



# MODELOS PREDICTIVOS

---

MAPAS AUTORGANIZATIVOS DE KOHONEN

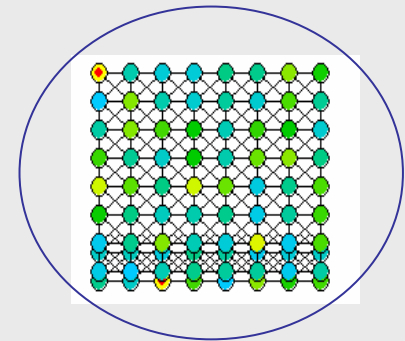
**Jaime Miranda**

Departamento de Ingeniería Industrial  
Universidad de Chile

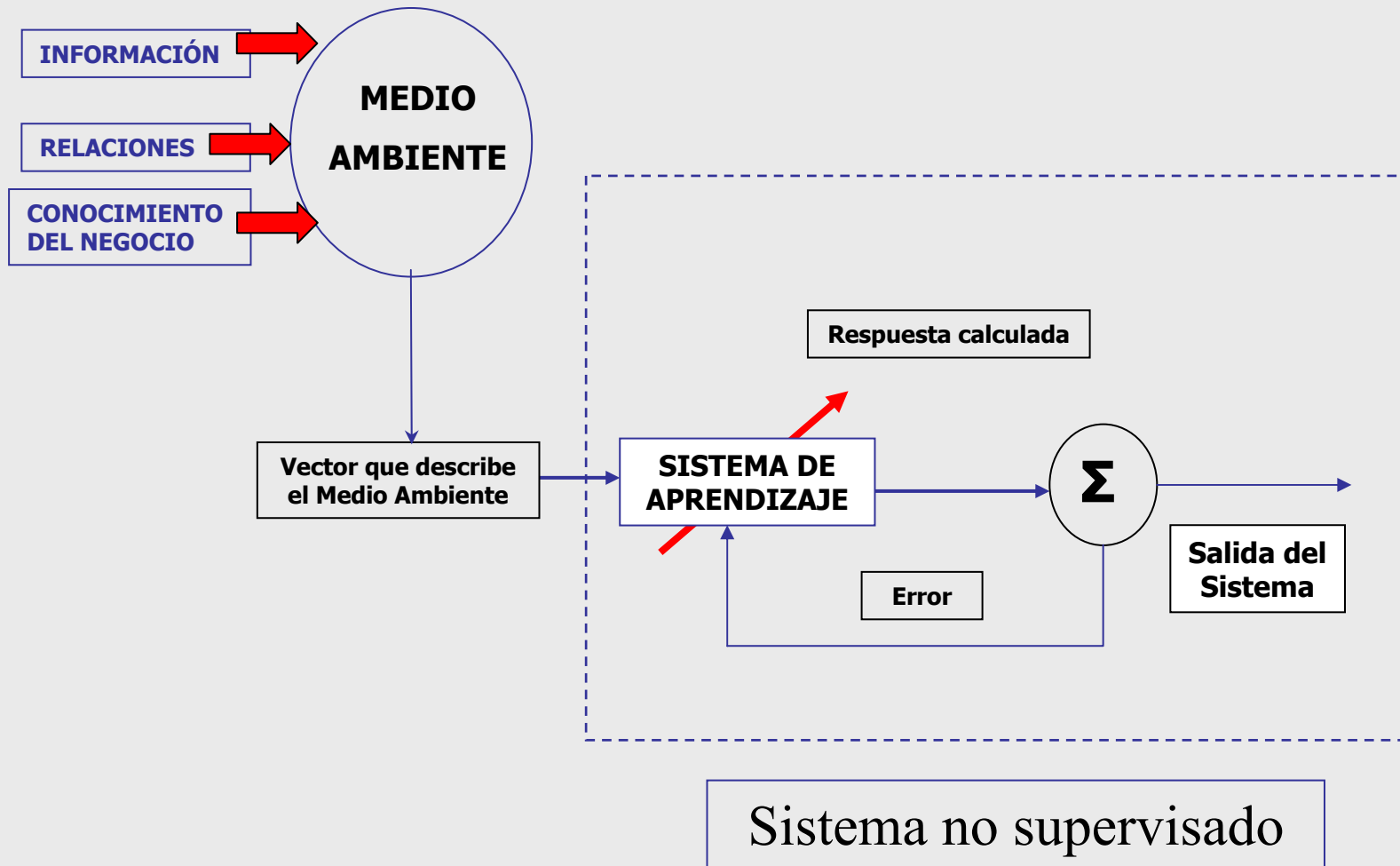
Sólo ocupan como información los atributos o variables de entrada

No se conocen las clases a priori

Se ocupan relaciones de similitud y distancia entre los objetos

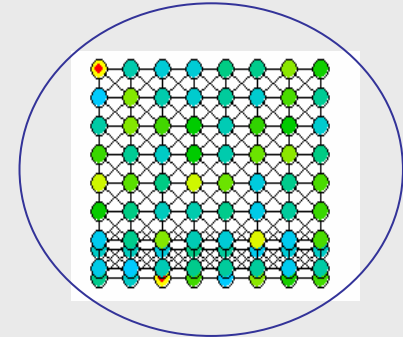


# DIAGRAMA DE APRENDIZAJE NO SUPERVISADO



## APLICACIONES

- **Marketing**
  - Segmentación de clientes
  - Ofertas focalizadas



## FORTALEZAS

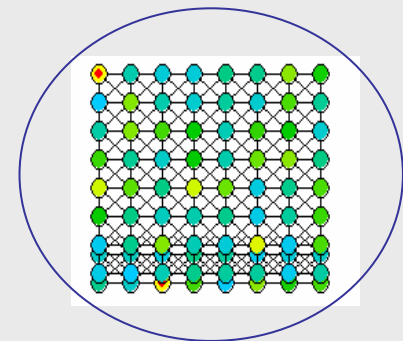
- Reduce la dimensionalidad del espacio.
- Los resultados son mas intuitivos y fáciles de entender e interpretar

## DEBILIDADES

- Sólo nos da una visión espacial de los resultados
- No trabajan bien con variables categóricas

## Proviene de la cuantización vectorial

- Aproximación a la función de densidad de probabilidad  $P(x)$  de las variables de entrada ( $x$ ) usando un número finito de prototipo
- La idea es relacionar un objeto con su prototipo más cercano.
- Trata de buscar una aplicación topológicamente ordenada del espacio de entrada
- Los hiperplanos de separación son lineales en trozos

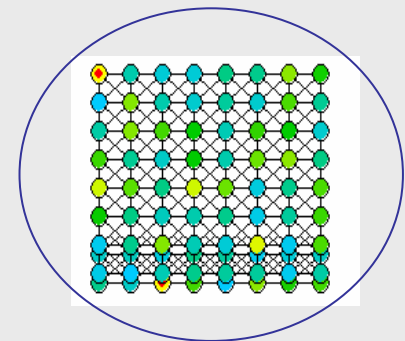


Método no supervisado

Especializado para encontrar relaciones entre subconjuntos

Reduce la dimensionalidad del espacio de entradas

Tiene forma de grilla (2D) o caja (3D)



## CLUSTERING

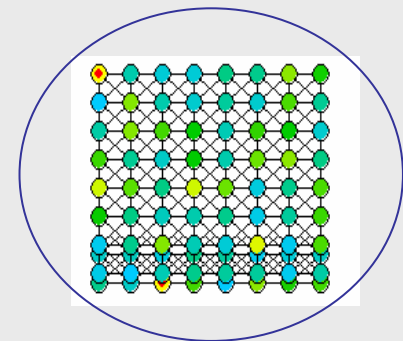
- **Aproximación:** es posible agrupar datos del conjunto de entrada, atendiendo a diferentes criterios con el fin de asignarle una clase a cada objeto del conjunto de entrada

## VISUALIZACIÓN

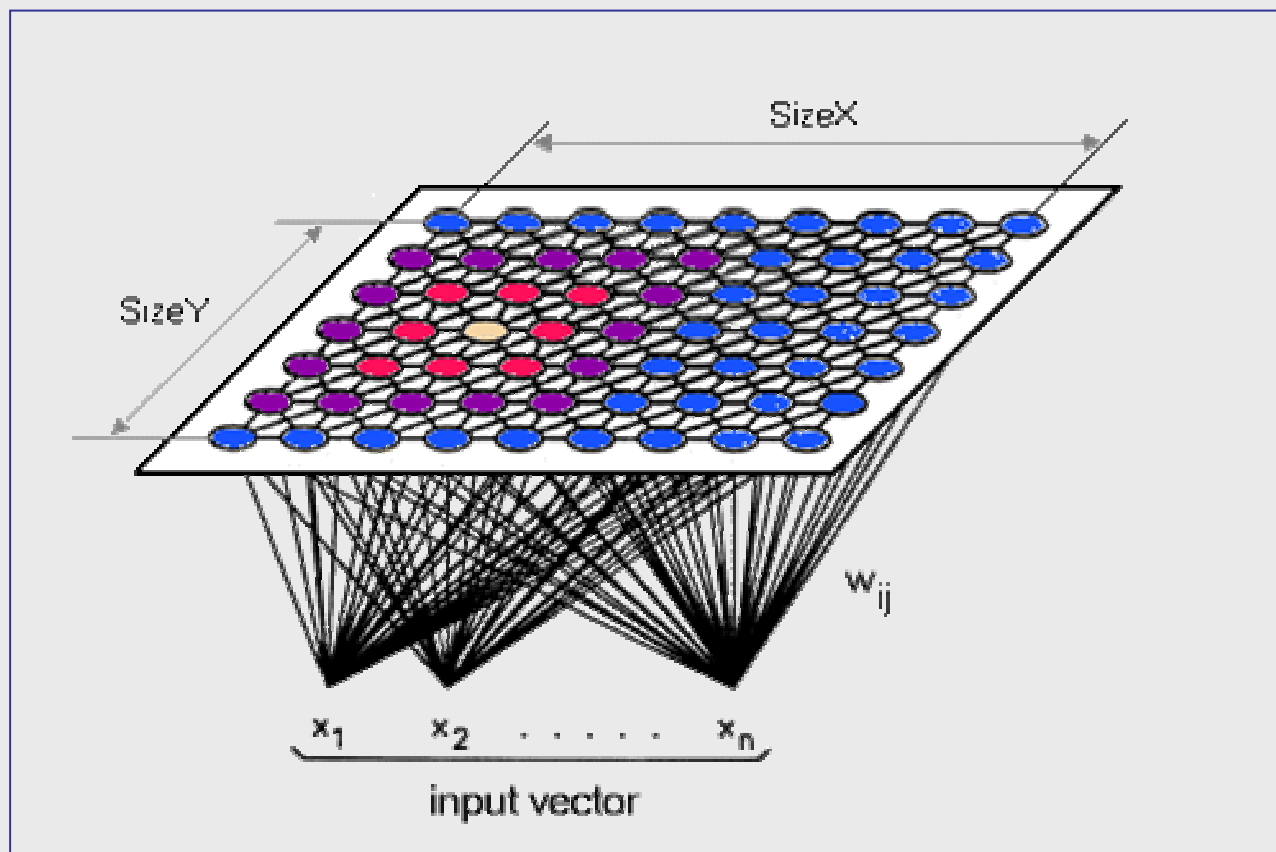
- Este agrupamiento, trata de mostrar en forma grafica las similitudes a través de la mantención de las características topológicas

## CLASIFICACIÓN:

- Una vez calibrado el mapa, o asignada algún tipo de etiqueta a cada clusters, se puede usar para clasificar sobre datos desconocidos



# EN FORMA GRÁFICA ...





1. Inicializar los prototipos

2. Encontrar neurona ganadora (BEST MATCHING)

$$m_i = m_i(0)$$

3. Modificar los vectores prototipos

$$\|x(t) - m_c(t)\| = \min \{ \|x(t) - m_i(t)\| \}$$

4. Volver a 3 hasta que no ocurran cambios significativos en el mapa **para  $i \in N_c$**

$$m_i(t+1) = m_i(t) + \alpha [x(t) - m_i(t)]$$

$$m_i(t+1) = m_i(t)$$

**Tasa de aprendizaje**

## El algoritmo básicamente realiza:

- Definir la estructura de la red (cantidad de neuronas y dimensionalidad)
- Inicializar neuronas aleatoriamente
- Presentar un ejemplo y determinar la neurona ganadora
- Modificar la neurona ganadora y las de la vecindad de acuerdo al ejemplo
- Volver a (3) hasta que no se produzcan cambios significativos

## Dimensiones del mapa

- 2D ó 3D
- Número de nodos y clases

## Geometría del mapa

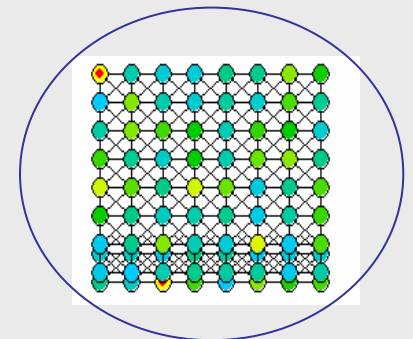
- Definición de la función de vecindad y radios de convergencia

## Tasa de aprendizaje

- Ajuste de los pesos

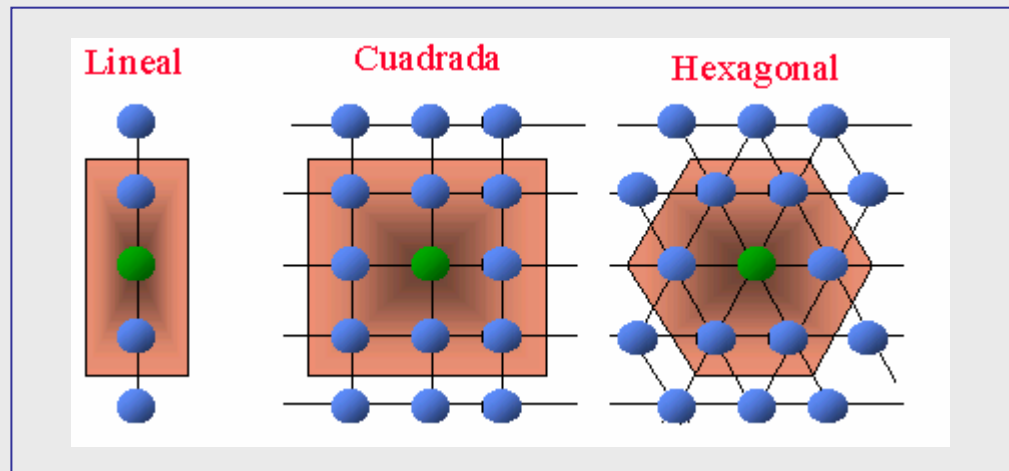
## Épocas

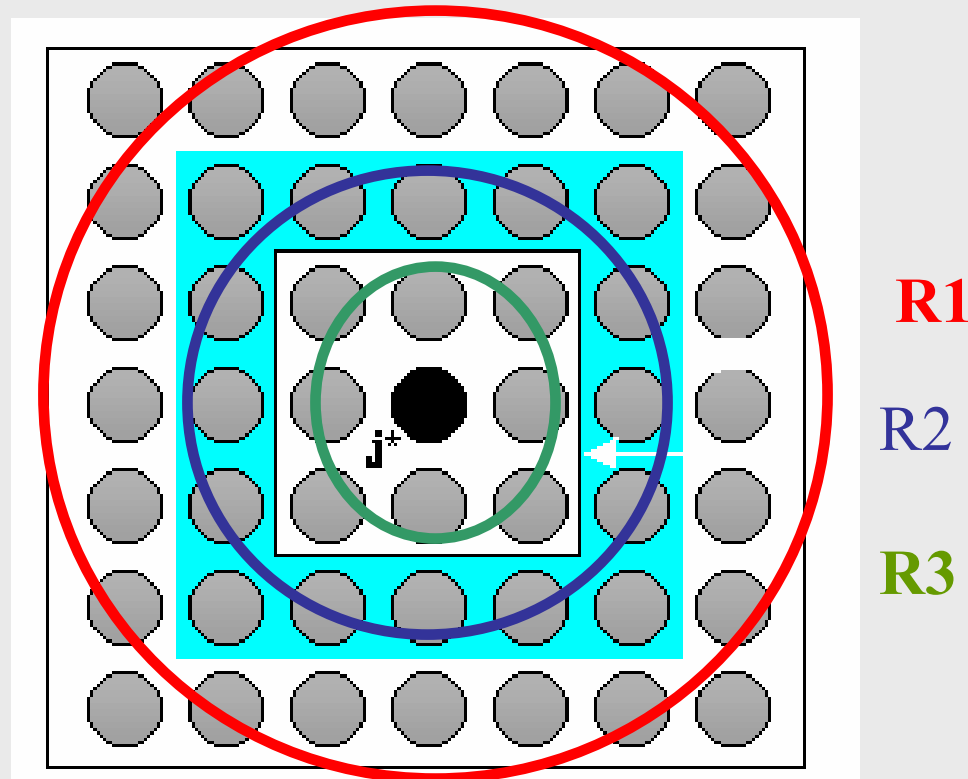
- Número de veces que se “muestran” los datos al método



## OBJETIVO:

- Nos dice que neuronas serán modificadas en cada iteración del algoritmo
- Esta vecindad decrece a mitad que pasan la épocas o iteraciones





En cada etapa de la iteración se disminuye el radio de pertenencia  $R_i$

## PROBLEMA:

- Clasificar a un grupo de países a través de variables socioeconómicas
- Conjunto de 23 países
- Cada país esta caracterizado por 18 atributos (4 grupos)

ATRIBUTOS DE LOS PAISES	
CLASES	DESCRIPCIÓN
1	Consumo energético y condiciones ambientales
2	Desarrollo económico.
3	Desarrollo tecnológico.
4	Población y densidad demográfica.

## DESARROLLO ECONÓMICO

→ Países cercanos

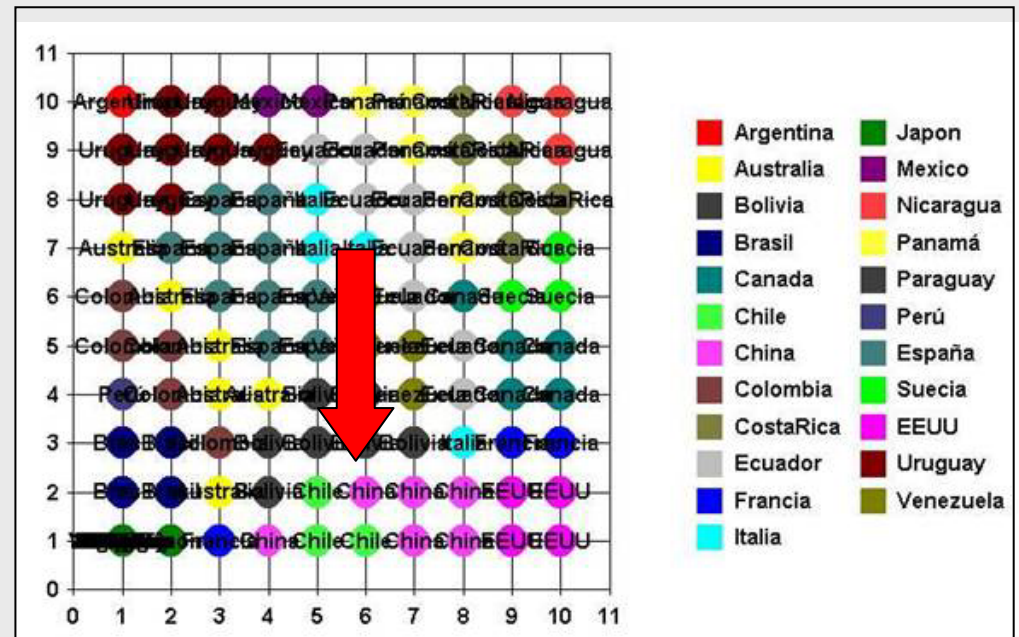
→ China.

→ Francia

→ Australia

→ Chile

→ EE.UU.



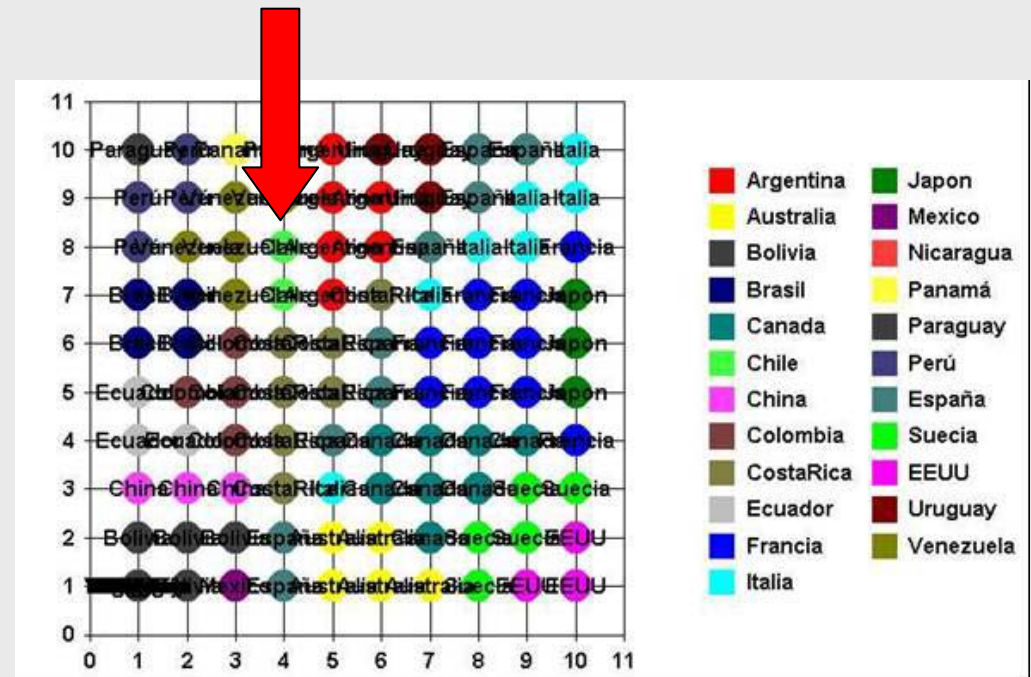
## DESARROLLO TECNOLÓGICO

→ Países vecinos:

- Argentina
- Venezuela
- Costa Rica
- Uruguay

→ Otros países vecinos:

- EE.UU.
- Suecia
- Australia



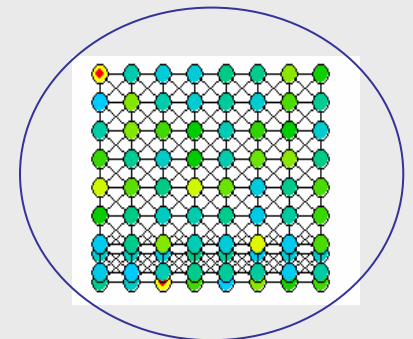


Entrenamiento no supervisado, lo que da independencia a supuestos fuertes (iid)

No necesita pares de Entrada/Salida (sólo datos de entrada), lo que ayuda a estudios exploratorios

Disminuye la dimensionalidad del problema de un espacio de alta dimensionalidad a 2 o 3 dimensiones

Visualmente intuitivo y “representable”

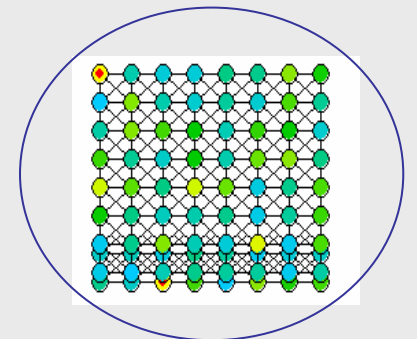


Sólo entrega información sobre la ubicación espacial de los objetos

No trabaja bien con variables categóricas – medidas de distancia

Los vértices del mapa muchas veces son poco explicables

Es necesario interpretar la información encontrada (PCA)





# MODELOS PREDICTIVOS

---

MAPAS AUTORGANIZATIVOS DE KOHONEN

**Jaime Miranda**

Departamento de Ingeniería Industrial  
Universidad de Chile