

# Índice

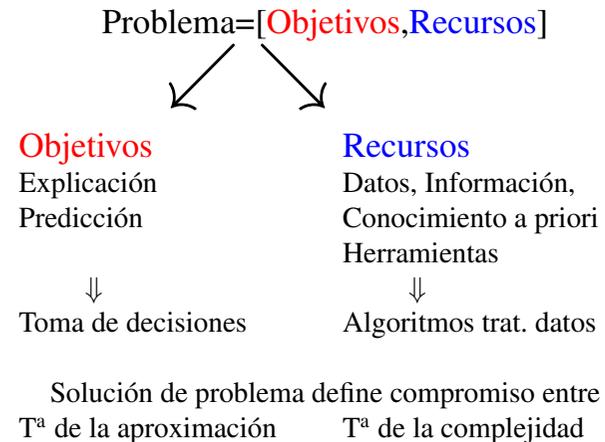
## Aprendizaje estadístico. Fundamentos y aplicaciones

Pedro J. Zufiria

- 1 Problemas y datos
  - Caracterización general de un Problema
  - Tratamiento de datos
- 2 Análisis estadístico
  - Construcción de modelos
  - Métodos de dependencia e interdependencia
    - Reducción de dimensionalidad
    - Estimación de densidades. Clustering
- 3 Panorámica del aprendizaje estadístico y tipos de problemas

# Caracterización general de un Problema

- 1 Problemas y datos
  - Caracterización general de un Problema
  - Tratamiento de datos
- 2 Análisis estadístico
  - Construcción de modelos
  - Métodos de dependencia e interdependencia
    - Reducción de dimensionalidad
    - Estimación de densidades. Clustering
- 3 Panorámica del aprendizaje estadístico y tipos de problemas



# Datos, información, conocimiento, sabiduría

# Ejemplo 1

## Datos

- Mera representación simbólica  
No tienen sentido semántico → no transmiten mensaje

## Información

- Datos organizados y/o procesados (+ contexto)  
Tienen un significado → transmiten un mensaje  
*Quién/Qué/Dónde/Cómo/Cuántos/Cuándo/Por qué*

## Conocimiento

- Valor añadido que permite tomar decisiones  
Saber *cómo* resolver un problema

## Sabiduría

- Saber qué problemas debo resolver  
*Por qué hacerlo*

## Datos

Registro de una persona: secuencia de ceros y unos (incluso codificados)

## Información

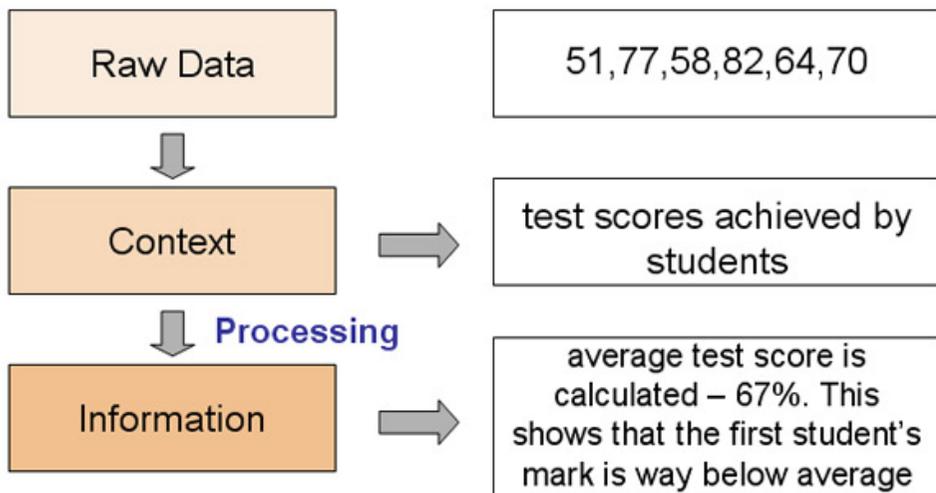
datos + contexto: nombre, edad, teléfono, dirección

Podemos completar información: país (a partir de dirección), sexo (a partir del nombre)

Podemos ordenar, comparar, etc. con datos de otros clientes (y obtener datos derivados)

# Ejemplo 2

# Ejemplo 3



Dato: medida de un sensor en un sistema

Información: añadimos contexto (significado y, tal vez, más datos):

- magnitud física que representa (presión de una válvula)
- valores habituales (está por encima de rango habitual)
- serie temporal (creciente)

# Conocimiento

# Ejemplo 3 (continuación)

Lo que te permite tomar decisiones a partir de la información. Se obtiene de

- Experiencia directa
- Transferencia (de conocimiento) desde otras personas: directa, manuales, procedimientos
- Entrenamiento (camino híbrido entre experiencia y transferencia)

Saber *cómo* resolver un problema.

## Conocimiento:

Nos permite tomar una decisión: abrir la válvula, modificar parámetros del sistema, parar el sistema, evacuar el lugar. . .

*Cómo* resolver un problema.

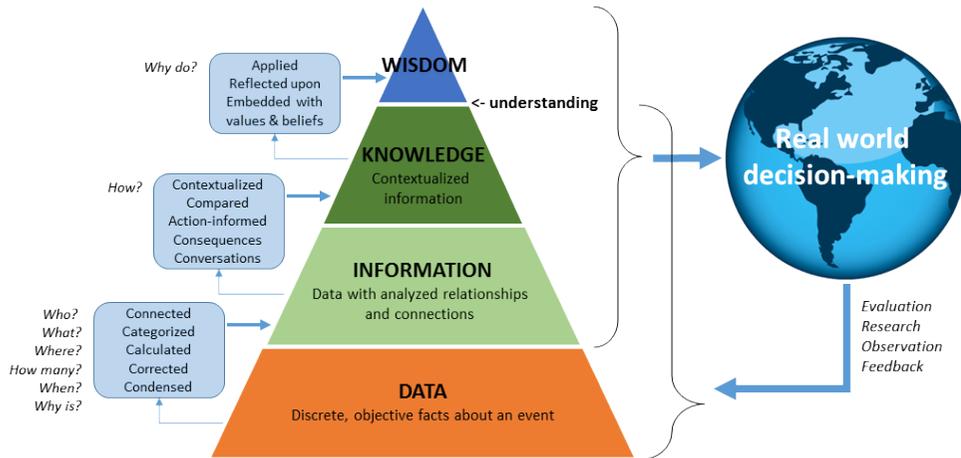


# Sabiduría

# Escenarios de captura de datos

Saber qué problemas son los que debo resolver.

*Por qué hacerlo*



- Investigador no interviene sobre variables:  
Estudio observacional (retrospectivo, prospectivo...)

Datos recopilados sin objetivo concreto



Problema a resolver no definido



Objetivos se plantean a posteriori (ajustados a los datos disponibles)



(pseudo)problema definido a medida

- Investigador interviene sobre variables:  
Diseño de experimento



# Fases o etapas del tratamiento de datos

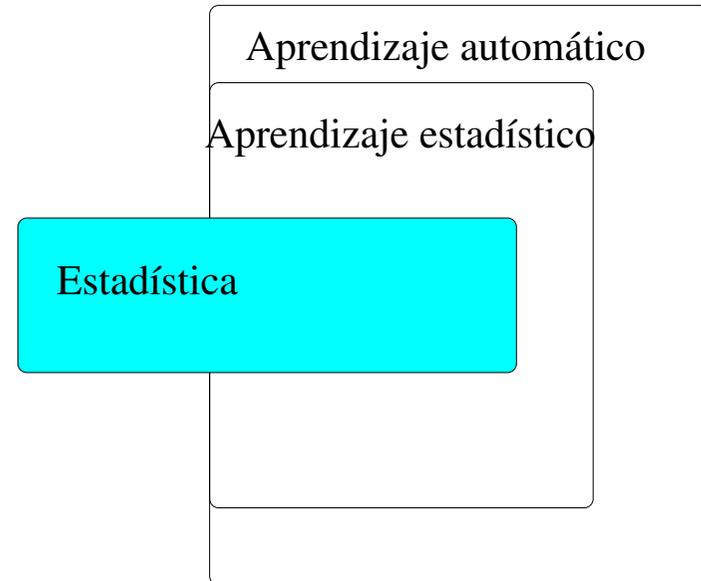
- Tratamiento preliminar (trabajo “sucio”)
  - Validación: datos limpios, correctos, útiles
  - Organización
- Análisis: resumen (agregación), exploración, modelado
- Presentación: visualización...

# Fases o etapas del tratamiento de datos

- Tratamiento preliminar (trabajo “sucio”)
  - Validación: datos limpios, correctos, útiles
  - Organización
- Análisis: resumen (agregación), exploración, modelado ←
- Presentación: visualización...

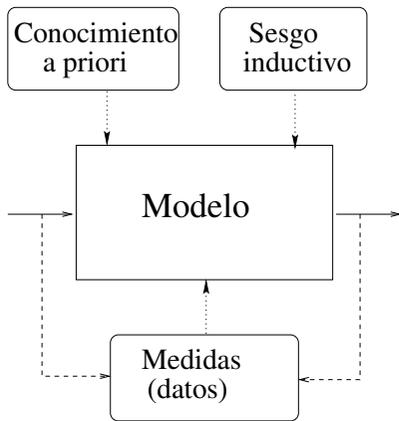
# Aprendizaje automático/estadístico y estadística

- 1 Problemas y datos
  - Caracterización general de un Problema
  - Tratamiento de datos
- 2 Análisis estadístico
  - Construcción de modelos
  - Métodos de dependencia e interdependencia
    - Reducción de dimensionalidad
    - Estimación de densidades. Clustering
- 3 Panorámica del aprendizaje estadístico y tipos de problemas



# Construcción de un modelo

# Utilidad y tipos de modelos



Disponibilidad de **modelo**  $\Rightarrow$  permite razonamiento deductivo  
( $\approx$ ley)

Utilidad fundamental:

- descripción  $\Leftarrow$  ciencia
- predicción  $\Leftarrow$  ingeniería

(Algunos) tipos de modelos

- Mecanicista puro: basado en leyes deterministas conocidas
- Probabilista: recoge incertidumbre, aprovechando cierto conocimiento
- Empíricos: combinan elementos deterministas y probabilistas  
Fuertemente basados en datos (“poco” conocimiento a priori)

# Utilidad y tipos de modelos

# Métodos de dependencia

Disponibilidad de **modelo**  $\Rightarrow$  permite razonamiento deductivo  
( $\approx$ ley)

Utilidad fundamental:

- descripción  $\Leftarrow$  ciencia
- predicción  $\Leftarrow$  ingeniería

(Algunos) tipos de modelos

- Mecanicista puro: basado en leyes deterministas conocidas
- Probabilista: recoge incertidumbre, aprovechando cierto conocimiento
- Empíricos: combinan elementos deterministas y probabilistas  
Fuertemente basados en datos (“poco” conocimiento a priori)

## Asumimos dependencia (y prelación) entre variables

Modelos empíricos:

- conocemos que existe relación funcional  $y = f(x)$  entre variables;
- desconocemos la forma de  $f$
- Hipótesis estructural: linealidad, polinomial, etc.
- $\epsilon$ : término de error (variable aleatoria) caracterizable a partir de **datos**

$$y = \tilde{f}(x) + \epsilon$$

Dos tipos (según naturaleza de  $Y$ ):

- Regresión
- Clasificación

# Métodos de dependencia

# Modelo de regresión (método de mínimos cuadrados)

## Asumimos dependencia (y prelación) entre variables

datos:  $(x_1, y_1), \dots, (x_n, y_n)$

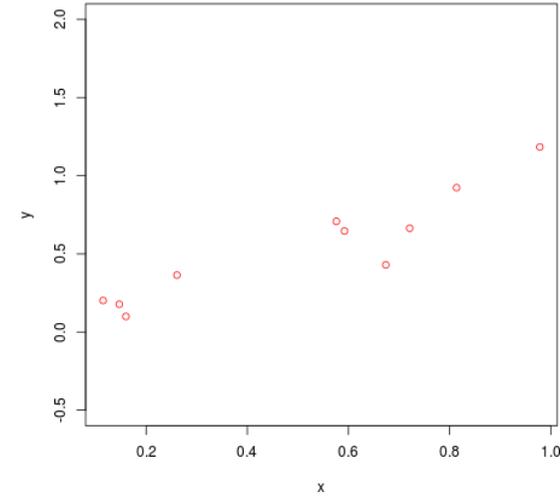
Modelos empíricos:

- conocemos que existe relación funcional  $y = f(x)$  entre variables;
- desconocemos la forma de  $f$
- Hipótesis estructural: linealidad, polinomial, etc.
- $\epsilon$ : término de error (variable aleatoria caracterizable a partir de **datos**)

$$y = \tilde{f}(x) + \epsilon$$

Dos tipos (según naturaleza de  $Y$ ):

- **Regresión**  $\leftarrow$
- Clasificación

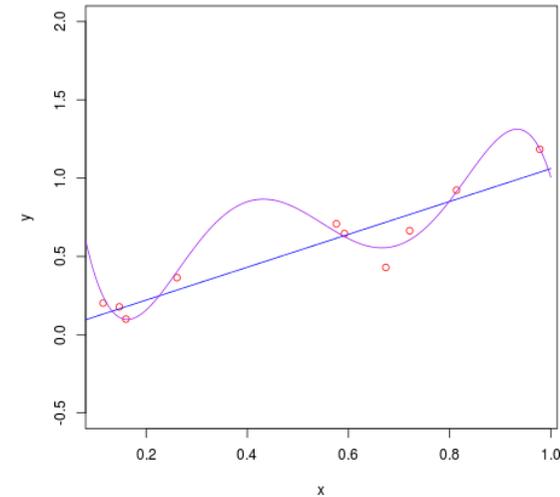
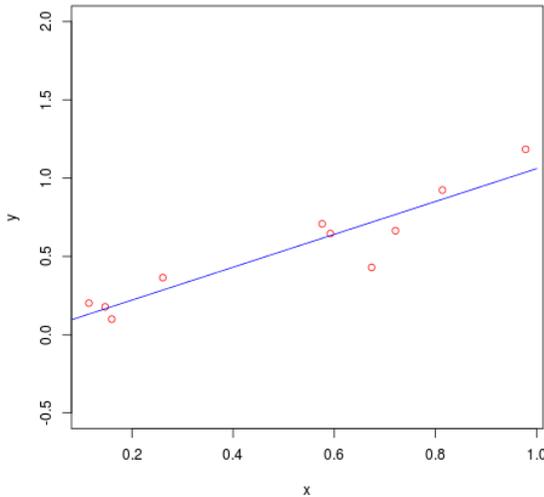


# Modelo de regresión (método de mínimos cuadrados)

# Modelo de regresión (método de mínimos cuadrados)

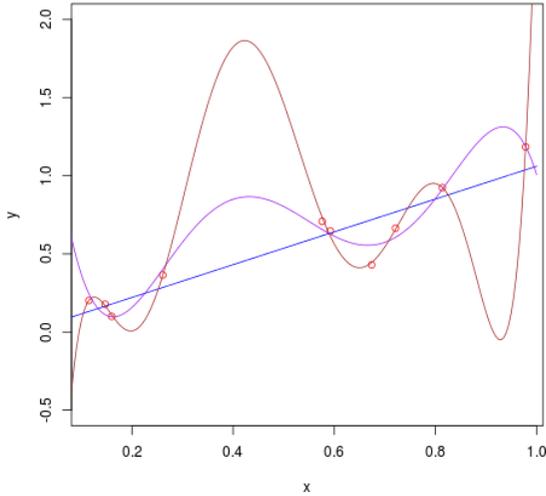
$$y = \beta_0 + \beta_1x + \epsilon$$

$$y = \beta_0 + \beta_1x + \beta_2x^2 + \beta_3x^3 + \beta_4x^4 + \beta_5x^5 + \epsilon$$



# Modelo de regresión (método de mínimos cuadrados)

$$y = \beta_0 + \beta_1x + \beta_2x^2 + \beta_3x^3 + \beta_4x^4 + \beta_5x^5 + \beta_6x^6 + \beta_7x^7 + \epsilon$$



# Sobre la elección de modelo

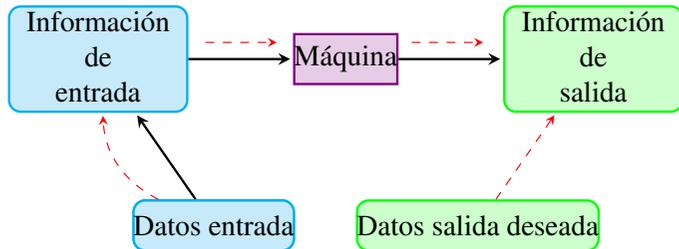
Capacidad de **generalización** del modelo

Compromiso entre **sesgo** (underfitting) versus **varianza** (overfitting)

- elección adecuada de espacio de funciones (LM, PMC, RBF, procesos gaussianos, etc.)
  - técnicas de regularización (meta-parámetros)
  - gestión inteligente de los datos. Subconjuntos:
    - Training
    - Validation: estima error de generalización. Aplicabilidad (herencia histórica/fases/disponibilidad de nuevos datos):
      - evaluación de modelo  $\equiv$  test
      - selección de modelo (ajuste-parada parámetros / meta-parámetros)  $\rightarrow$  se complementa con training
    - Test (evaluación de modelo)
- Validación cruzada (pocos datos): partición dinámica de conjuntos (evaluación/selección de modelo)
- Ensemble learning

Compromiso entre **precisión** versus **interpretabilidad** del modelo

# Métodos de dependencia



Formulación de aprendizaje automático (supervisado)  
(Fase de entrenamiento -rojo- y fase de ejecución.)

# Métodos de interdependencia

No asumimos dependencia con prelación entre variables

Buscamos detectar estructuras en los datos:

- Reducción de dimensionalidad
- Detección de grupos (variables latentes)
- Detección de anomalías

(Sucedáneos de estimación de densidades)

# Métodos de interdependencia

# Reducción de dimensionalidad para visualización de datos

No asumimos dependencia con correlación entre variables

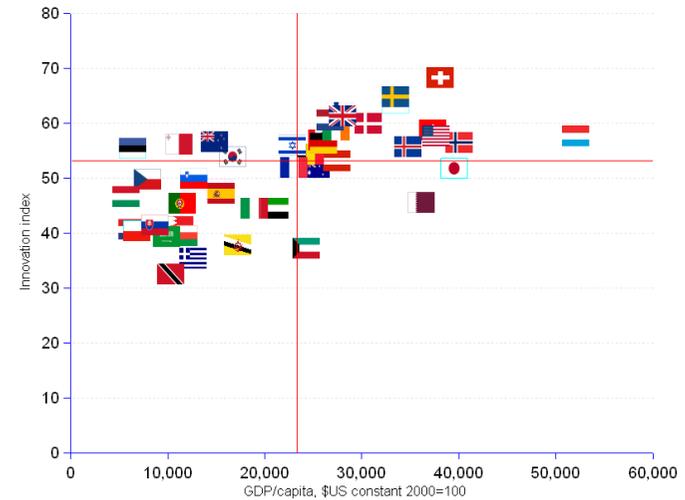
Buscamos detectar estructuras en los datos:

- Reducción de dimensionalidad
- Detección de grupos (variables latentes)  $\Leftarrow$  Algoritmos de clustering
- Detección de anomalías

(Sucedáneos de estimación de densidades)

Caso general: formulación de aprendizaje automático (no supervisado)

Ejemplo: características definitorias de países:



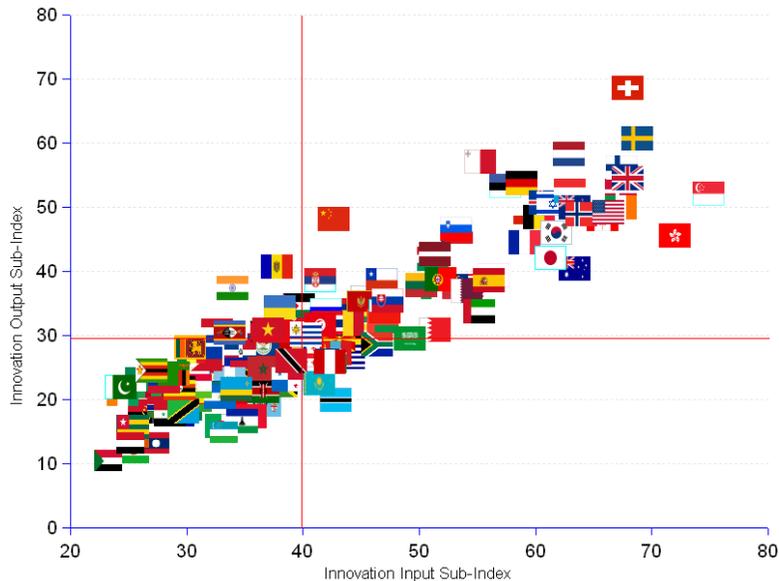
Resumimos en 2 características (sin significado preciso: relacionadas con tamaño y la actividad económica por persona)

Podemos dibujarlo

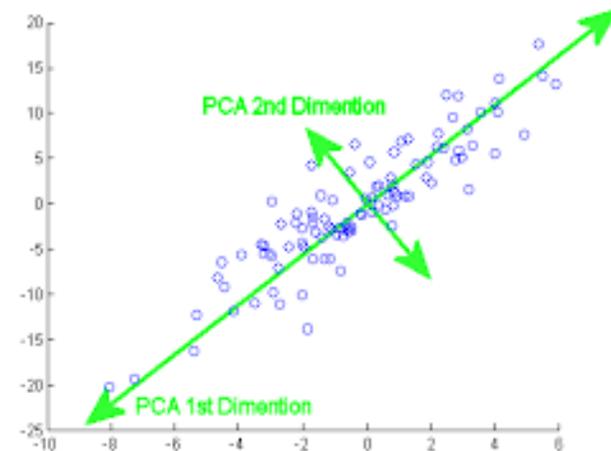


# Reducción de dimensionalidad para visualización de datos

# Análisis de Componentes Principales



Muestra  $\{(x_{11}, x_{21}), \dots, (x_{1s}, x_{2s})\}$  de variables aleatorias  $(X_1, X_2) \in \mathbb{R}^2$



Determinar esas direcciones: caracterización más simple ( $\Rightarrow$  compresión de datos)

## Aplicaciones habituales de reducción de dimensionalidad

- Compresión de datos
  - Almacenamiento, transmisión
  - Implementación computacionalmente eficiente de problema supervisado
- Visualización: con  $M = 2$  o  $M = 3$

- $K$ -means
- Clustering jerárquico

Elegimos aleatoriamente los  $K$  centroides:  $\mu_1, \mu_2, \dots, \mu_K \in \mathbb{R}^n$ .

Repetimos dos bucles:

- de 1 a  $m$  etiquetamos cada dato con el número del centroide más cercano (busco  $c^{(i)} = \arg \min_k \|x^{(i)} - \mu_k\|^2$  para  $k = 1, \dots, K$ )
- De 1 a  $K$  “movemos” cada centroide a la media de los datos asignados a su cluster  
Si un cluster no tiene asignados datos, lo eliminamos o (si queremos mantener  $K$ ) lo reinicializamos aleatoriamente

Puede ser un problema ambiguo

Lo más común es visualizar (si se puede); es siempre útil

Método del codo: representar para diferentes valores de  $K$ , mín  $J(K)$  y elegir  $K$  allá donde se halla el “codo”

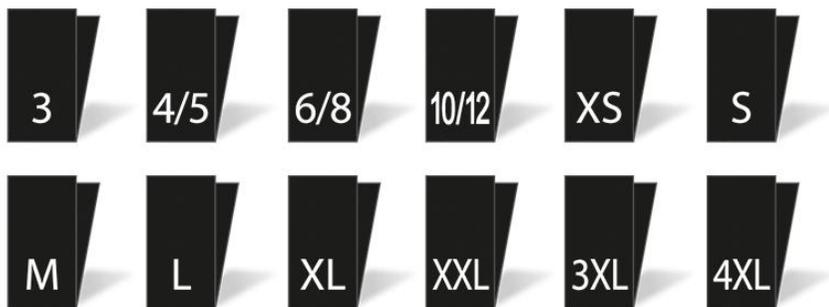
(Problema: puede no haber un “codo” claro; pero merece la pena intentarlo)

Objetivo para el que realizamos el clustering puede orientarnos:

el valor de  $K$  que más ayuda para el siguiente paso (métrica de ese paso es un referente) puede ser la buena opción

# Ejemplo

Etiquetar tamaño de camisetas a partir de tamaño de gente:

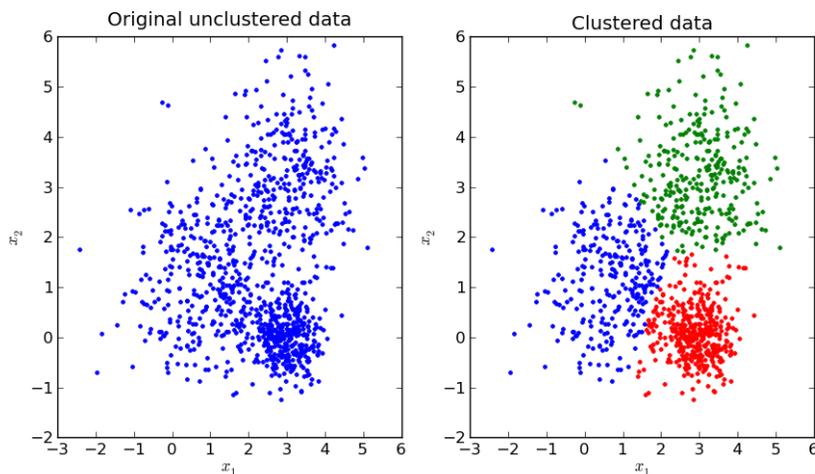


# Ejemplo

Imponemos  $K = 3$  para S, M, L.  
Los centroides nos dan las tallas



# K-means. Ejemplo



Muestras de tres distribuciones gaussianas bivariantes.

# Clustering jerárquico

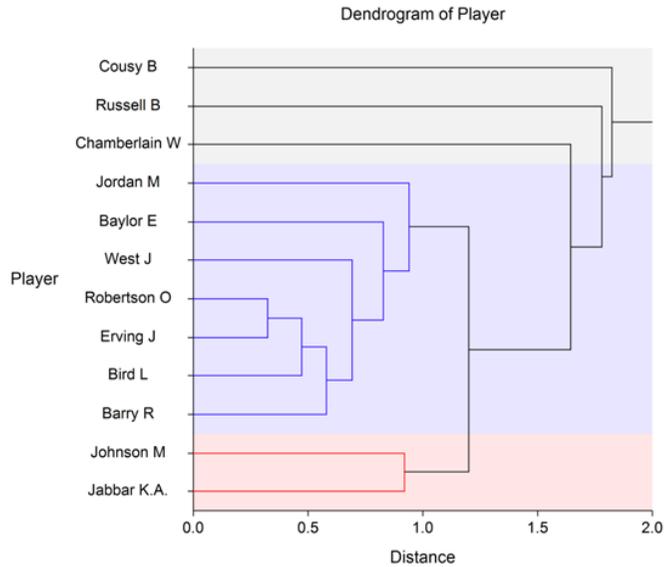
Aplicable cuando los datos no son vectores de  $\mathbb{R}^n$  (y también cuando lo son).  
Solamente necesitamos una distancia: euclídea, Minkowski, Manhattan, binaria, derivadas de medidas de similitud (correlación o coseno), edit distance, etc.

Pasos

- 1 Agrupar los dos datos más cercanos.
- 2 Sigüentes pasos: agrupar los grupos más cercanos (ya sean datos sueltos o datos agrupados).  
Clave: definir cómo se calcula la distancia entre grupos  
Hay varios métodos: complete, average, Ward, etc.
- 3 Finalmente proporciona un Dendrograma (árbol) que recoge las distancias entre grupos. Una manera de medir esas distancias es la cophenetic distance or similarity.

# Clustering jerárquico

# Mapa Topológico Auto-organizativo (MTA): fundamentos biológicos



Agrupación de jugadores según características

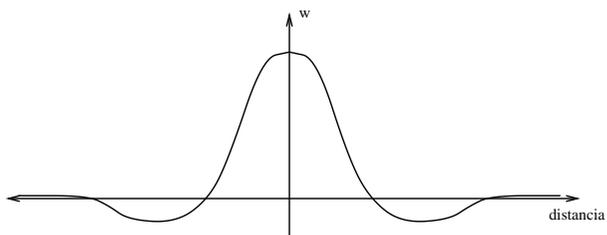
## Fundamentos biológicos (mapas computacionales):

- Procesamiento paralelo de información similar pero proveniente de diferentes regiones del espacio sensorial
- Preservación de contexto de cada pieza de información en cada etapa de procesamiento
- Existencia de **mapas**: áreas del cerebro dedicadas a tareas específicas o especializada en análisis de determinadas señales sensoriales
- Mapas contextuales como aplicaciones entre espacio sensorial y superficie cortical.

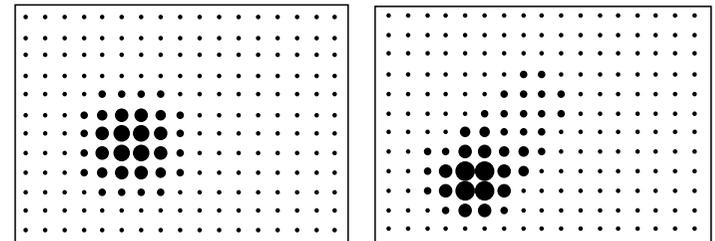
# MTA: fundamentos biológicos (II)

# MTA: fundamentos biológicos (III)

## Interacción lateral entre neuronas: ejemplo de *sombrero mexicano*



- Agrupaciones de respuestas activas: **burbujas**



- Actualización de características en la burbuja de activación

## MTA: principios de formación

Principio de formación de mapa topológico: la localización espacial de una neurona de salida en un mapa topológico corresponde a un dominio o característica de los datos de entrada

⇒ **Feature mapping models**

MTA o SOM

- Activación neuronal por **similitud** con entrada (correlación entre entrada y vector de pesos de la neurona)
- Actualización de pesos en vecindarios: aproximación del vector de pesos al de entrada.

Consecuencia: **Mapa topológicamente ordenado**

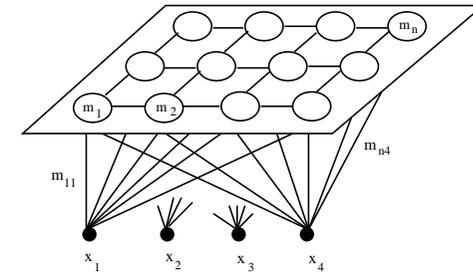
## MTA: modelo formal

### Topología

Entrada  $x = (x_1, \dots, x_K) \in \mathbb{R}^K$

Vector de Pesos  $m_i = (m_{i1}, \dots, m_{iK}) \in \mathbb{R}^K$

Ejemplo para  $K = 4$



## Dos espacios diferenciados de representación

- Espacio físico-topológico (bidimensional)
- Espacio de pesos (K-dimensional)

Desvinculados a priori.

Aspecto clave: vinculación a posteriori.

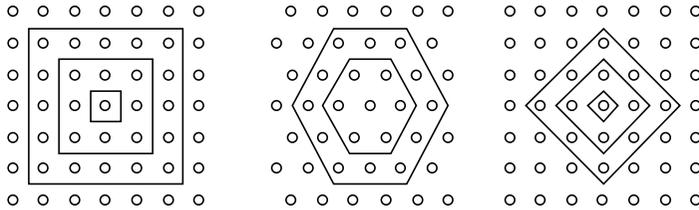
## MTA: aprendizaje

### Características del Algoritmo de Entrenamiento:

- **No supervisado**: único dato de entrenamiento  $x$   
⇒ estimación de densidades (o heurístico asociado)  
Comportamiento Auto-organizativo
- **Competitivo**: Actualización de neurona activada por la entrada (ganadora) y vecinas topológicas

# MTA: ingredientes de algoritmo

- Competición: **distancia Euclídea**
- Interacción lateral: **criterio de vecindad  $N_c(t)$**



- **Actualización:** incremento proximidad de  $m$  a  $x$

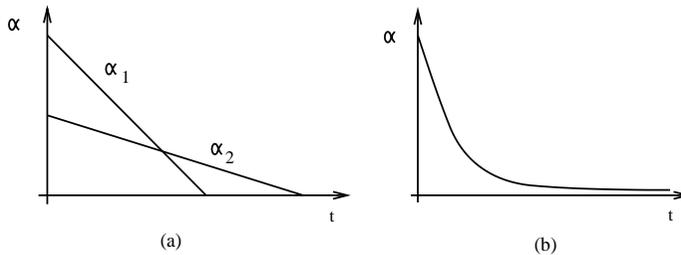
$$m_i^{new} = m_i^{old} + \alpha(x - m_i^{old}) \quad 0 < \alpha(t) < 1$$

$\alpha(t)$  decreciente en  $t$



# MTA: parámetros del algoritmo

## Ley de Ganancia.



# MTA: algoritmo

## Algoritmo

1. Inicializar  $m_i(0)$ ,  $N_c(0)$  y  $\alpha(0)$
2. Presentar una entrada  $x(t)$
3. Seleccionar  $m_c$

$$\|x - m_c\| = \min_i \{\|x - m_i\|\}$$

4. Actualizar pesos

$$m_i(t+1) = \begin{cases} m_i(t) + \alpha(t)(x(t) - m_i(t)) & \text{si } i \in N_c(t) \\ m_i(t) & \text{si } i \notin N_c(t) \end{cases}$$

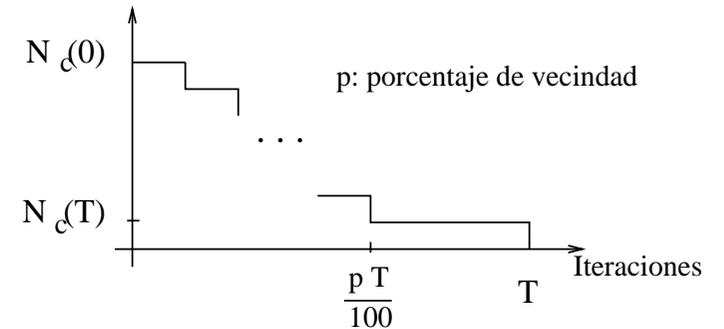
con  $0 < \alpha(t) < 1$  y decreciente en  $t$ .

5. Incrementar  $t$ . Volver a 2.



# MTA: parámetros del algoritmo

## Ley de Vecindad.



# MTA: características

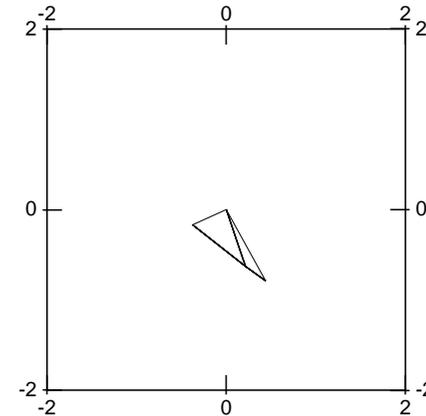
# Ejemplos de ordenación: mapa bidimensional

**Aprendizaje.** Dos fases:

- Distribución global de neuronas
- Ajuste fino

### Características del Mapa (doble o triple funcionalidad)

- 1 Define **reducción de dimensionalidad** mediante proyección de datos a espacio de baja dimensión (permite visualización)  
Topológicamente ordenado: proyecta funcionalidad de espacio multidimensional en espacio de baja dimensión  
(Recordar conceptos de topología y continuidad de funciones)
- 2 Según número de neuronas:
  - Muchas neuronas: define **aproximación a fdp** (estimación de densidades -problema tipo) de los datos
  - Pocas neuronas: define **partición** del espacio del conjunto de entrenamiento (clustering y cuantificación vectorial)

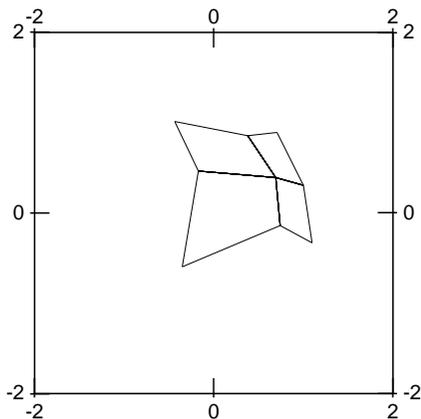


10 It.

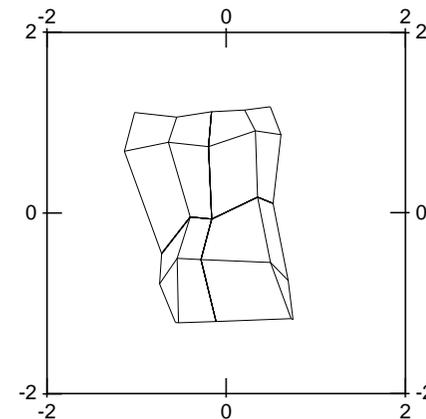


# Ejemplos de ordenación: mapa bidimensional

# Ejemplos de ordenación: mapa bidimensional



100 It.

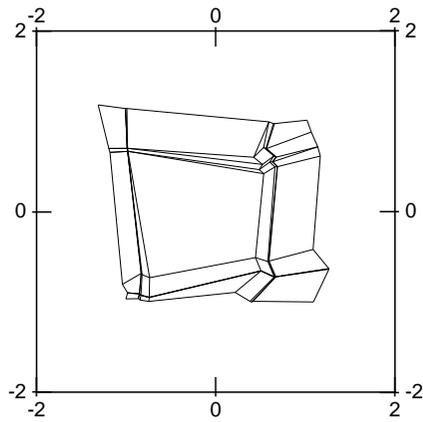


1000 It.

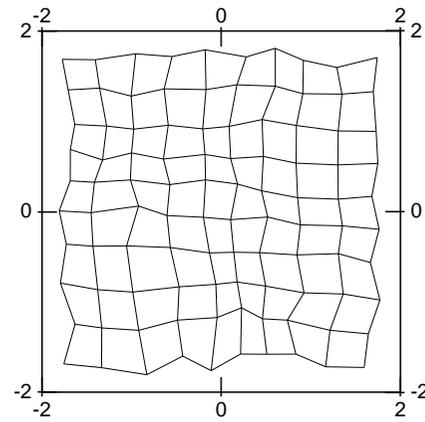


# Ejemplos de ordenación: mapa bidimensional

# Ejemplos de ordenación: mapa bidimensional



3000 lt.

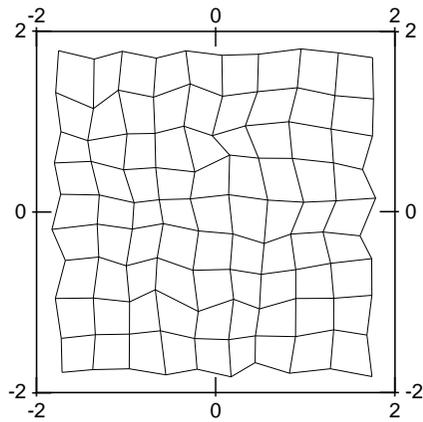


7000 lt.

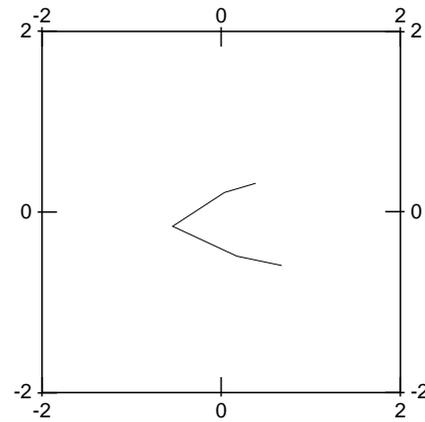


# Ejemplos de ordenación: mapa bidimensional

# Ejemplos de ordenación: mapa lineal



10000 lt.

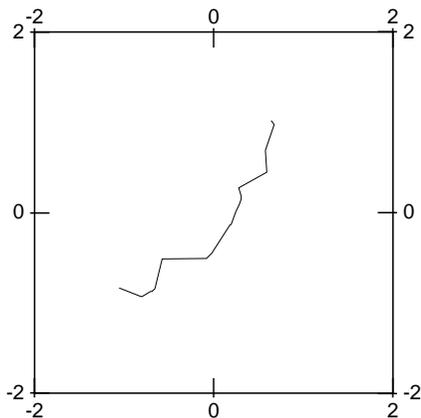


10 lt.

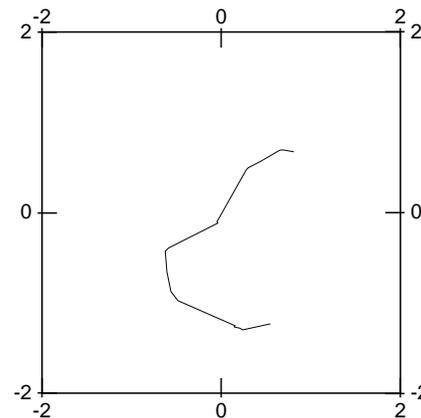


# Ejemplos de ordenación: mapa lineal

# Ejemplos de ordenación: mapa lineal



100 It.

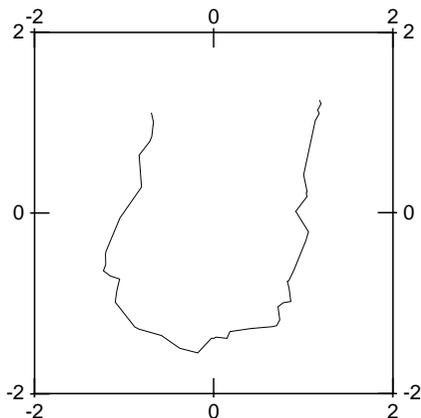


1000 It.

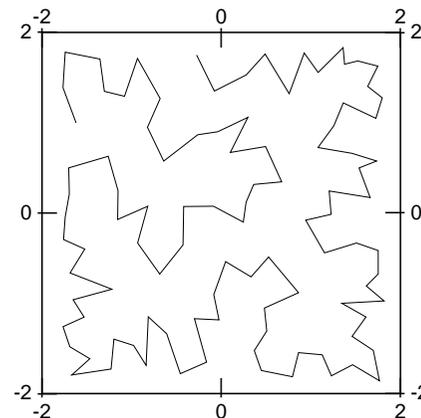


# Ejemplos de ordenación: mapa lineal

# Ejemplos de ordenación: mapa lineal



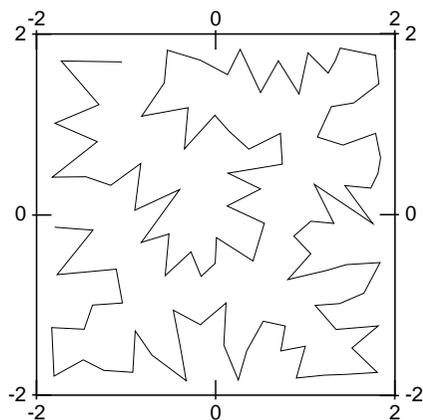
3000 It.



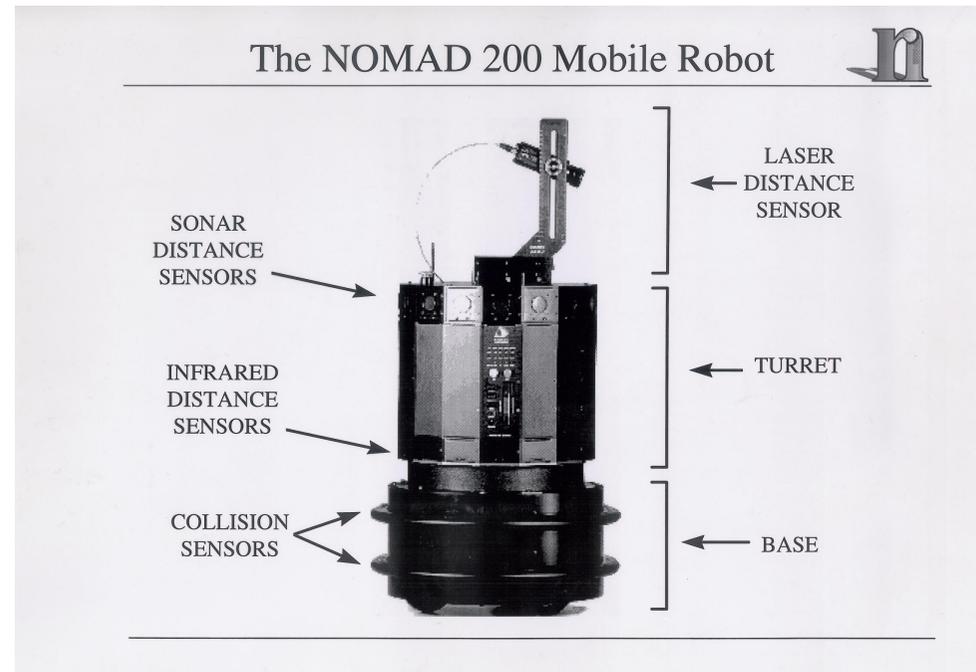
7000 It.



# Ejemplos de ordenación: mapa lineal



10000 It.



The Self-Organizing Map as a Perceptual Level for Mobile Robot Control

## Introduction



- Nomad 200 mobile robot
  - Sonar
  - IR sensors
  - Laser
  - Vision
  - Bumpers
  - Compass



- Real-world unstructured environment

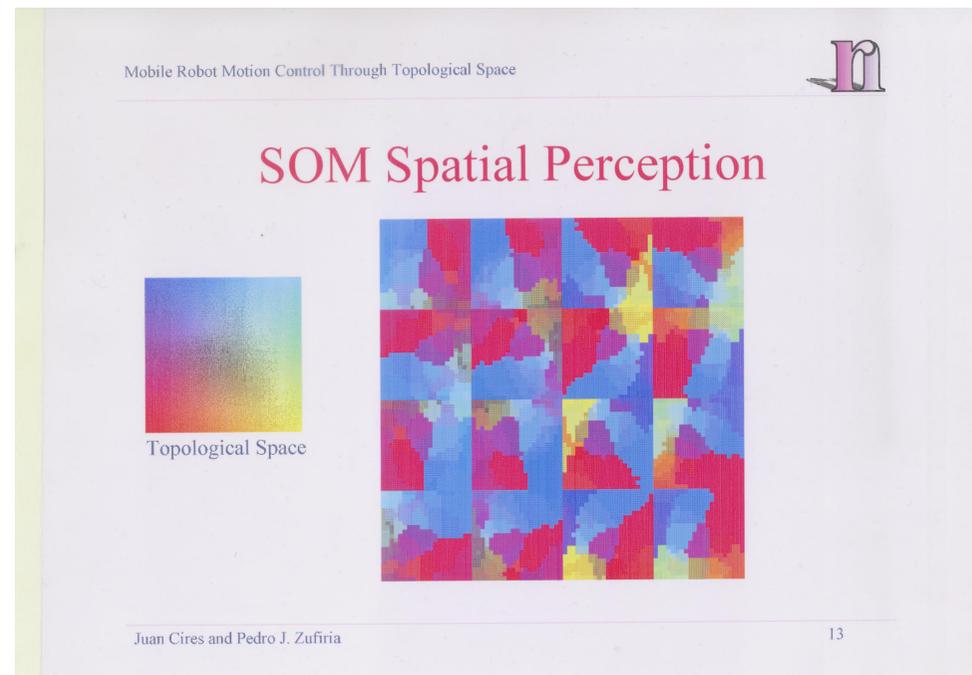
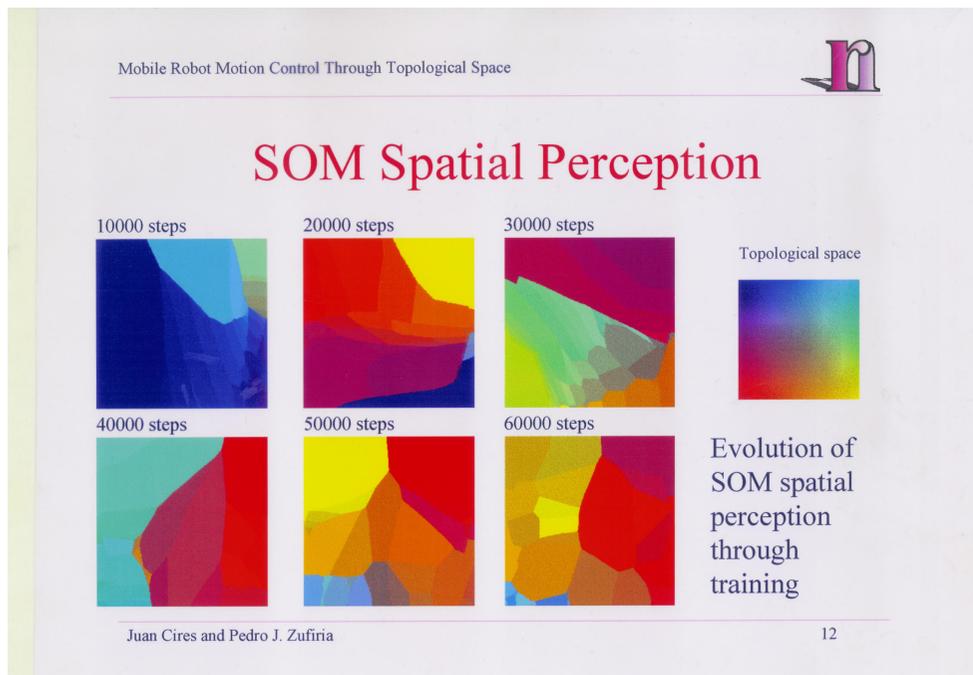
Sistema híbrido neuro-simbólico para el control multisensorial de un robot móvil

## El mapa topológico auto-organizativo

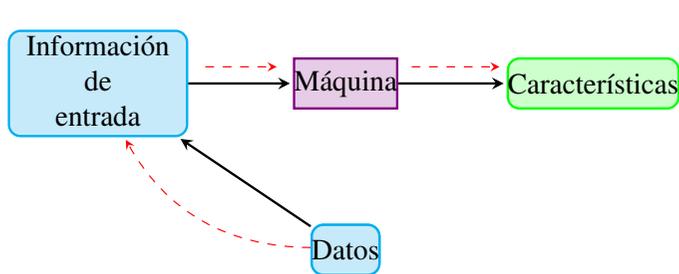
- Activación utilizada como representación de las señales de entrada

$$\|w^r - x\| = \min_i \|w^i - x\|$$

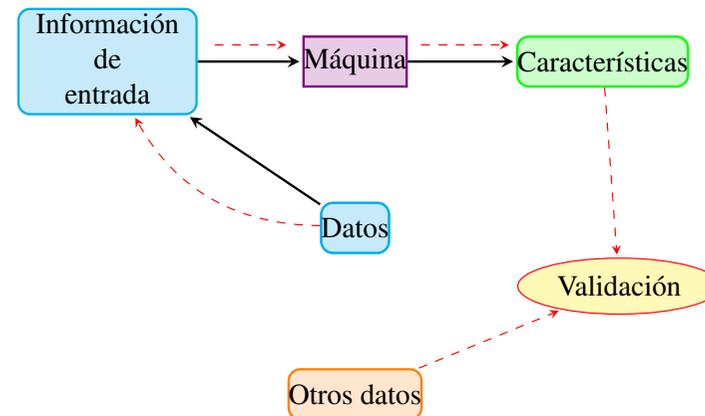
- Elementos importantes en el aprendizaje:
  - Codificación del vector de entrada
  - Política de presentación de patrones



## Métodos de interdependencia. Validación



Formulación de aprendizaje automático (no supervisado)  
(Fase de entrenamiento (rojo) y fase de ejecución.)





# Paréntesis: aprendizaje de conceptos; conocimiento

# Aprendizaje estadístico y niveles de Marr

Semántica asociada al concepto

Supervisado	Adquisición	Concepto
vs. No supervisado	Formación	

Semántica asociada al conocimiento

Adquirir	Conocimiento
Extraer	
Descubrir	
Inferir	

Aprendizaje estadístico	Niveles de Marr
Problema de aprendizaje	Computacional
Máquina de aprendizaje	Algorítmico Implementacional

} Sistema dinámico con modularidad específica

Redes neuronales proporcionan diferentes modelos de máquina de aprendizaje (niveles algorítmico e implementacional) para los diferentes tipos de problemas

# Aprendizaje estadístico. Tipos de problemas

# Aprendizaje estadístico. Tipos de problemas

## Inferencia clásica (básica)

- Estimación paramétrica: subcategoría de problema de estimación de densidades (conocida estructura de distribución)
- Test de hipótesis: subcategoría de problema de estimación de densidades (relevancia de características) paso previo a problemas de regresión o clasificación
- Regresión lineal

## Inferencia clásica (avanzada)

- Estimación no paramétrica (de densidades): kernel estimation methods (Parzen window, etc.). Kernel aproxima función de densidad. (No confundir con “kernel trick” -en técnicas de clasificación, PCA, etc.- que aproxima relaciones entre variables en análisis multivariante.)

## Técnicas de análisis multivariante

(solución a subproblemas de aprendizaje estadístico)

Caracterizan comportamiento conjunto de varias variables (vectores aleatorios)

Diferentes tipos de caracterización ⇒ Diferentes tipos de problemas resueltos

Técnicas de análisis multivariante

Métodos de dependencia (variable dependiente de otras)

- **Regresión** multivariable, regresión logística (**clasificación**)
- Análisis discriminante (determinar variable categórica **-clasificación-** mediante variables independientes continuas o binarias)  
Análisis discriminante lineal (clasificación lineal binaria con datos gaussianos)
- ANOVA, MANOVA: dependencia de variable con respecto a grupo (datos agrupados a priori)  
Interpretable como problema de **regresión**: determinar variable continua a partir de variables categóricas

Técnicas de análisis multivariante

Métodos de interdependencia (conjunto de variables)

- PCA: determinar propiedades entre las variables de una distribución conjunta
- Análisis factorial (caracterizar relaciones con variables latentes)
- Análisis de correspondencias (caracterizar relaciones con variables sintéticas)
- Análisis de correlación canónica (reformular correlación entre dos grupos de variables)
- Análisis cluster

Interpretables como aproximaciones a la **estimación de densidad** conjunta

Técnicas de análisis multivarianteTécnicas de análisis multivariante

Modelos de ecuaciones estructurales (asunciones cualitativas sobre causalidad + datos estadísticos).

Incorporan información estructural adicional

Definen un modelo general que recoge modelos particulares

- Regresión lineal
- Análisis factorial
  - Confirmatorio: falsar modelo o hipótesis previa variante del análisis factorial exploratorio (clásico)
  - Exploratorio



¿Técnicas multivariantes no lineales?

Redes neuronales (y paradigmas asociados):

- Regresión: PMC, RBF...
- Clasificación:
  - PMC
  - Kernel-based-methods: Support Vector Machines (SVM), Nonlinear Discriminant Analysis (extension of LDA)
- Estimación de densidades (o aproximaciones):
  - SOM
  - Kernel-based-methods: nonlinear PCA, spectral clustering



# Aprendizaje estadístico. Tipos de problemas

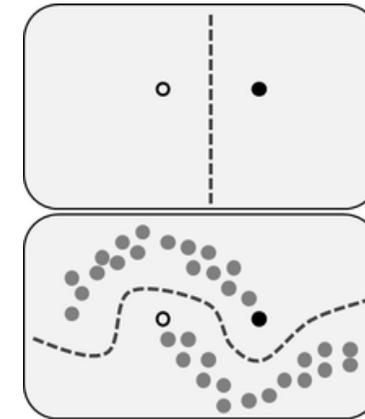
Objetivo: descripción, predicción  
 Recursos (datos): supervisado, no supervisado

	Modelo predictivo	Modelo descriptivo
Apr. supervisado	<b>Clasificación, regresión</b>	Descubrimiento de subgrupos, clustering ( $\equiv R^2$ , ANOVA)
Apr. no supervisado	Clustering predictivo	<b>Clustering descriptivo, descub. reglas de asociación</b>
Apr. semi-supervisado	Clasificación, regresión	
Apr. transductivo	Clasificación	
Apr. on-line	Decisión	
Apr. por refuerzo	Decisión	
Apr. activo	Regresión, clasificación	

# Aprendizaje estadístico. Tipos de problemas

## Problemas derivados, variantes o combinaciones

### Semi-supervised learning

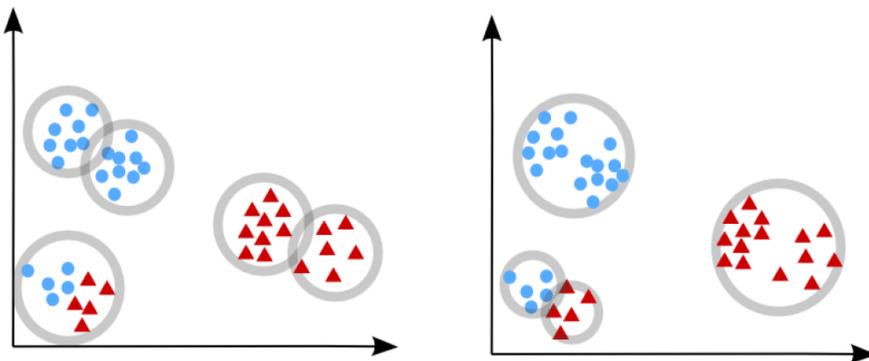


# Aprendizaje estadístico. Tipos de problemas

# Aprendizaje estadístico. Tipos de problemas

## Problemas derivados, variantes o combinaciones

### Supervised-clustering



## Problemas derivados, variantes o combinaciones (cont)

- Aprendizaje de conceptos es un caso particular de problema de clasificación
- Clustering y/o reducción de dimensionalidad pueden ser vistos como aproximaciones al problema de estimación de densidades (reducción mediante proyección a dimensión 2 o 3 para visualizar datos)
- Cuantificación vectorial puede combinar clustering y reducción de dimensionalidad mediante etiquetado de vectores
- Predicción combina regresión con dependencias o relaciones temporales
- Aprendizaje por refuerzo incorpora elementos adicionales de dinámica y control  $\Rightarrow$  plantea un nuevo paradigma
- *Recommender systems* emplean aspectos de aprendizaje supervisado, pero el problema es más complejo
- Reconocimiento de patrones: familia de problemas muy diferentes. (Ambigüedad semántica)

## Aprendizaje estadístico. Tipos de problemas

Ejemplo de ambigüedad semántica: “Reconocimiento de patrones”

Asignar etiqueta a un dato de entrada. Puede corresponder a muchos problemas: clasificación (supervisado), clustering (no supervisado), regresión (asignar valores reales en lugar de etiquetas), etiquetado de secuencias, etc. Incluso si nos centramos en clasificación:

- Problema de clasificación con modelo probabilístico completo disponible  
Sub-categoría: problema de detección
- Problema de clasificación sin modelo probabilístico disponible, pero con ejemplos de clasificación: datos de pares características↔clase
- Problema combinado
  - extracción de características (aproximación a estimación de densidades) sin modelo probabilístico disponible +
  - clasificación con ejemplos de clasificación sobre datos totales (no sobre características)
- Problema de extracción de características + clasificación con información específica (de experto)

Problema diferentes: mismos objetivos, pero recursos diferentes

## Marcos de evaluación de modelos de aprendizaje en máquinas

- Teoría de aprendizaje estadístico
  - Consistencia: ley grandes números en espacios de funciones
  - Cotas/Capacidad:
    - complejidad de Rademacher: riqueza de una clase de funciones
    - Annealing entropy
    - Growth function
    - dimensión de Vapnik-Chervonenkis (VC): capacidad de algoritmo de clasificación (cardinalidad del mayor conjunto que puede particionar).  
Ventaja: es más fácilmente computable y acota a las anteriores
  - Control de generalización: nuevos principios inductivos  
Structural Risk Minimization (SRM), regularization, cross validation, ensemble learning, etc.
  - Algoritmos
- Teoría de aprendizaje computacional: complejidad algorítmica o computacional
- Teoría PAC (Probably Approximately Correct)= cotas + complejidad computacional  
(Efficiently PAC learnable class)

## Aprendizaje estadístico. Tipos de problemas

Ejemplo de competición IJCNN 2011 (estimación de enlaces en red):

- Datos (iniciales): 7.237.983 enlaces (entrada o salida) en red social
- Objetivo: Predecir si 8960 enlaces adicionales son verdaderos o falsos

Conocimiento (adicional) disponible es clave para catalogar problema:

- Población total = *training set* + *test set*  
(podría ser que *training set* + *test set* < Población total)
- Generación de *test set* y existencia de *validation test*:
  - Procedimiento muestral para generar *test set* puede ser o no conocido
  - Con *validation test*: ejemplos de enlaces falsos, muestreados según el proceso (no necesariamente conocido). Permite aplicar clasificador (p.e., SVM) con validación cruzada (problema de clasificador binario estándar con muchos datos, balanceados)
  - Sin *validation test*: problema de qué hace que un enlace sea real o falso ⇒ problema de ingeniería inversa del proceso de muestreo
- Conocimiento de red real: desanonización