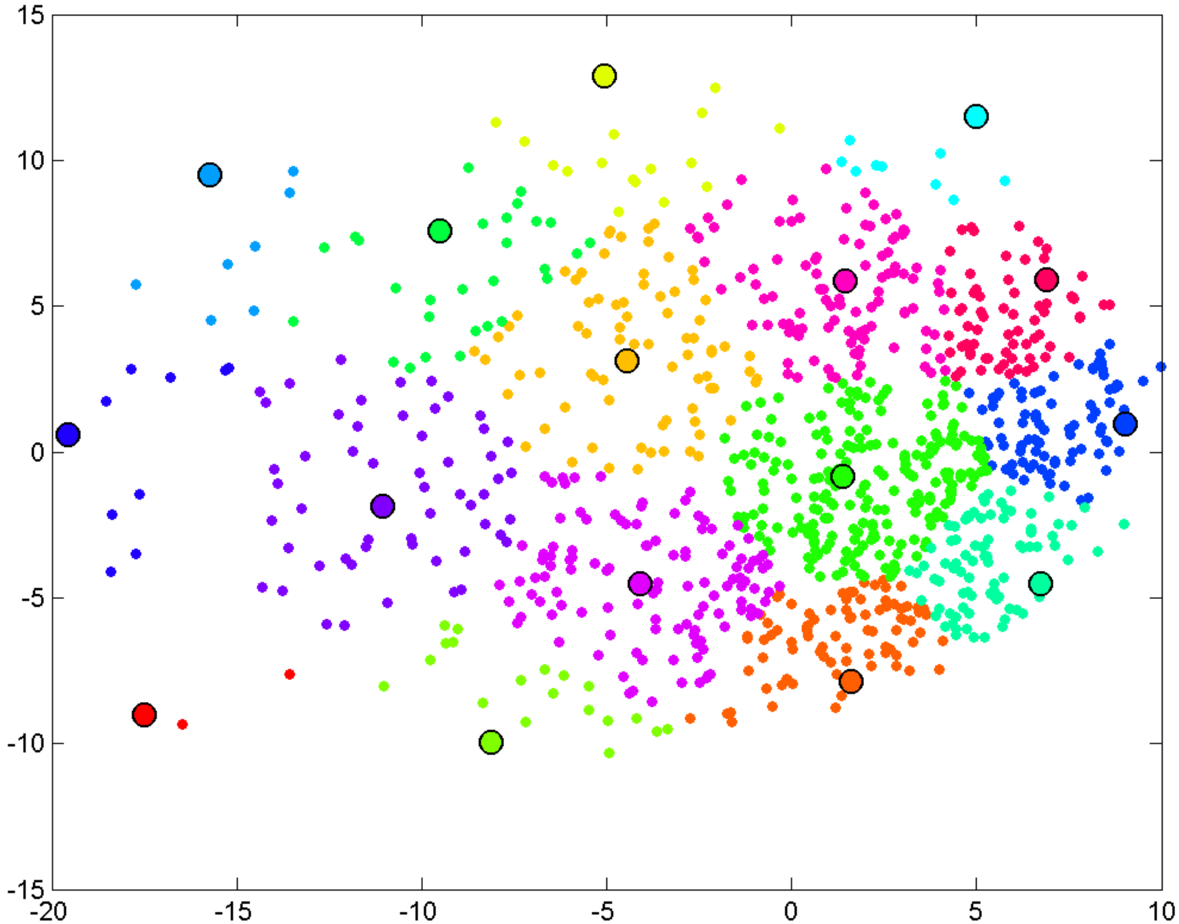
Subconjunto: $\{v_1, v_2\}$

Nuevo datos del subconjunto $\{v_3\}$:

$$d_{ij} = \left\| x_i - v_j \right\|; i = 1, \ldots, N-2; j = 1, \ldots, 2$$

$$d_{i,subconjunto} = \min\left\{\left\| x_i - v_j \right\|; j = 1, \ldots, 2\right\}; i = 1, \ldots, N-2$$

$$\max\left\{d_{i,subconjunto}; i = 1, \ldots, N-2\right\}$$



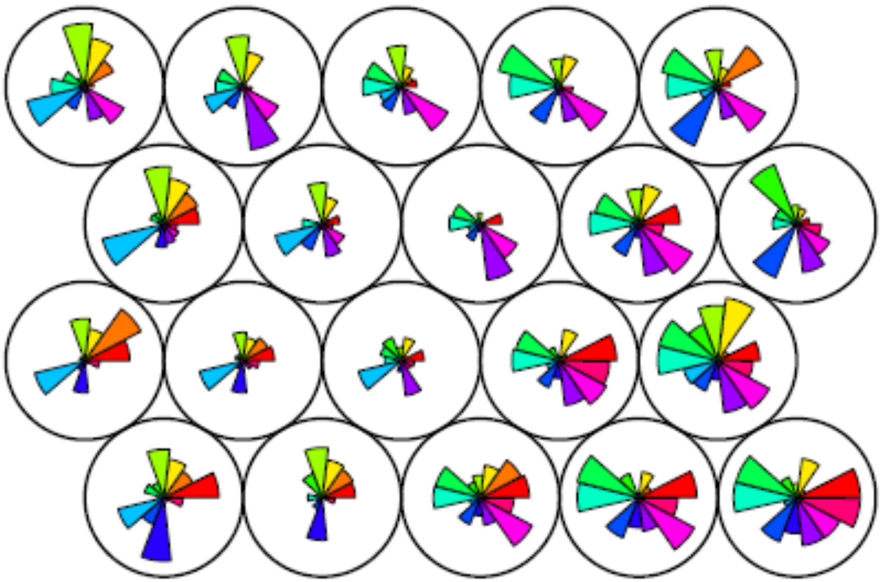
Wine Data Set

Download: Data Folder, Data Set Description

Abstract: Using chemical analysis determine the origin of wines

https://archive.ics.uci.edu/ml/datasets/wine

Data Set Characteristics:	Multivariate	**Number of Instances:**	178	**Area:**	Physical
Attribute Characteristics:	Integer, Real	**Number of Attributes:**	13	**Date Donated**	1991-07-01
Associated Tasks:	Classification	**Missing Values?**	No	**Number of Web Hits:**	831234



Crear una red auto-organizativa para crear una segmentación del problema de los vinos (178 vinos en Italia).

- alcohol
- malic acid
- ash
- ash alkalinity
- magnesium
- tot. phenols
- flavonoids
- non−flav. phenols
- proanth
- col. int.
- col. hue
- OD ratio
- proline

Uso de paquetes más avanzados:

https://en.wikibooks.org/wiki/Data_Mining_Algorithms_In_R/Clustering/Self-Organizing_Maps_(SOM)