



VNIVERSITAT ID VALÈNCIA

# MASTER DE INGENIERÍA BIOMÉDICA.

## Métodos de ayuda al diagnóstico clínico.

### Tema 3: Preprocesado de los datos.

# Objetivos del tema

**Conocer los problemas que nos podemos encontrar en los datos proporcionados por los clínicos.**

**Aprender algunos métodos para completar patrones incompletos**

**Conocer los métodos más extendidos de recodificación de variables.**

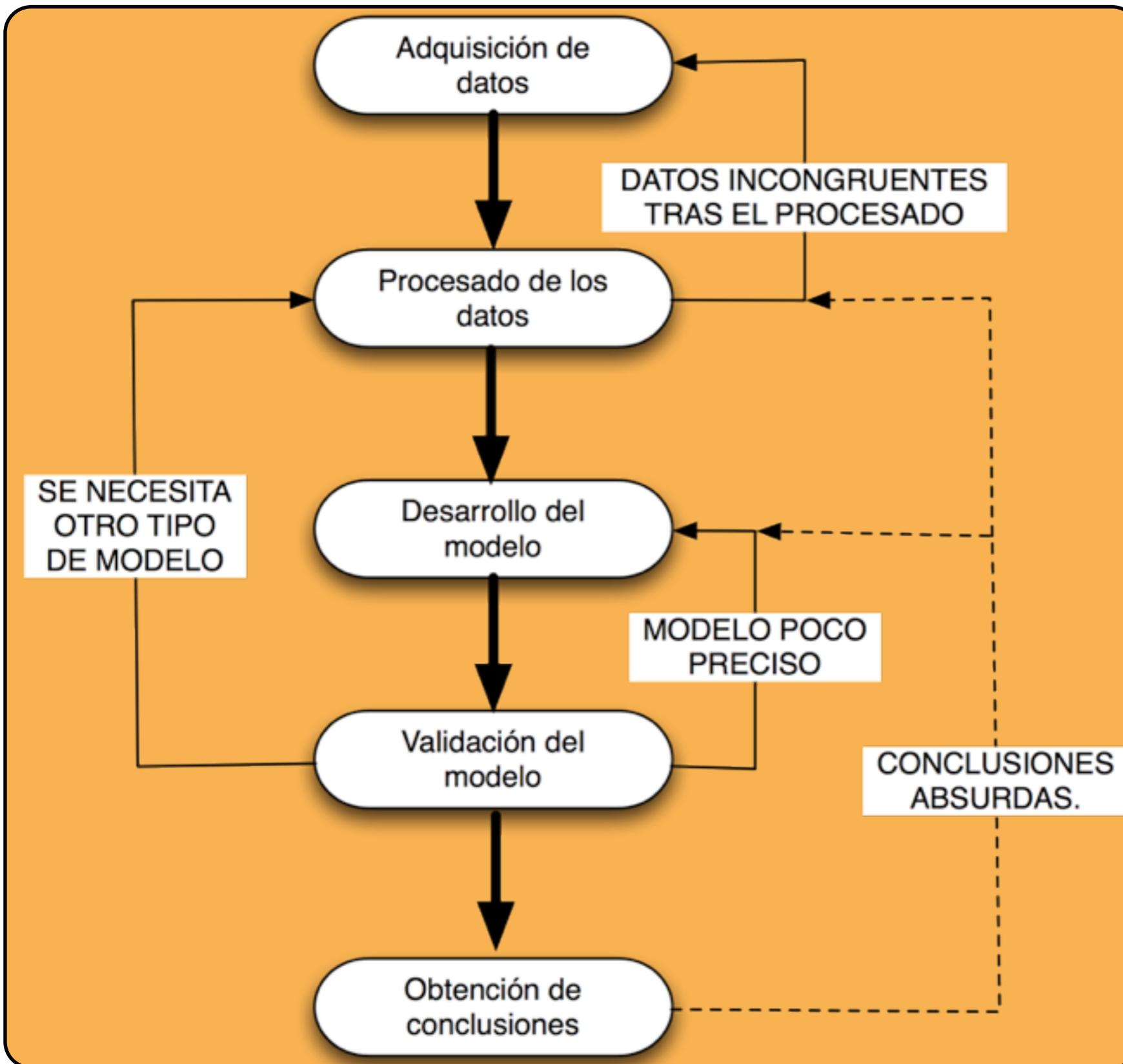
**Aprender lo que se denomina la “maldición de la dimensionalidad” y saber la relación con los problemas clínicos.**

**Conocer el método de análisis de componentes principales para reducir la dimensionalidad de los datos.**

**Aprender los métodos de agrupamiento (clustering) y el uso que podemos darle en el análisis de datos de clínica.**

**Conocer nuestra primera estructura neuronal artificial, los mapas autoorganizados, y su uso en el agrupamiento de datos.**

# ¿Qué tenemos que hacer con los datos?



Hay que tener en cuenta SIEMPRE que todo proceso de construcción de un modelo de ayuda al diagnóstico clínico es un proceso realimentado, todas las etapas dependen de las anteriores o influyen en las posteriores. Todas las etapas son, entonces, igual de importantes

# Preprocesado de datos.

Según el libro que se consulte las tareas de un preprocesado de datos las engloba, o no, dentro del desarrollo del modelo que se va a utilizar (modelo lineal, redes neuronales, árbol de decisión, etc). Nosotros vamos a entender como preprocesado de los datos como todo aquel conjunto de operaciones que se aplican a los datos antes de desarrollar el modelo.

TAREA	OBJETIVOS
Detección de datos anómalos e incompletos	Obtener un conjunto de datos completo sin incoherencias
Representación visual de los datos	Obtención de posibles relaciones de forma gráfica
Caracterización estadística de las variables	Detección de outliers. Identificación de las relaciones elementales entre datos, etc
Selección y reducción de características	Reducción del número de variables de un modelo
Algoritmos de clustering	Reducción del número de variables del modelo y/o del número de datos

# Definiciones.

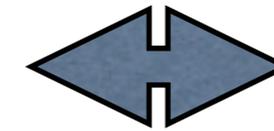
Consideraremos los datos descritos por un conjunto de variables o características. Hablaremos de patrones (muestras) para referirnos a cada elemento individual de la base de datos.

En cuanto al tipo de variables se pueden tener variables cualitativas (en principio no se pueden hacer operaciones numéricas con ellas aunque se codifiquen con un número) y cuantitativas. Las cualitativas pueden ser variables **ordinales** (sus valores se pueden ordenar, ¿grado de dolor?) y **nominales** (no se pueden ordenar, ¿religión?). Las variables cuantitativas pueden ser o bien **discretas** (¿número de hijos?) o bien **continuas** (concentración plasmática de fármaco?).

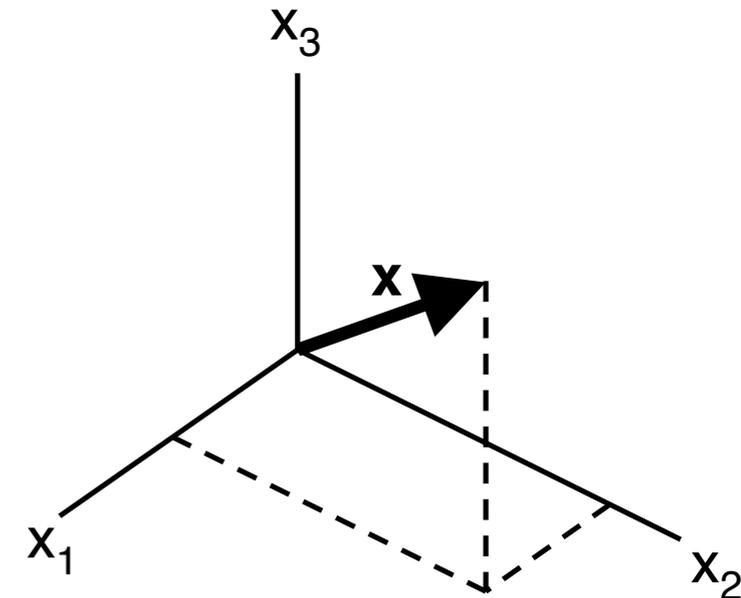
	V1	V2	V3
Patrón 1	75	grado medio	si
Patrón 2	25	grado inferior	no
Patrón 3	35	grado superior	si

Patrón  $x$

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix}$$



Representación vectorial



# Patrones incompletos (I).

El problema que nos encontramos es que tenemos patrones a los que les faltan los valores de determinadas variables; ¿qué hacemos en esos casos?.

La respuesta a esta pregunta es como todo en análisis de datos...¡depende!. Si el dato tiene bastantes variables incompletas lo mejor es no considerar dicho dato o patrón. En nuestro caso pac.5 no se tendría en cuenta.

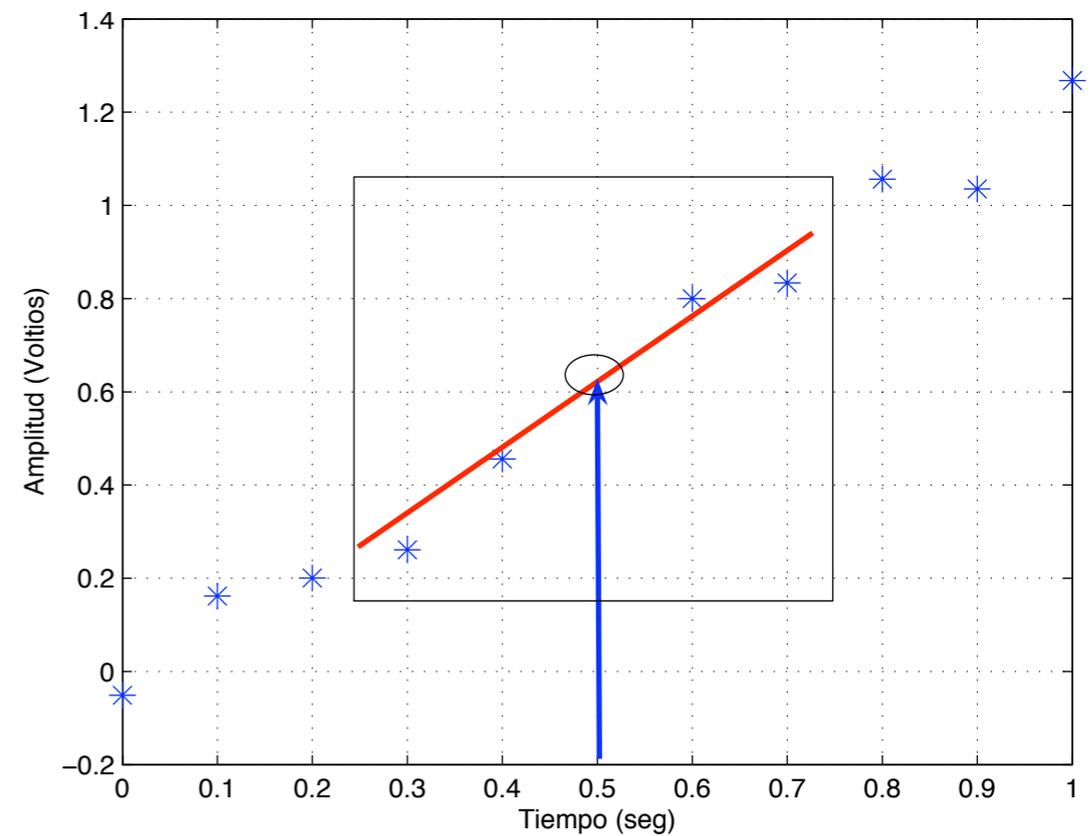
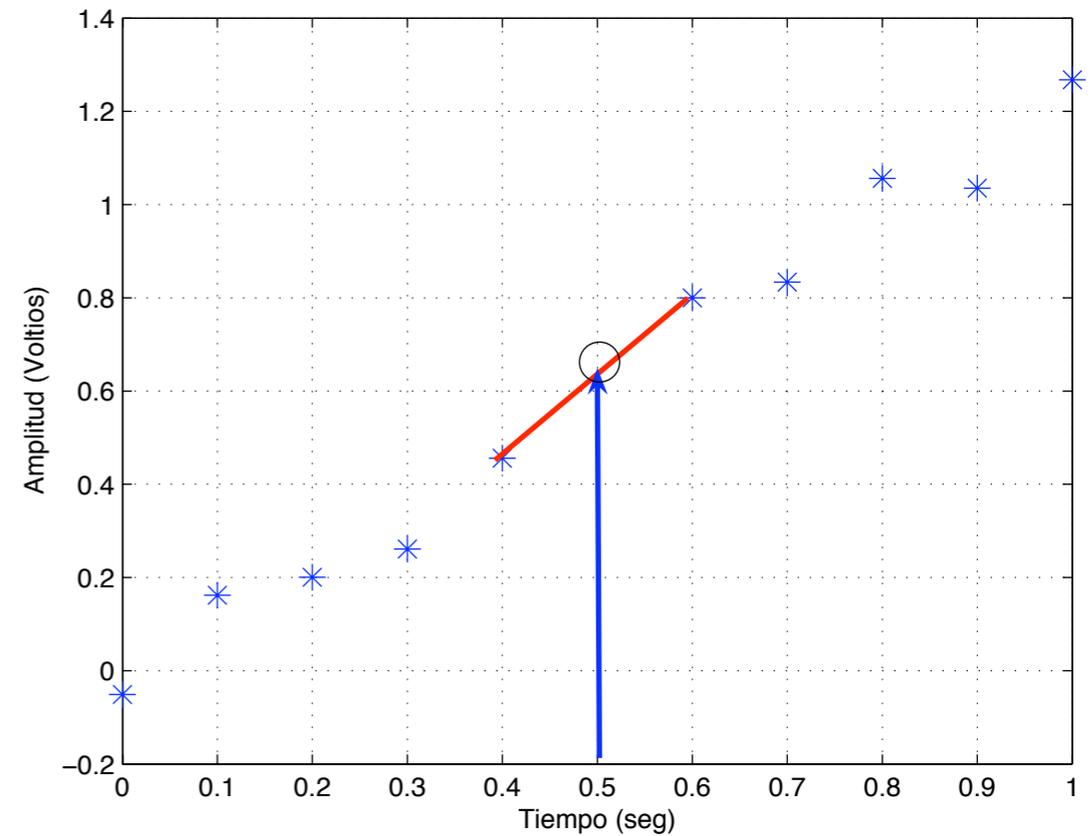
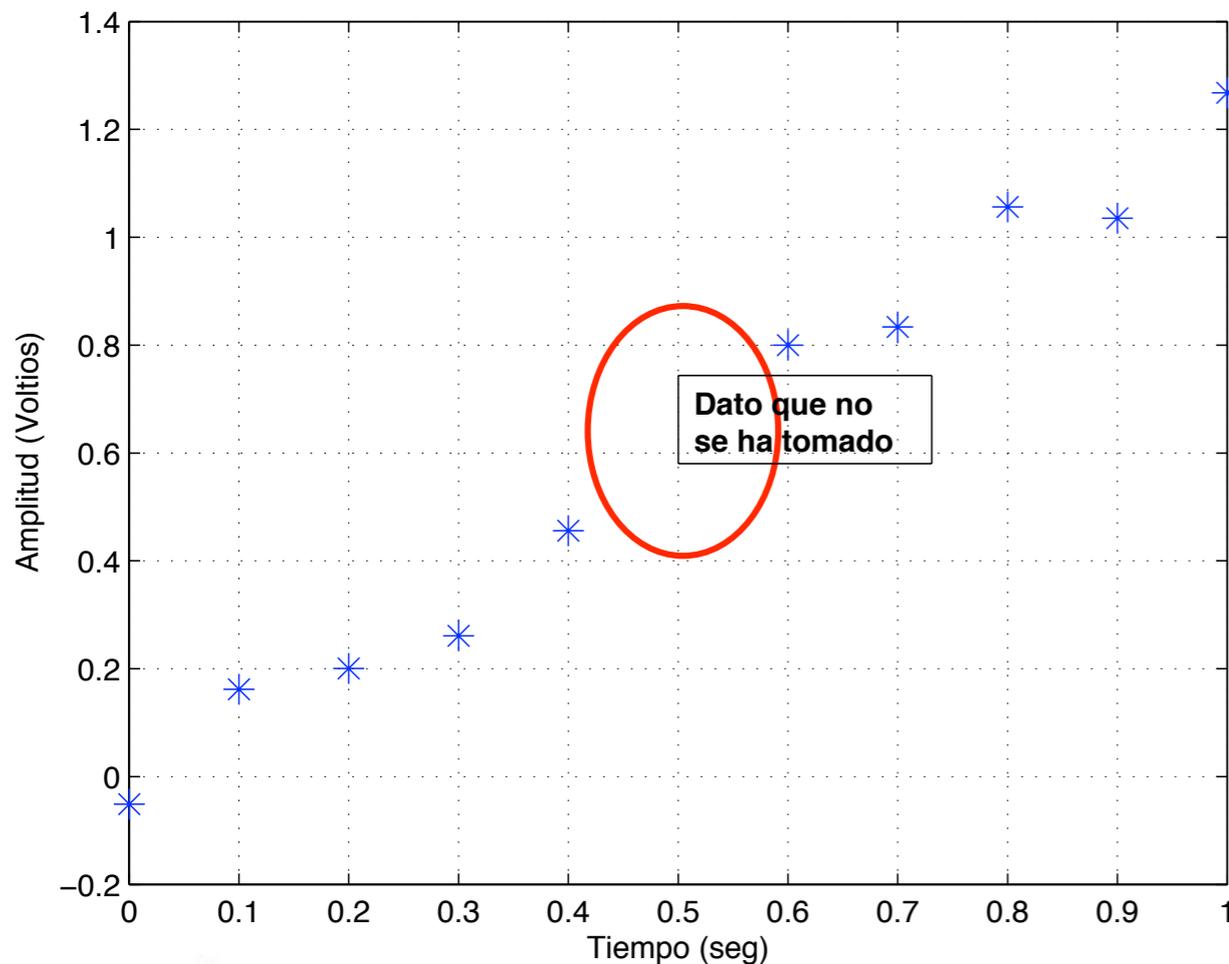
Si tenemos un patrón que *se parece a otro* (¿cómo podemos determinar su parecido?) podemos usar el valor de la variable que aparece en dicho patrón. Por ejemplo el paciente 2 se parece al paciente 4; podemos usar como variable 5 del paciente 2 el valor 4,6.

Otra cuestión importante es el considerar o no una determinada variable; en nuestro caso no sería muy astuto considerar la variable 5 porque faltan bastantes valores. No hay que olvidar que toda información incorporada PUEDE NO REFLEJAR LA REALIDAD. Por eso lo mejor es no abusar.

	V1	V2	V3	V5
Pac1	1,1	2,3	4,5	8,9
Pac2	2,0	3,2	9,7	
Pac3	1,5	1,2	5,6	
Pac4	2,1	3,3	9,9	4,6
Pac5	5,6			
Pac6	5,4	6,7	9,2	1,1
Pac7	3,4	2,2	6,6	

# Patrones incompletos (II).

Otro tipo de datos muy usados y que aparecen en clínica son las series temporales; variables en las que existe una dependencia con el tiempo (Señales e Imágenes Médicas.....). Aquí podemos aprovecharnos de esa dependencia para establecer otro procedimiento para procesar los datos incompletos.



# Codificación de los datos. Transformaciones.

Se conoce como **codificación** al proceso por el cual a las variables cualitativas se les asigna un determinado valor. En general no hay reglas para establecer esta codificación pero hay que intentar, siempre que se pueda, tener una codificación simétrica con valor central 0; evitando en la codificación el valor de 0

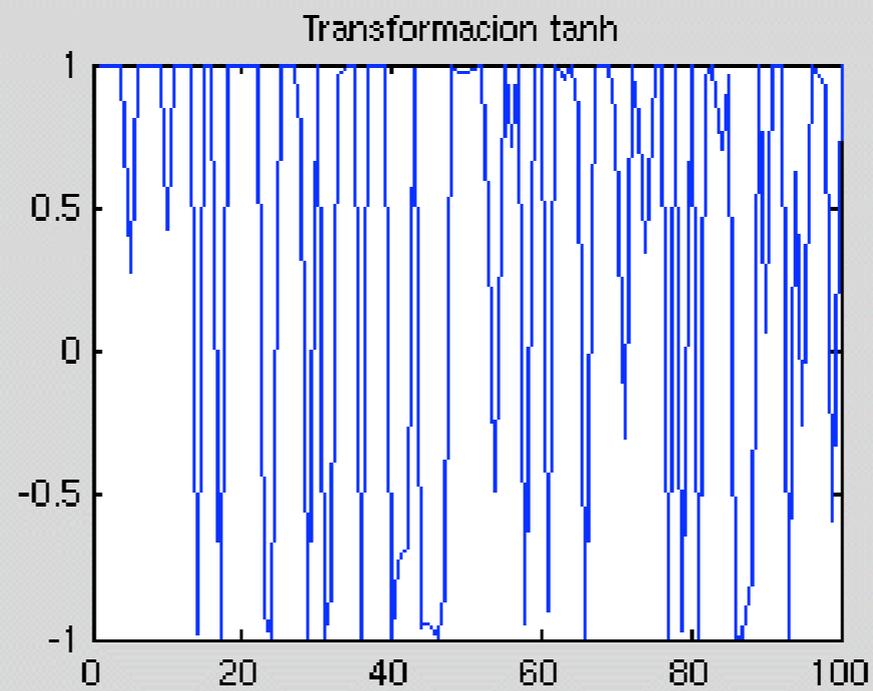
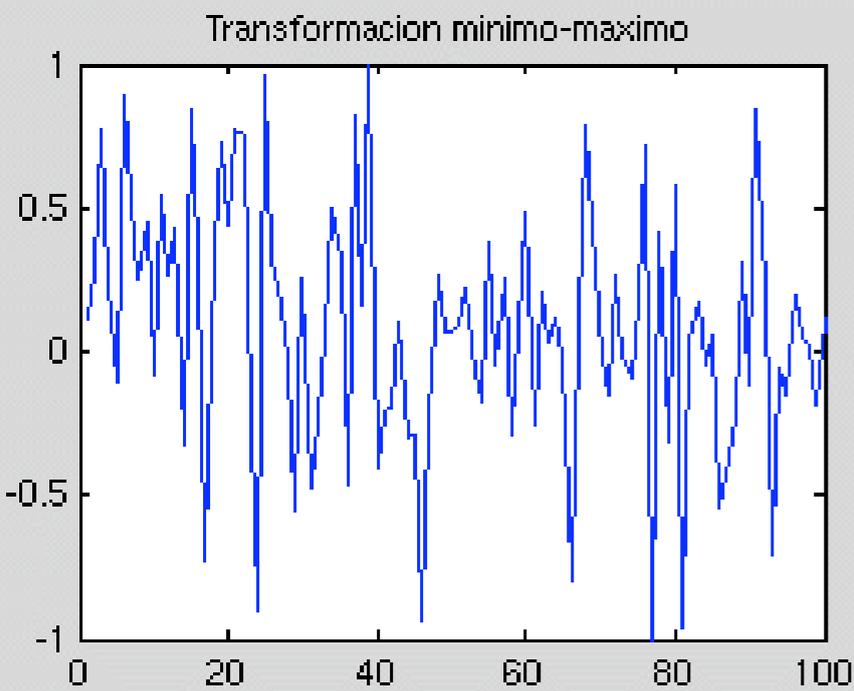
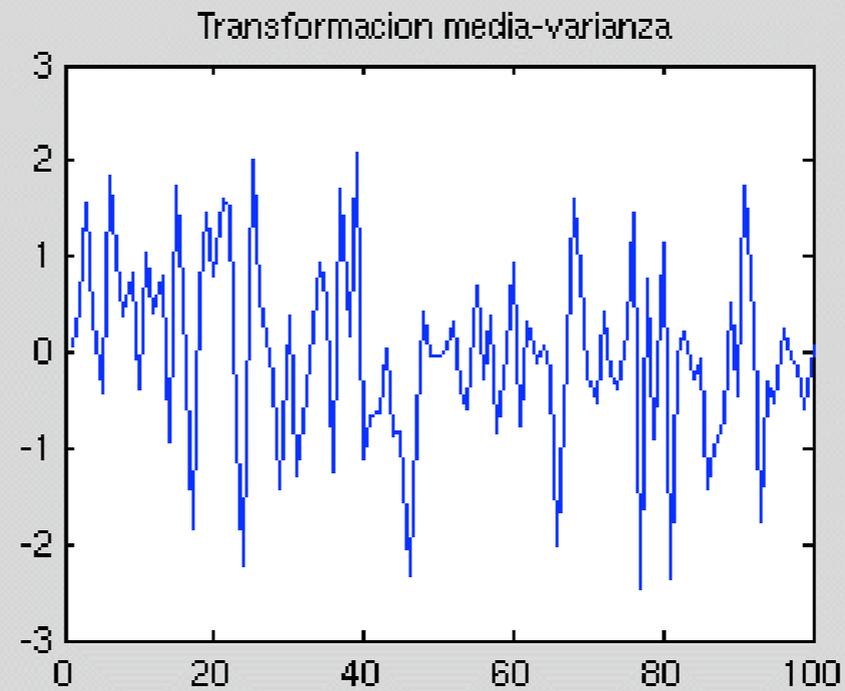
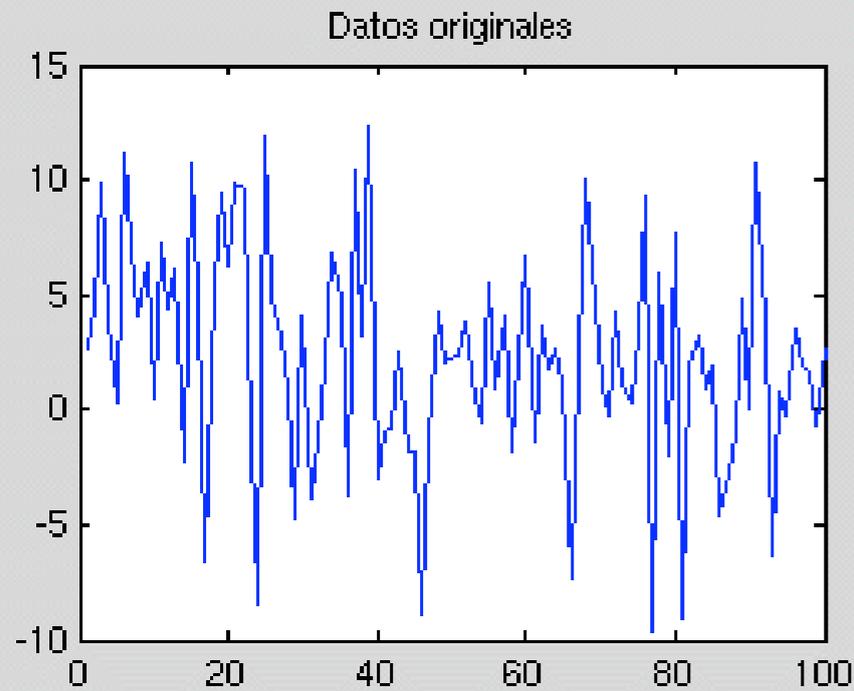
Se conoce como **transformación** al proceso por el cual se aplica a las variables una determinada función para obtener otras variables que cumplen unas determinadas propiedades. Se aplica a variables cuantitativas.

Función	Misión
$y_k = \frac{x_k - m_x}{\sigma_x}$	La nueva variable tiene valor medio 0 y varianza 1.
$y_k = \frac{x_k - x_{mi}}{x_{ma} - x_{mi}} \cdot [y_{ma} - y_{mi}] + y_{mi}$	La variable está en $[y_{mi}, y_{ma}]$
$y_k = \tanh\left(\frac{x_k - m_x}{\sigma_x}\right)$	La variable está en $[-1, 1]$

Aquí  $m_x$  y  $\sigma_x$  son el valor medio y la desviación estándar de la variable  $x$ ;  $x_{mi}$  y  $x_{ma}$  son, respectivamente, el valor mínimo y máximo de dicha variable.

Otra transformación muy usada es la **discretización**; a partir de una variable continua se llega a una variable discreta. En ocasiones es más fácil trabajar con una variable discreta (modelización-clasificación).

# Ejemplos de transformaciones.



El ejemplo que se presenta es el de una variable aleatoria que tiene de media 3 y de desviación estándar 5. Destacar que, de todas las transformaciones la más usada suele ser la de media-varianza aunque esta transformación no me permite tener un rango fijo de la variable como ocurre con la transformación máximo y mínimo. La transformación tangente hiperbólica se usa cuando se tienen casi todos los datos concentrados en 0 y algunos toman un valor más o menos elevado (outliers).

# Análisis Exploratorio de Datos (EDA).

Se conoce como **Análisis Exploratorio de Datos (EDA)**, al uso de representaciones gráficas junto con el de estadísticos para obtener conocimiento acerca de las variables que forman nuestros datos.

Estadístico	Uso
Valor medio, mediana, moda	Tendencia central
Frecuencias	Distribución de las variables cualitativas
Cuartiles, Percentiles, IQR	Localización
Varianza, Desviación estándar	Dispersión.
Sesgo, Curtosis	Forma de la distribución.
Test-estadísticos	Comprobación de valores medios, tipos de distribución.

Gráfica	Uso
Barras	Descriptivo de los datos
Sectores	Determinación de frecuencias
Líneas	Tendencias de las variables
Histograma	Determinación de la distribución
Steam & Leaf	Determinación de la distribución
Boxplot	Representación de las medidas de posición más usadas
QQ-plot PP-plot	Comprobación del tipo de distribución.

# Problema de la dimensionalidad (I)

En los puntos anteriores hemos tratado como procesar los datos y extraer unas primeras conclusiones.

Ya estamos preparados para desarrollar el modelo....¿cuántas y qué variables tiene que considerar el modelo?. La respuesta más evidente sería todas ya que así tenemos toda la información recogida en el modelo. Sin embargo esta estrategia no es la óptima porque no podemos coger entradas sin límite, existen problemas.

Al aumentar el número de entradas hay que aumentar el número de patrones para ajustar el modelo

Se reduce la velocidad de convergencia del algoritmo de construcción del modelo

La precisión de los modelos es menor al aumentar el número de entradas

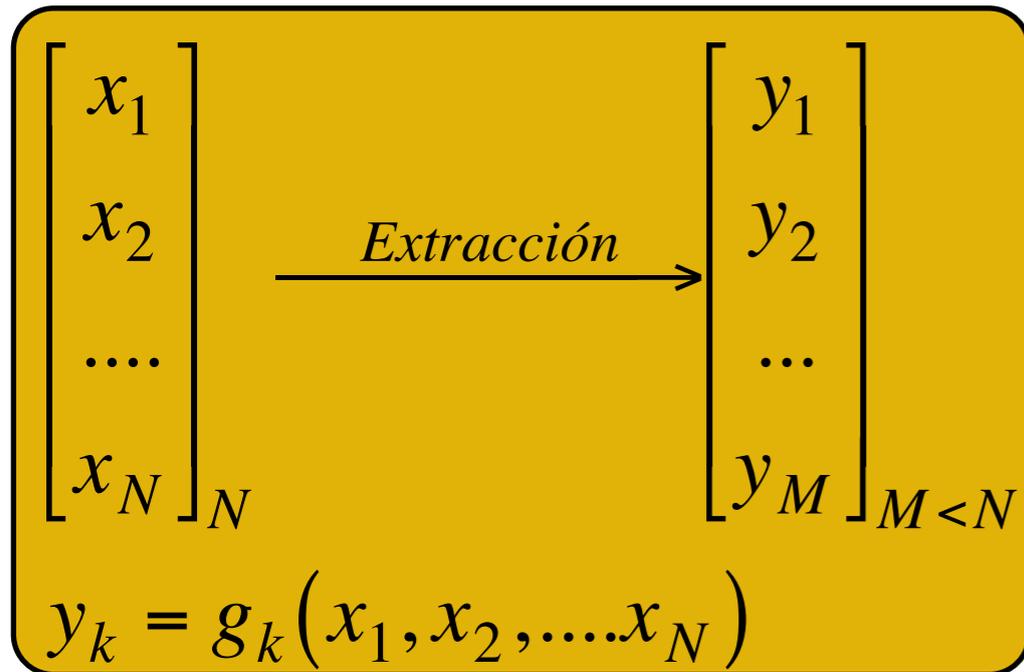
El análisis final del modelo es más complicado

Las entradas no necesarias para el modelo actuarán como ruido

**EL MEJOR MODELO NO ES SIEMPRE EL QUE CONTIENE TODAS LAS VARIABLES QUE TENEMOS**

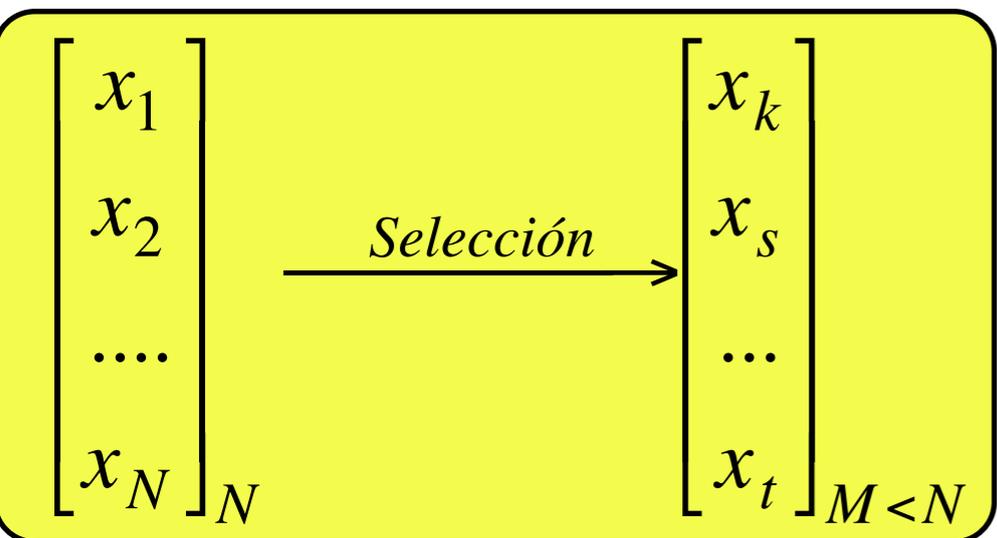
# Problema de la dimensionalidad (II)

Se hace necesario tener el número mínimo de entradas en el modelo que se vaya a desarrollar. Aquí se tienen dos estrategias posibles; se puede realizar una selección de las entradas (a este método se le conoce como **selección de características**) o bien podemos transformar el conjunto de variables que se tienen en un conjunto donde hay menos variables (**extracción de características**).



El problema de selección de características es un proceso muy complejo por el gran número de posibilidades que se tienen. Por ejemplo, si suponemos que tenemos N variables tenemos  $2^N$  posibilidades de elección (o no) de las entradas

Actualmente no existe un algoritmo de selección que funcione bien en todas las situaciones y problemas. Es importante que el experto defina, con la mayor precisión posible, los factores más importantes en el problema que se intenta resolver. Se puede obtener nuevos modelos con menos entradas analizando modelos obtenidos.

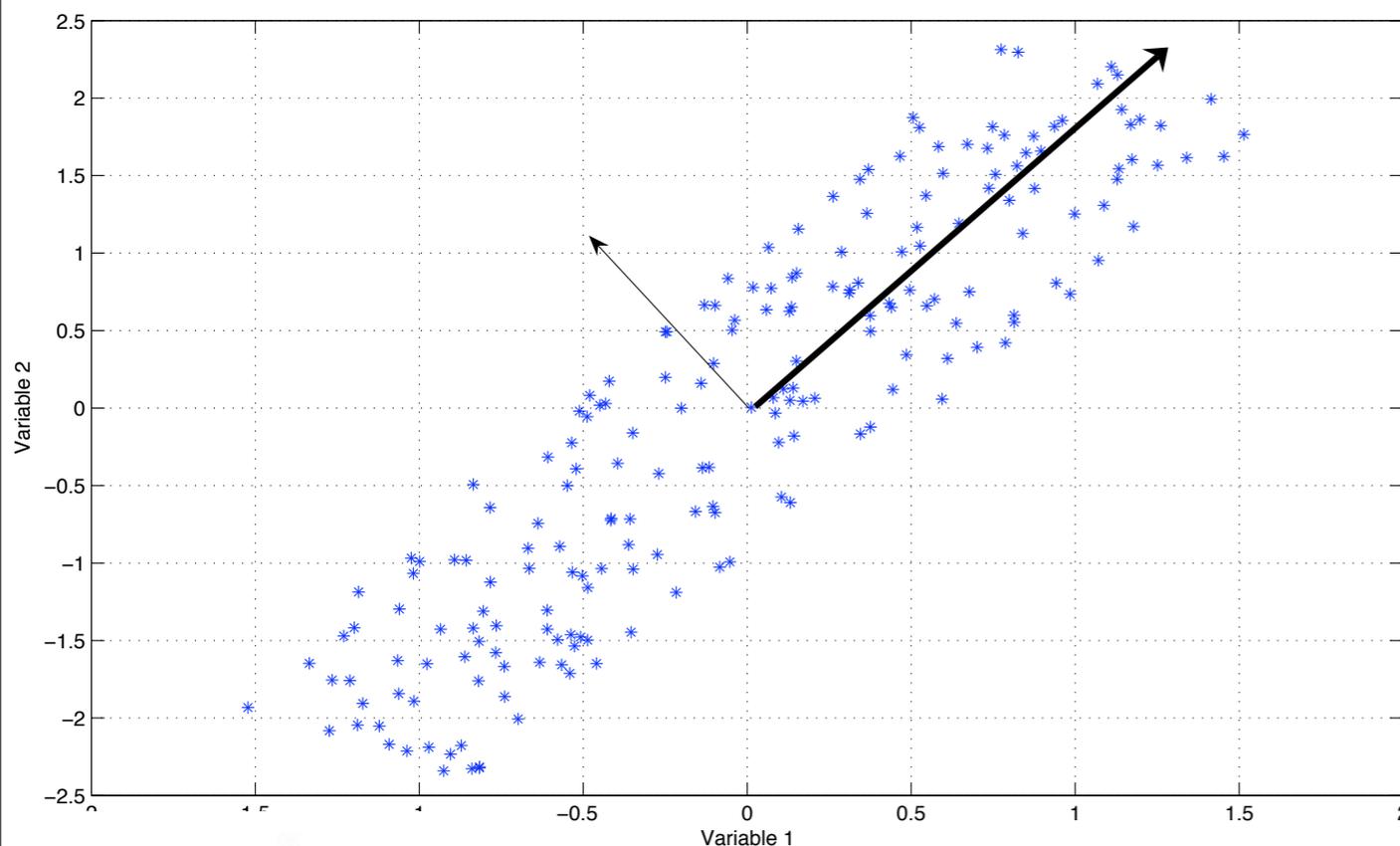


# Análisis por componentes principales, PCA (I)

La idea que hay detrás de este método es “puedo tener la misma información (o casi) en el conjunto inicial de variables que en otro conjunto con menos variables y que resulte de combinar linealmente el conjunto inicial”?

Hay que hacerse dos preguntas; ¿qué entendemos por información en este procedimiento? y cómo calculamos los coeficientes de la transformación.

En el ejemplo gráfico se observa que en la dirección de la flecha de mayor grosor es donde tenemos **mayor variabilidad de los datos**; para nosotros esa dirección es muy importante porque es la que contiene la mayor parte de información sobre la codificación de nuestros datos. A esta dirección se le denomina **componente principal**. Si tuviéramos un ejemplo de 3 dimensiones podría obtener 3 direcciones principales ordenándolas de acuerdo a la diferente variabilidad que presentan las variables. La idea es hacer un cambio de ejes y codificar las variables de acuerdo a las componentes principales que presentan mayor variabilidad.



# Análisis por componentes principales, PCA (II)

## Pasos para realizar una PCA.

Eliminamos el valor medio de los vectores que se tienen (1)

Calculamos la matriz de autocorrelación (2)

Calculamos la diagonalización de esa matriz (3)

Determinamos los vectores propios asociados al mayor autovalor (4)

Proyectamos cada vector en esos valores propios (5)

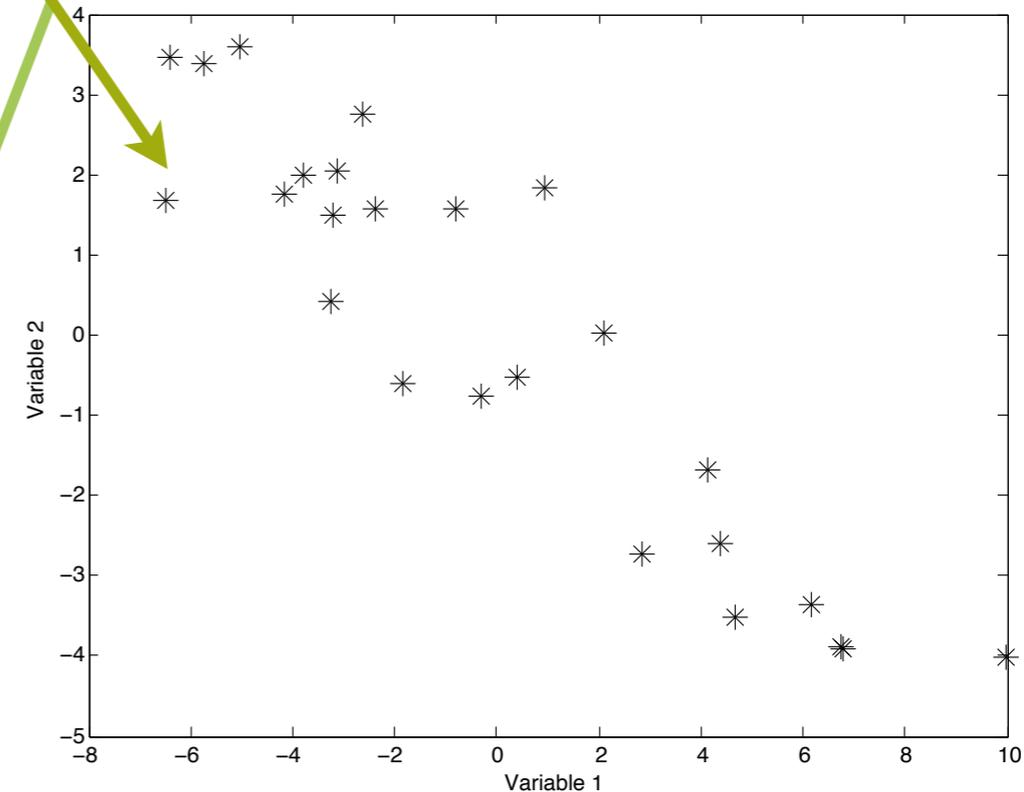
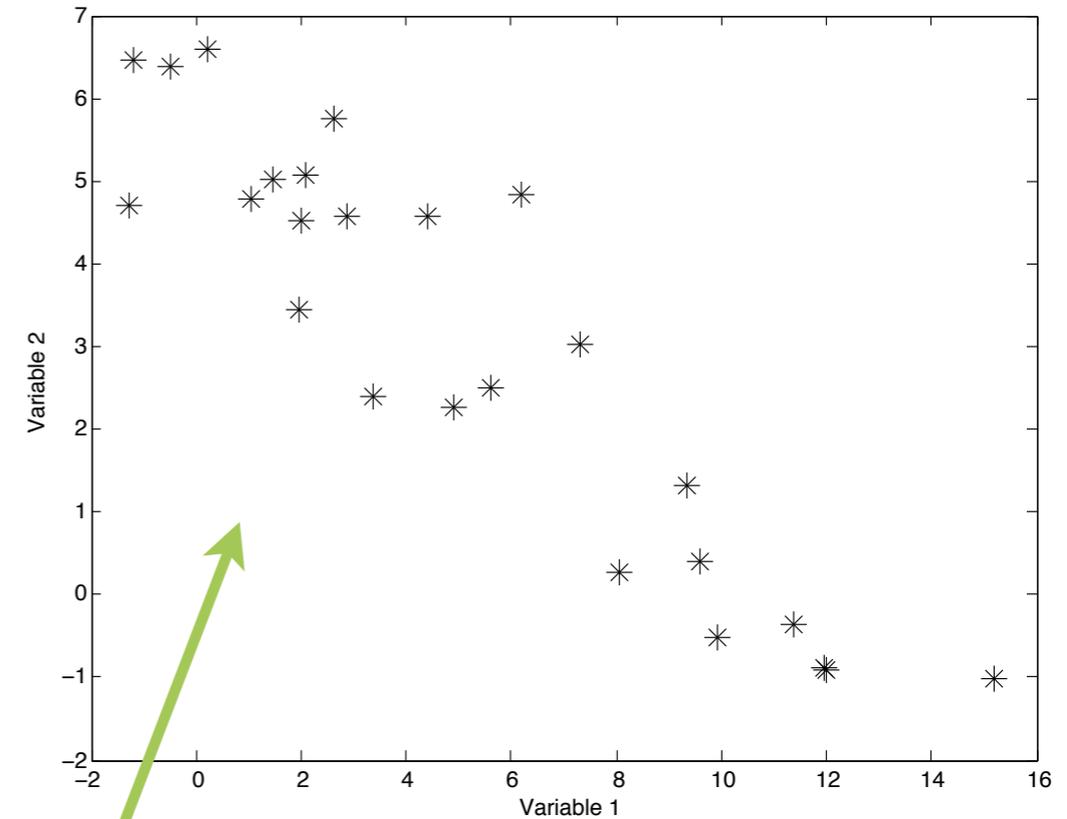
$v_k = [p_{1k} \quad p_{2k}]$  Datos originales (2 variables)

$$M = \frac{1}{N} \cdot \sum_{k=1}^N v_k \quad \text{Cálculo del valor medio}$$

$$u_i = v_i - M \quad \text{Eliminación del valor medio}$$

$$C = \frac{1}{N} \cdot \sum_{i=1}^N (u_i)^t \cdot (u_i) \quad \text{Matriz de autocorrelación.}$$

Aquí el valor medio es de 5.1 para la variable 1 y 3.0 para la variable 2.



# Análisis por componentes principales, PCA (II)

Con los datos que se tienen la matriz de autocorrelación a la que se llega es

$$C = \begin{bmatrix} 20,83 & 20,83 \\ -10,5 & 6,29 \end{bmatrix}$$

Aquí la operación complicada es la de diagonalizar la matriz pero todos los paquetes informáticos numéricos tienen esta operación; así en Matlab es una sola instrucción  $\gg [\text{vectp}, \text{valp}] = \text{eig}(C)$ . Nos devuelve dos matrices. La primera son los vectores propios y la segunda es una matriz diagonal (valores propios).

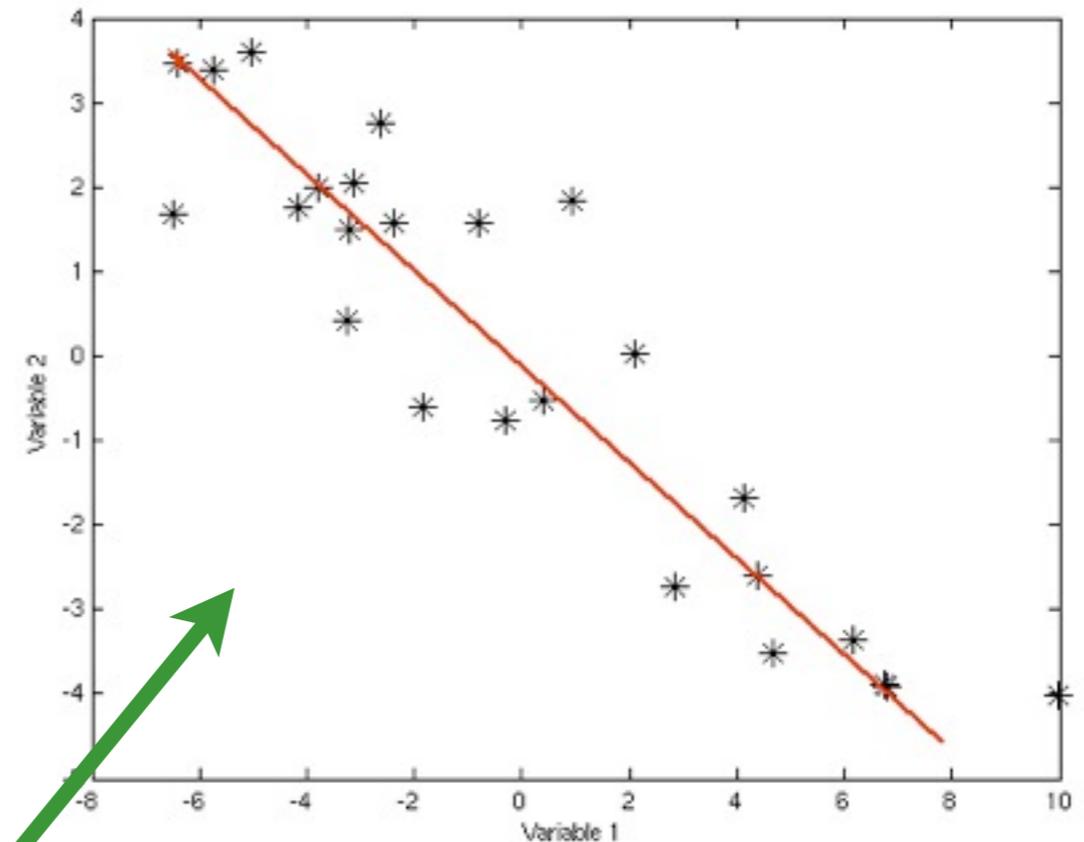
Vectores propios

-0,46	-0,88
-0,88	0,46

Valores propios

0,76	0
0	26,3

La dirección de mayor variabilidad es la dirección  $[-0,88, 0,46] = \text{VP}$



El último paso es “proyectar” todos los datos sobre vector mediante

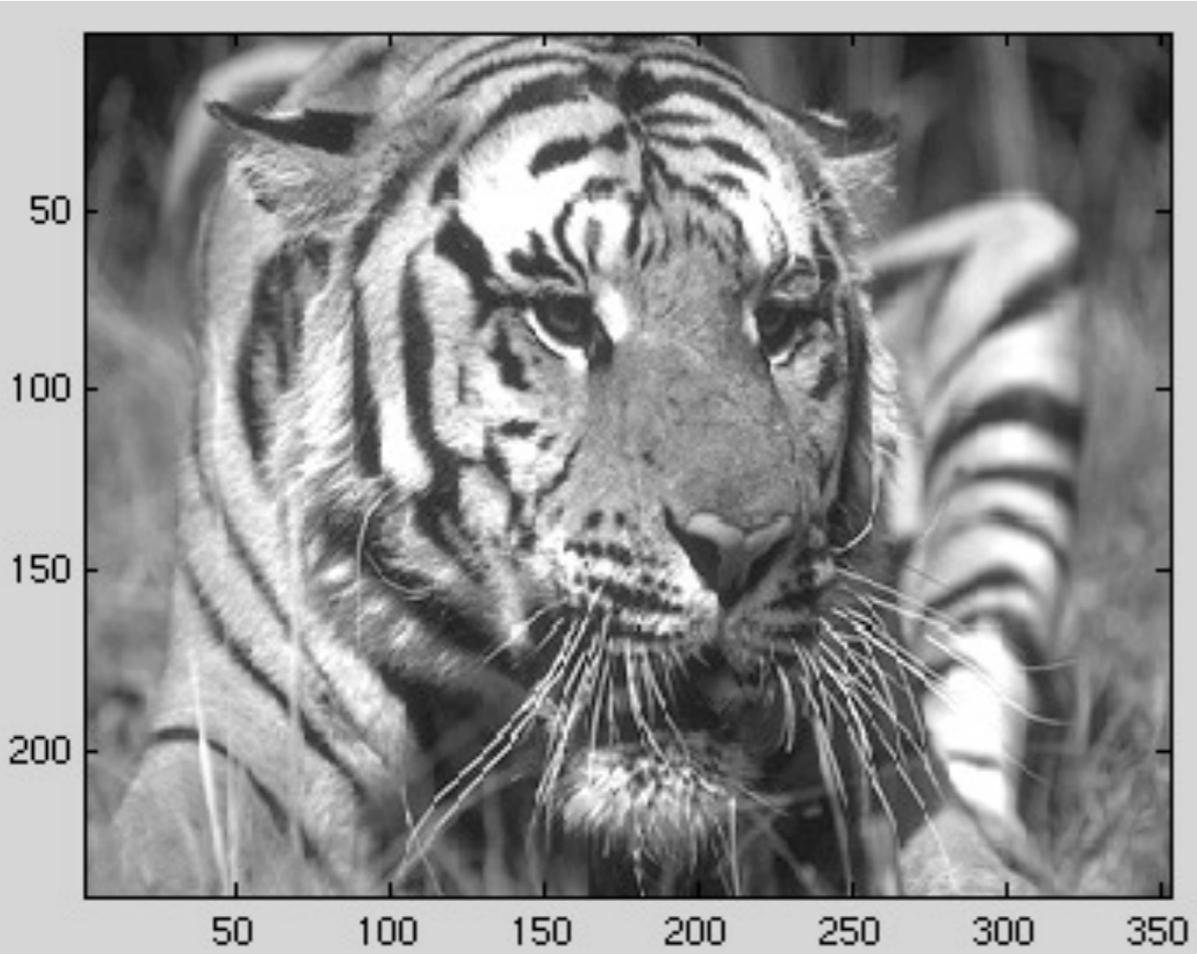
$$m = u_k \cdot \text{vp} = p_{1k} \cdot \text{vp}_1 + p_{2k} \cdot \text{vp}_2$$

Esa m es la coordenada sobre la dirección escogida.

v1	v2	m
-4,15	1,76	4,46
4,68	-3,5	-5,74
-6,4	3,4	7,2

# Ejemplo de PCA (II).

En este ejemplo aplicamos la PCA a un problema de compresión de imágenes. La imagen que tenemos tiene un tamaño de 240x352. Dicha imagen la dividimos en ventanas de 8x8 píxeles. Dicha ventana la transformamos en un vector de 64 componentes. Seguidamente calculamos las dos cantidades que necesitamos que son: el valor medio de cada componente del vector así como la matriz de autocorrelación de los vectores que se tienen. Seguidamente se diagonaliza la matriz (es una instrucción en Matlab). Se calculan los vectores con los mayores valores propios asociados y se proyectan todos los vectores en esos vectores. Cada nuevo vector queda representado por esas componentes.



$$w_k = \begin{bmatrix} p_{1k} & p_{2k} & \dots & p_{8k} \\ p_{9k} & \dots & \dots & p_{16k} \\ \dots & \dots & \dots & \dots \\ p_{57k} & \dots & \dots & p_{64k} \end{bmatrix}$$

$$v_k = [p_{1k} \quad \dots \quad \dots \quad p_{64k}]$$



$$M = \frac{1}{N} \cdot \sum_{k=1}^N v_k \quad C = \frac{1}{N} \cdot \sum_{i=1}^N (v_i - M)^t \cdot (v_i - M)$$

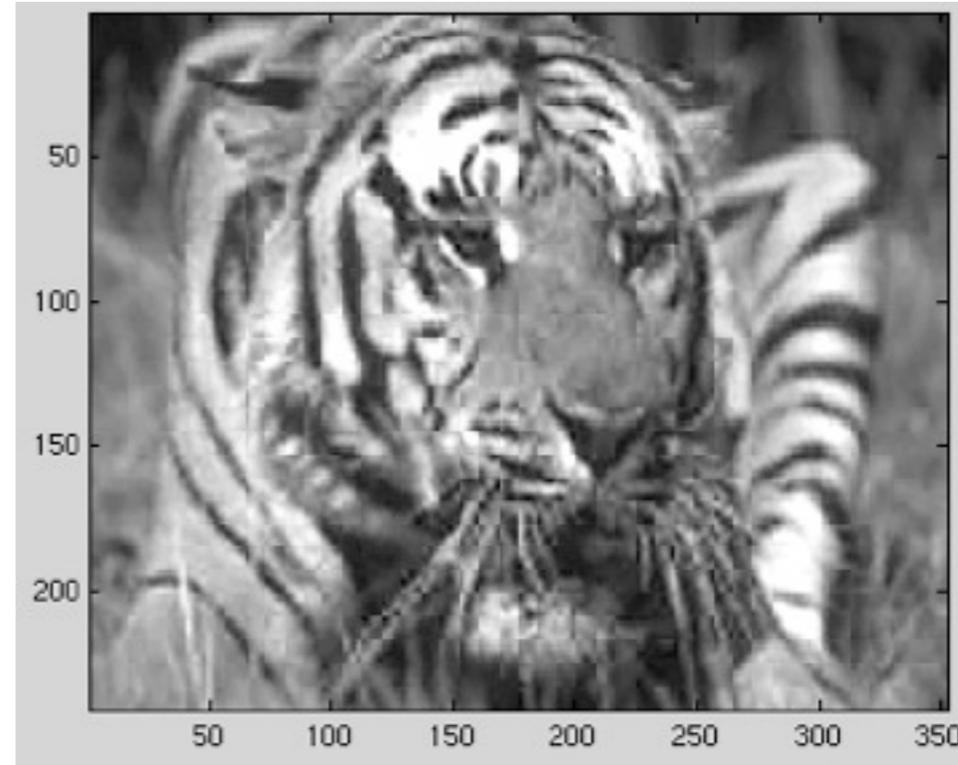
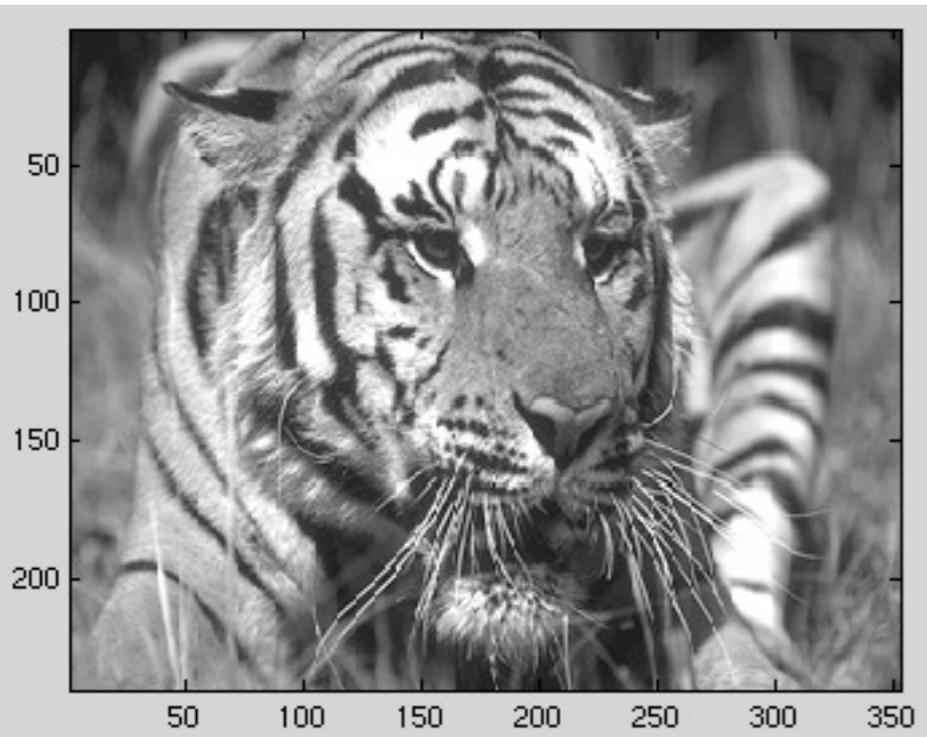
$$U = [u_1 \quad \dots \quad \dots \quad u_L]$$

$$a_k^i = u_k \cdot v_i = \sum_{s=1}^N u_{is} \cdot v_{is}$$

$$v_{approx_i} = \sum_{m=1}^L a_m^i \cdot u_m$$

**CLAVE L < 64**

# Ejemplo de PCA (II).



8:1



16:1



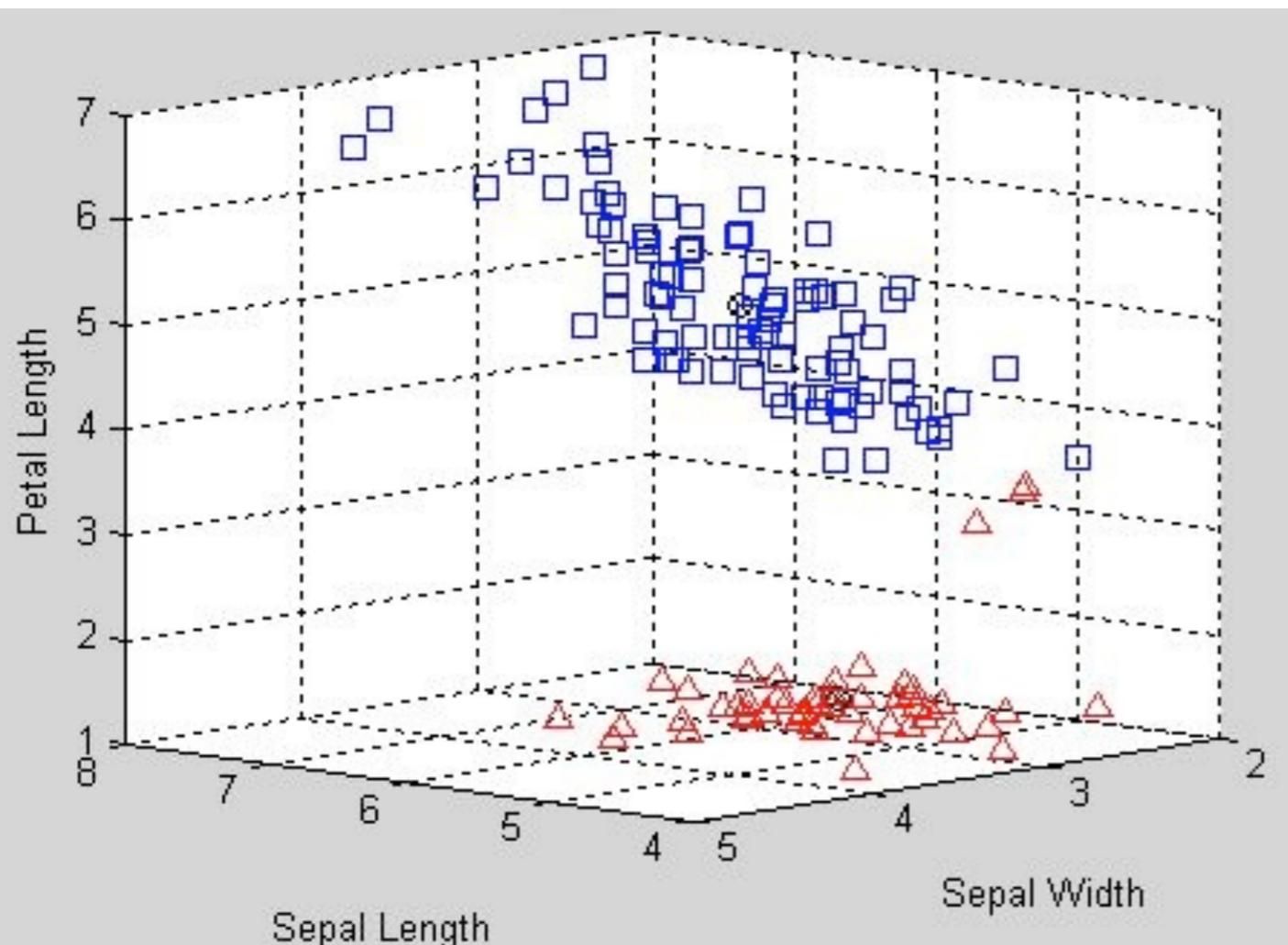
32:1

# Agrupamiento de datos (clustering).

En las transparencias anteriores hemos intentado reducir el número de variables a considerar en un modelo. Ahora el objetivo es doble: reducir el número de patrones a considerar y caracterizar dichos patrones mediante una serie de prototipos. El conjunto de técnicas de obtención de dichos prototipos se conoce como algoritmos de clustering.

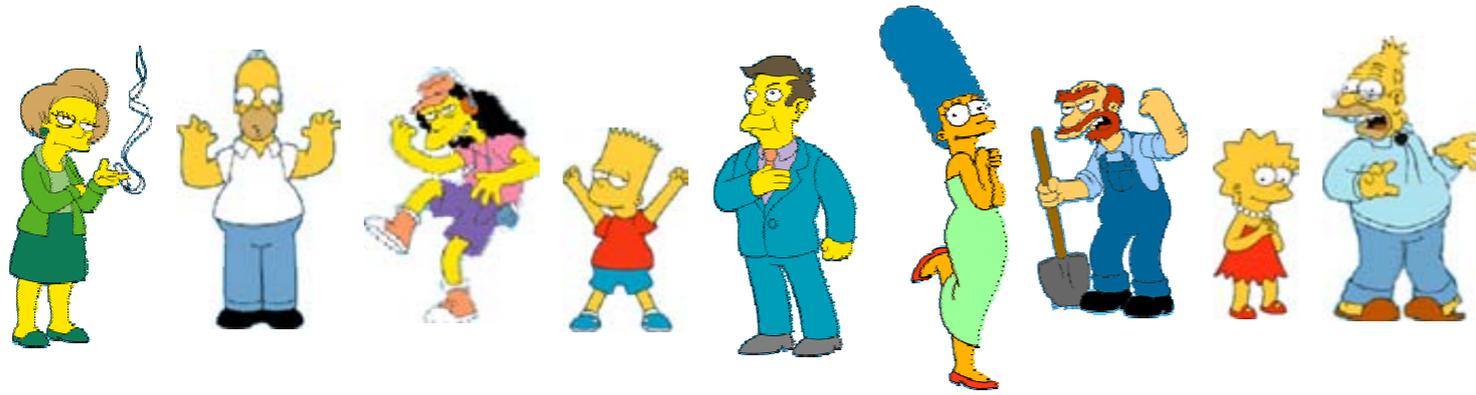
Las ventajas de determinar prototipos son varias: a) podemos conocer las diferentes tendencias que presentan los datos analizados, b) podemos codificar los datos en función de la distancia a los prototipos (menos información para almacenar) o, en el caso extremo, hacer corresponder los datos a sus prototipos, c) podemos tener un procedimiento de clasificación muy rápido (por cercanía) si este es el problema a resolver y d) podemos detectar outliers.

El problema principal que tienen la mayoría de estos algoritmos es que el número de prototipos hay que fijarlo al principio del algoritmo y ese número no se conoce a priori por lo que hay que aplicar un procedimiento de prueba y error

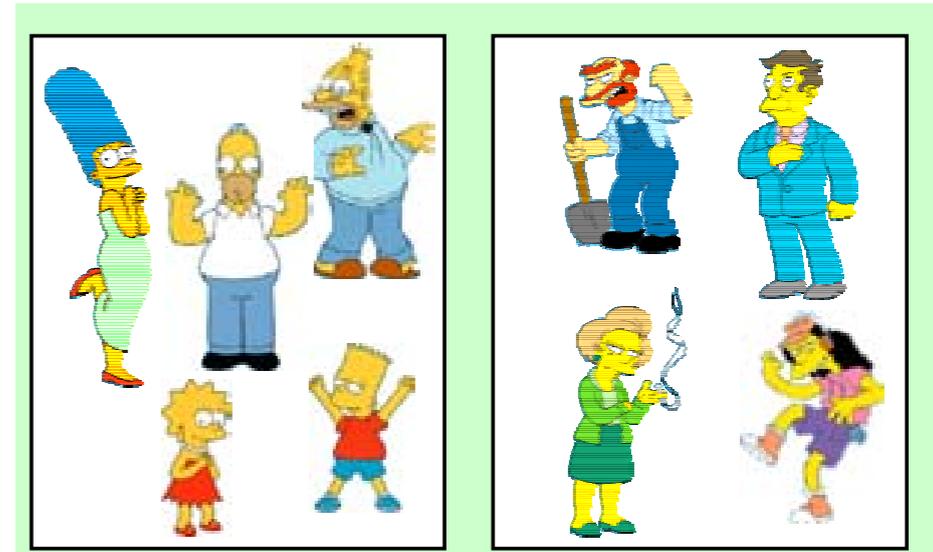
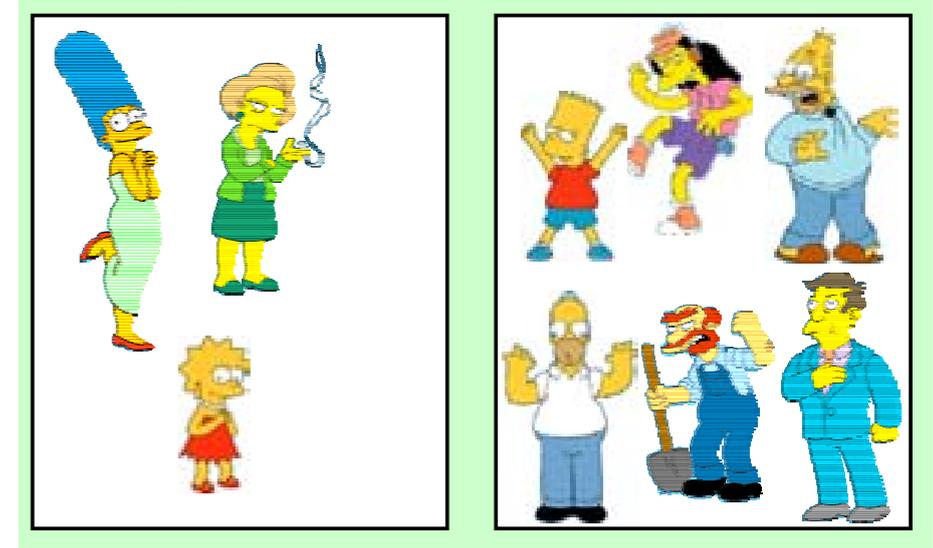


# Agrupamiento de datos (clustering).

El punto de partida de los algoritmos de clustering es el concepto de parecido entre dos patrones; ¿cómo podemos indicar que dos patrones se parecen?



Lo primero a tener en cuenta en las siguientes transparencias es que, dependiendo de la medida considerada de parecido (de ahora en adelante similitud) podremos llegar a diferentes clusters.



Otra cuestión importante a tener en cuenta es si la medida de similitud tiene sentido de cara a la obtención de prototipos dentro de nuestros datos

# Clustering

En nuestro caso los patrones quedarán definidos por el valor que toman las diferentes variables que forman dichos patrones. Es decir, cada uno de los patrones puede ser visto como un vector N-dimensional; ¿cómo puedo determinar el parecido entre dos vectores?, la respuesta se encuentra en la distancia que hay entre dichos vectores

$$d_1(A, B) = \sum_{k=1}^N |A_k - B_k|$$

$$d_2(A, B) = \sqrt{\sum_{k=1}^N [A_k - B_k]^2}$$

$$d_p(A, B) = \left[ \sum_{k=1}^N |A_k - B_k|^p \right]^{\frac{1}{p}}$$

Las distancias comentadas son las más utilizadas, sin embargo podemos usar cualquier medida que cumpla las siguientes condiciones (hay muchas!!!).

Nonnegativity:  $D(\mathbf{x}, \mathbf{y}) \geq 0$ .

Reflexivity:  $D(\mathbf{x}, \mathbf{y}) = 0$  if and only if  $\mathbf{x} = \mathbf{y}$ .

Symmetry:  $D(\mathbf{x}, \mathbf{y}) = D(\mathbf{y}, \mathbf{x})$ .

Triangle inequality:  $D(\mathbf{x}, \mathbf{y}) + D(\mathbf{y}, \mathbf{z}) \geq D(\mathbf{x}, \mathbf{z})$

En la expresión de las diferentes distancias queda reflejada la importancia de la normalización de las variables antes de aplicar los algoritmos de clustering. Si existen diferencias numéricas importantes entre las variables estos algoritmos no funcionarán.

# Clustering.

Existen diferentes tipos de algoritmos de clustering los cuales siguen, en su mayoría, los pasos comentados en la siguiente tabla.

0. Inicialización aleatoria de los prototipos
1. Calculo de los factores de proximidad
2. Actualización de los prototipos.
3. Si la diferencia entre los nuevos y antiguos prototipos es mayor que un umbral ir al paso 1.
4. FIN

Los dos algoritmos más extendidos son los que se conocen como HCM (Hard Clustering Mean) y FCM (Fuzzy Clustering Mean). La diferencia entre los dos se encuentra en el cálculo de los factores de proximidad. En el HCM este factor sólo puede valer 0/1 y en el FCM puede estar en el intervalo [0 1].

## HCM

$$u_{ij} = \begin{cases} 1 & \text{si } \| \mathbf{x}_j - c_i \|^2 \leq \| \mathbf{x}_j - c_k \|^2 \quad k \neq i, \\ 0 & \text{otro caso} \end{cases}$$

## FCM

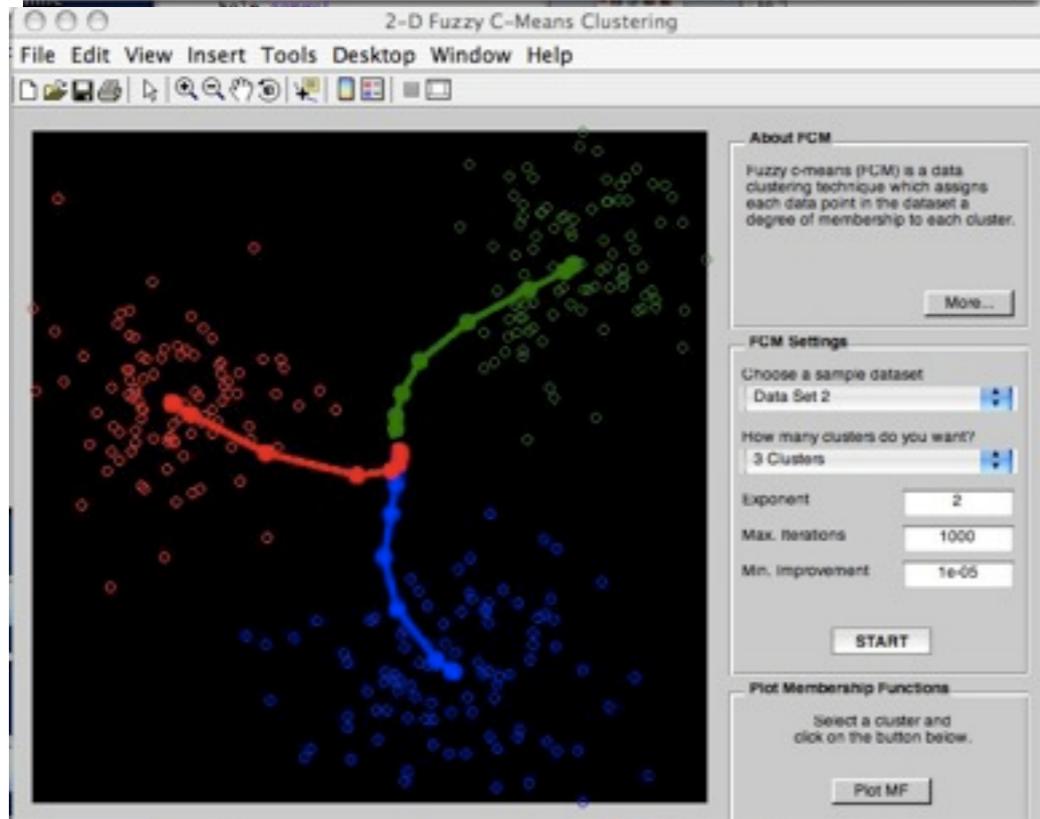
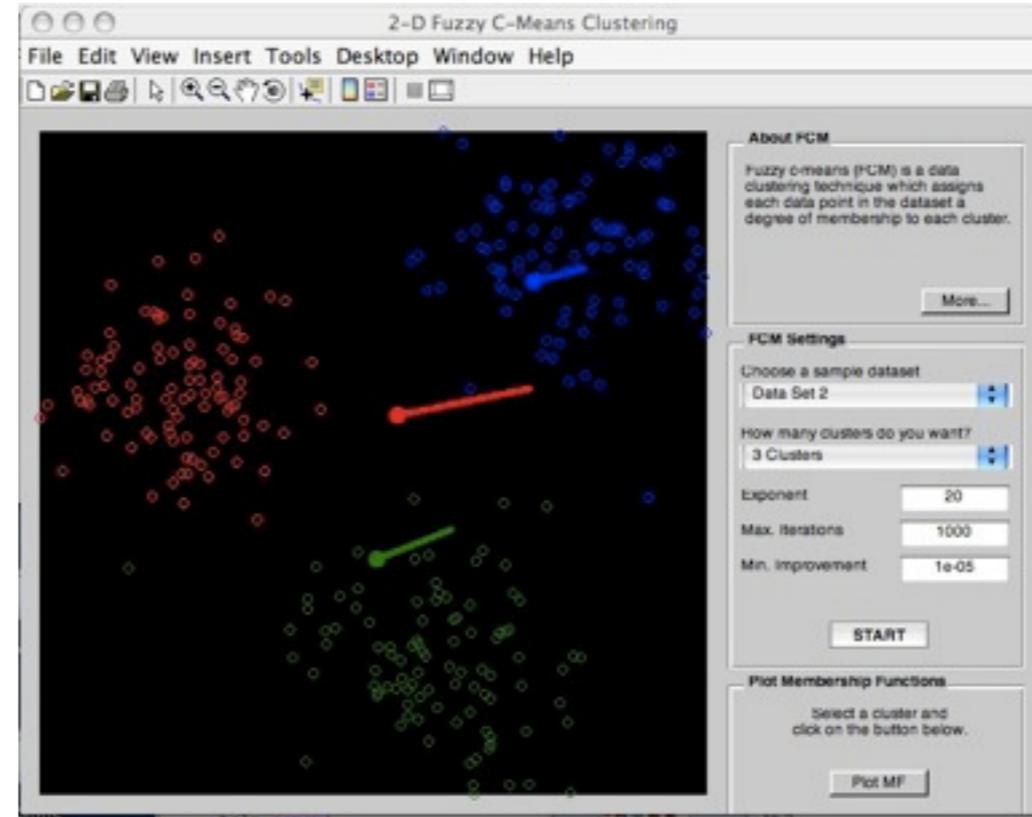
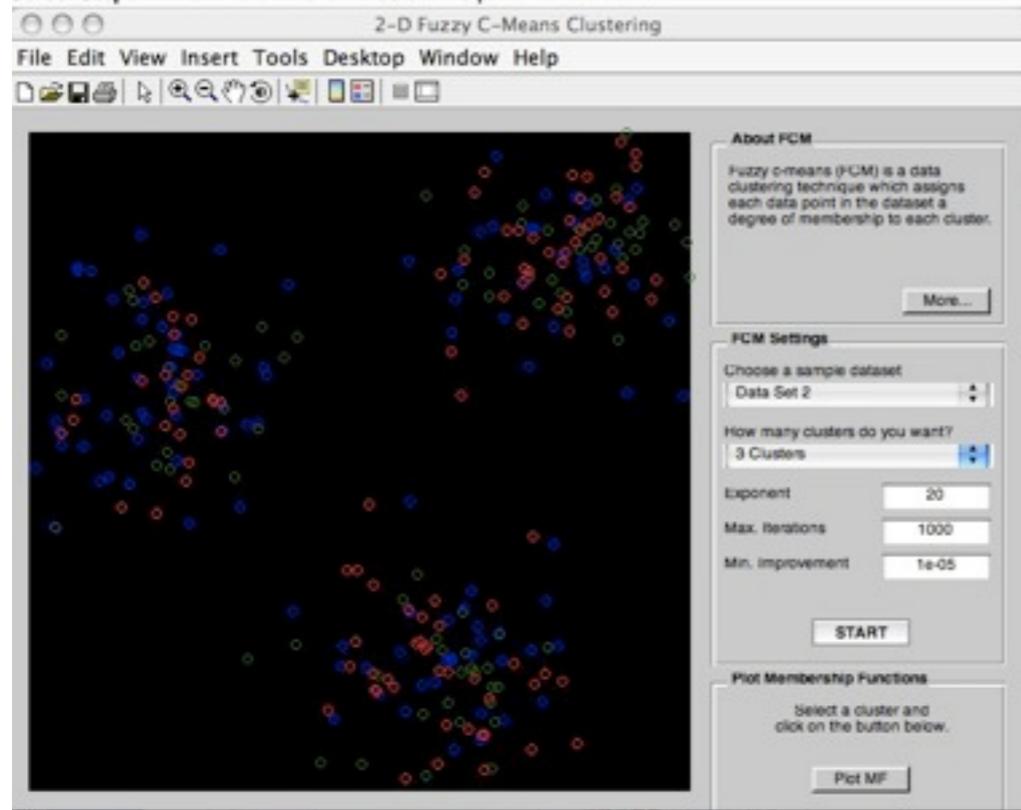
$$u_{ij} = \frac{1}{\sum_{k=1}^c \left( \frac{d_{ij}}{d_{kj}} \right)^{2/(m-1)}} \quad d_{ji} = \| c_j - x_i \|^2$$

**m es un factor > 1**

La actualización de los prototipos es la misma para los dos algoritmos.

$$c_i = \frac{\sum_{j=1}^n u_{ij}^m x_j}{\sum_{j=1}^n u_{ij}^m}$$

# Clustering. Demo



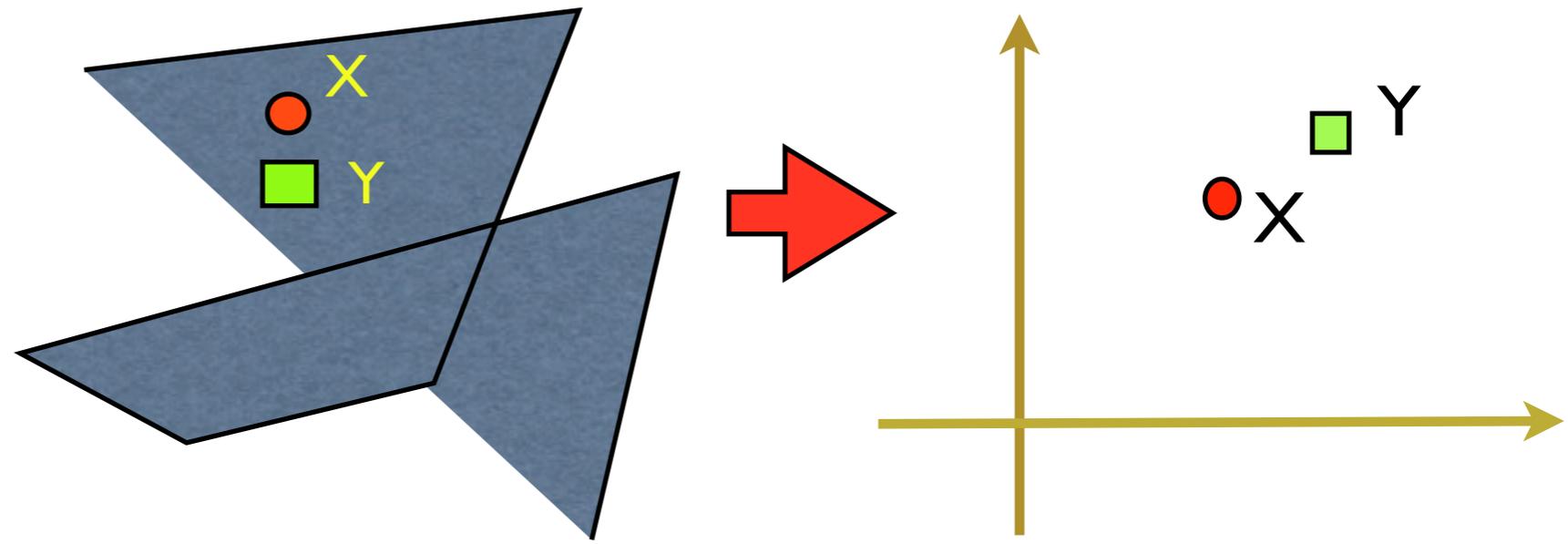
# Clustering. Mapas autoorganizados.

Esta estructura fue desarrollada por Teuvo Kohonen y su principal uso es la visualización de datos con un alta dimensionalidad.

La idea clave es mantener las relaciones de vecindad que se dan en el espacio original N-dimensional a otro espacio de 3, 2 o 1 dimensión (esto espacios sí lo podemos ver!!!).

$$X = \begin{bmatrix} x_1 \\ \dots \\ x_N \end{bmatrix}; Y = \begin{bmatrix} y_1 \\ \dots \\ y_N \end{bmatrix}$$

$$X \cong Y$$



ESPACIO N-DIMENSIONAL

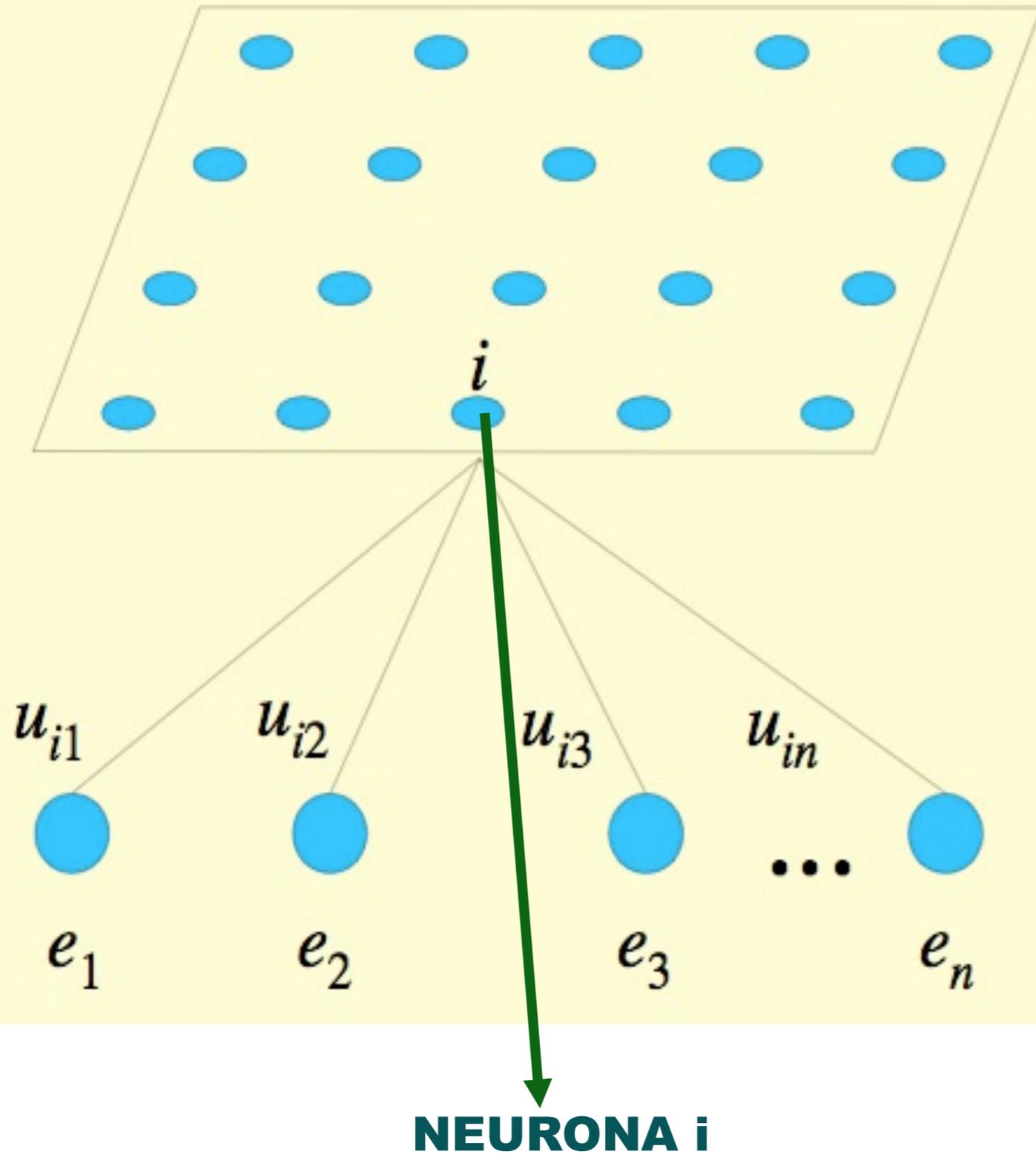
Es un modelo matemático *neuronal*, su punto de partida fue el modelo de memoria desarrollado por Kohonen en la década de los 70. Seguidamente se hablará de **neuronas** que, para nosotros serán los elementos elementales de cálculo de nuestra estructura. Así pues ahora habrá que diferenciar entre los datos de entrada, lo que denominaremos neuronas y lo que llamaremos **pesos sinápticos** que serán coeficientes asociados a cada una de las neuronas.

# Mapas autoorganizados (II)

Las neuronas se disponen en una estructura bidimensional. Es una representación en 2-D de las relaciones de cercanía en el espacio original de los patrones que se tienen. Ahora el clustering se realiza en esa estructura.

Cada neurona tiene un vector de coeficientes que la representa (pesos sinápticos). Tiene tantas componentes como el vector de entrada.

El problema es determinar los coeficientes que resuelven el problema planteado; para ello existen una gran cantidad de algoritmos iterativos (se tiene un problema de optimización no lineal).



# Mapas Autoorganizados (III)

El primer paso es determinar la arquitectura y los pesos sinápticos del mapa autoorganizado. En el primer caso existe una relación de partida entre el número de neuronas, el número de variables y la dimensión del mapa. En el segundo caso se puede aplicar uno de los numerosos algoritmos iterativos que existen. El paquete informático MATLAB dispone de una librería que crea y ajusta estos mapas autoorganizados.

## Algoritmo de aprendizaje "en palabras"

Inicializamos pesos

Escogemos un vector de entrada

Comparamos el vector de entrada con todos los del mapa.

El vector neuronal que más se parezca lo actualizamos para que, si vuelve a entrar un vector parecido a la entrada actual, se active dicha neurona de una forma más fuerte.

Hay que actualizar los vecinos de la neurona ganadora para que esos vecinos se activen con vectores de entrada parecidos.

Vuelta al paso 2.

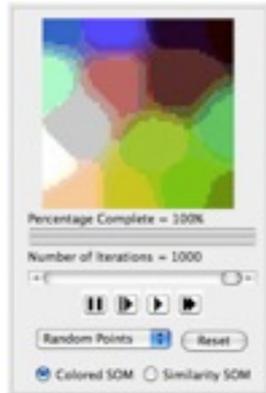
**Comparación;** dados 2 vectores existen dos maneras "naturales" de comprobar su parecido; si tienen diferente norma se utiliza la distancia entre ellos y si tienen la misma norma se puede usar el producto escalar.

**Actualización;** lo que se busca es que si se tuviera otro vector de entrada parecido al actual el vector de la neurona ganadora que es el más parecido (menor distancia) debería parecerse más.... simplemente hay que "acercar" al vector de entrada.

**Vecindad;** no solamente se actualiza el ganador sino que hay que actualizar las neuronas más cercanas; la actualización anterior hay que extenderla a los vecinos aparece la función vecindad.

$$w_{ij}(n+1) = w_{ij}(n) + \alpha(n)h(n)(x_j - w_{ij})$$

# DEMOSTRACIÓN DEL SOM.

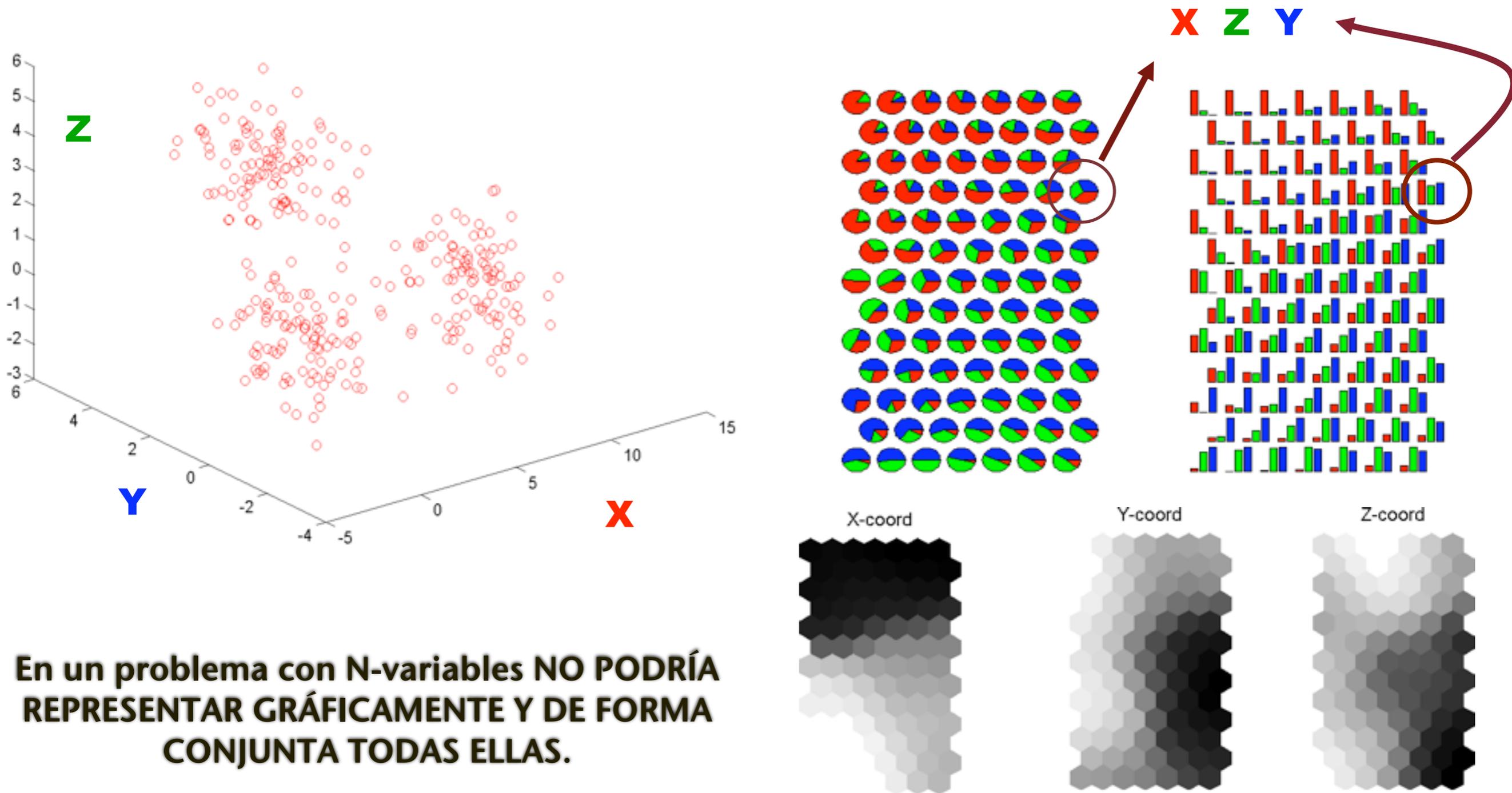


Percentage Complete = 0%  
Number of Iterations = 400  
Random Points [button]  
Reset [button]  
 Colored SOM  Similarity SOM



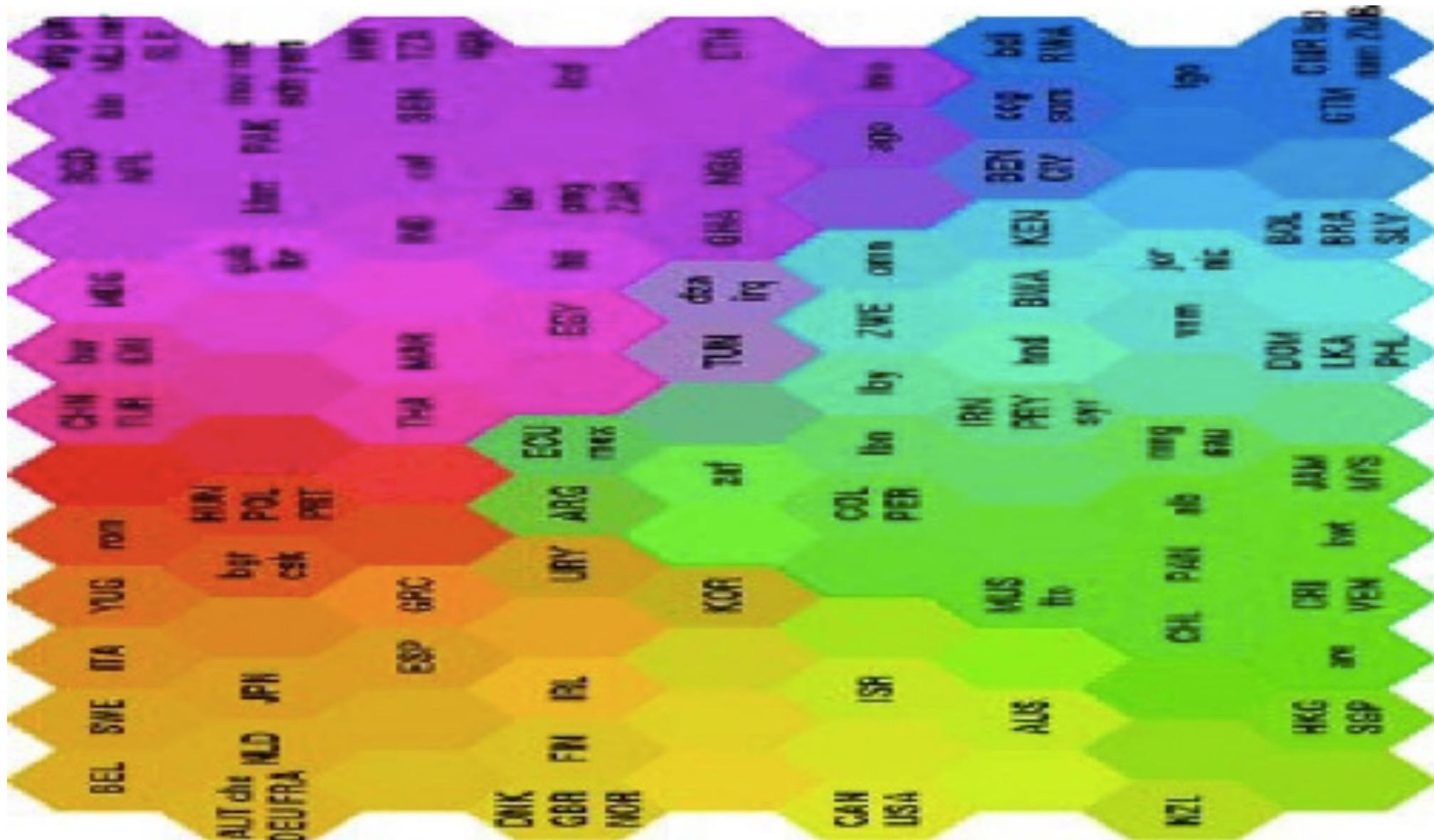
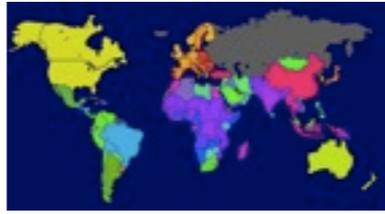
# Mapas autoorganizados (IV)

Se tienen una serie de datos que tienen tres variables, x,y ,z. En la serie que se tiene se aprecian diferentes grupos. Los mapas autoorganizados pueden representar las relaciones que existen entre las variables de una forma visual (a diferencia de los algoritmos de clustering clásicos).

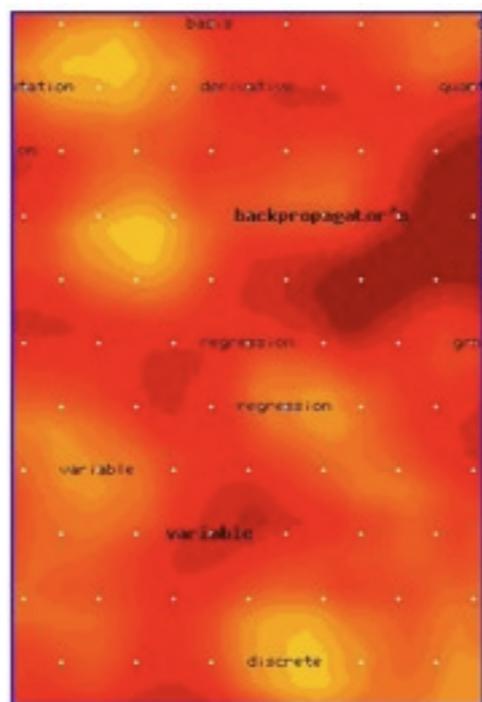
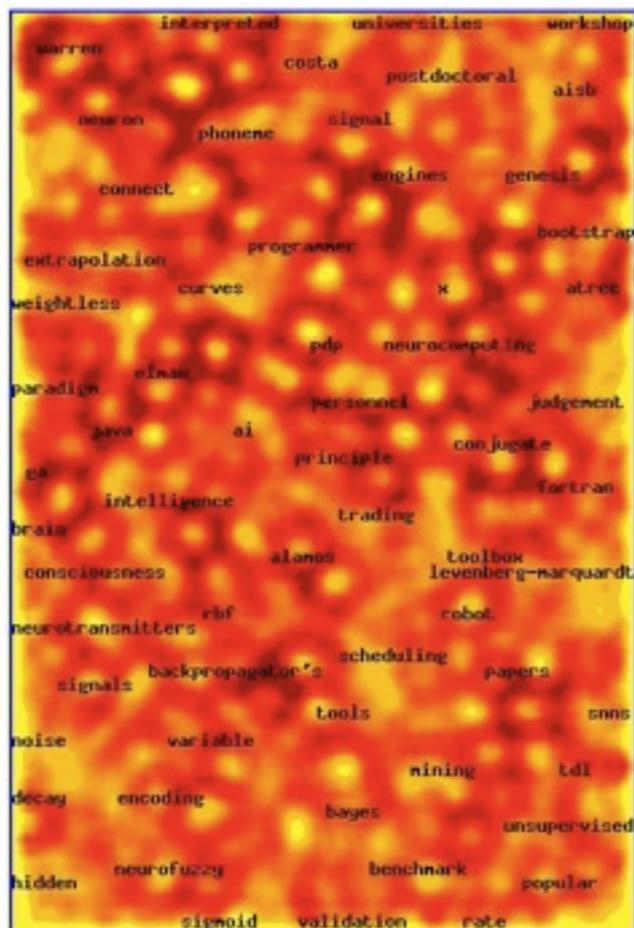


**En un problema con N-variables NO PODRÍA REPRESENTAR GRÁFICAMENTE Y DE FORMA CONJUNTA TODAS ELLAS.**

# Aplicación del SOM en la descripción de países.

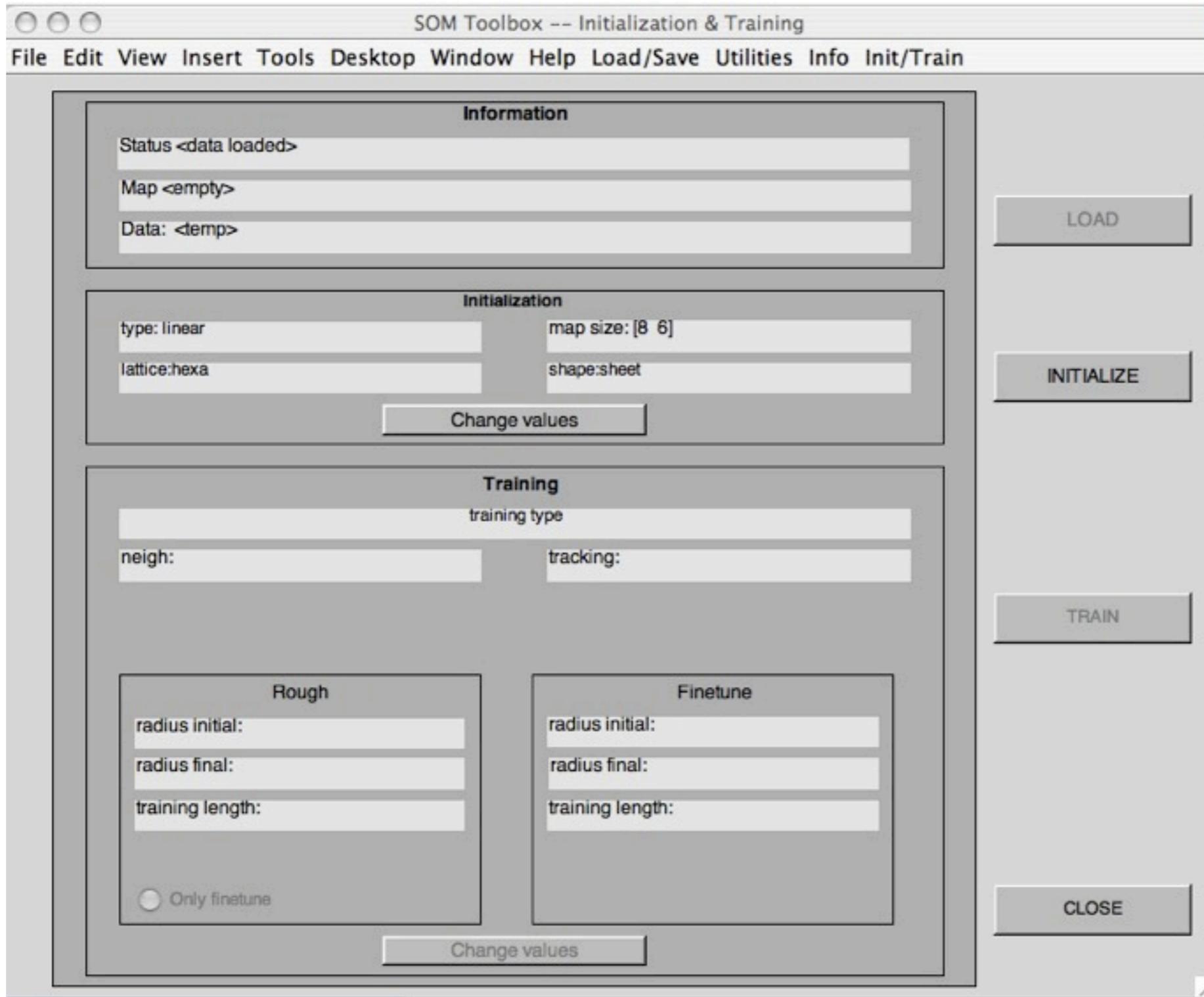


**De cada país se escogen una serie de indicadores económicos y sociales, para hacer un SOM.**



- [Advantages of NN over Stat. Regression?](#) ♦ Juan Matosas, 14 Dec 1995, Lines: 9.
- [Re: Advantages of NN over Stat. Regression?](#) ♦ Michael A. Bonnice, Fri, 15 Dec 1995, Lines: 32.
- [Re: Advantages of NN over Stat. Regression?](#) ♦ TiedNBound, 19 Dec 1995, Lines: 5.
- [Re: Advantages of NN over Stat. Regression?](#) ♦ Bill Armstrong, 19 Dec 1995, Lines: 19.
- [Re: The book I want to buy](#) ♦ David Livingstone, Wed, 20 Dec 1995, Lines: 170.
- [Re: Advantages of NN over Stat. Regression?](#) ♦ Warren Sarle, Wed, 20 Dec 1995, Lines: 19.

# SOM TOOLBOX



Librería de  
Matlab, de  
libre  
distribución.

# Mapas Autoorganizados (V)

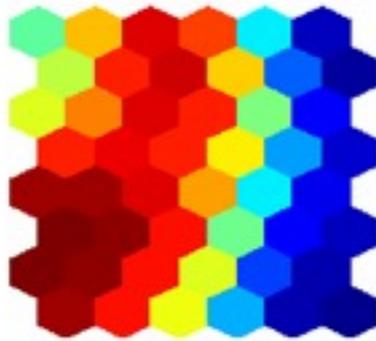
La representación conjunta de las variables en un mapa autoorganizado permite la visualización de las características comunes que se tienen dentro de nuestro conjunto de datos. Este tipo de visualización no se tiene en los algoritmos de agrupamiento tradicionales.

NOROESTE

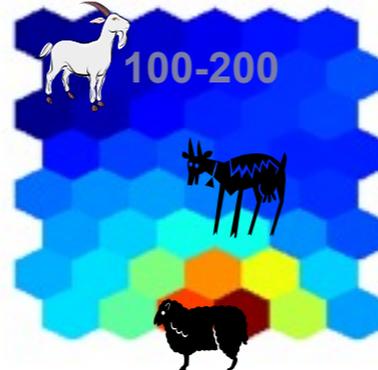
ALTIPLANO  
CARTAGENA

MULA  
GUADALENTIN  
SEGURA

Area de Estudio



Tamaño Rebaño



Alim. Reposición



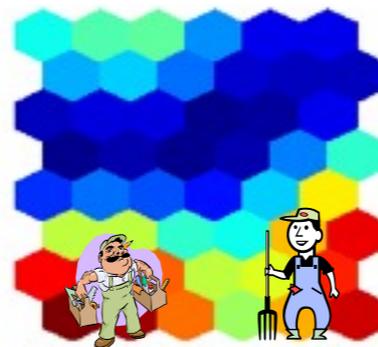
Alim. Ordeño



Alim. Gestación



Ases. Nutrición



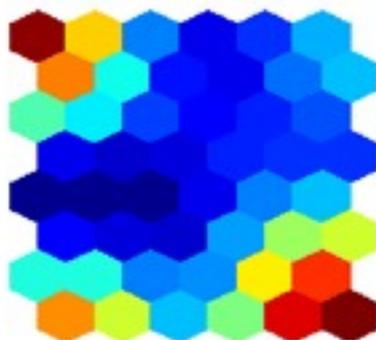
Tipo Alim.



Alim. Pico Lactac.



Condición Corporal



Ingresos Leche



Uso de Subprod.



**AZUL= valor bajo o negativo**  
**ROJO = valor alto o positivo**

## Un ejemplo en veterinaria



VNIVERSITAT ID VALÈNCIA

# MASTER DE INGENIERÍA BIOMÉDICA.

## Métodos de ayuda al diagnóstico clínico.

### Tema 3: Preprocesado de los datos.