

Aprendizaje no Supervisado

a lazo abierto → sistema autoorganizado → no hay realimentación

Objetivo: *descubrir patrones o características significativos en los datos de entrada*

Dos subparadigmas clásicos: **Aprendizaje hebbiano**
Aprendizaje competitivo

	<i>familiaridad</i>
	<i>ACP</i>
Diversas arquitecturas,	<i>Clustering</i>
distintos fines	<i>Prototipos</i>
	<i>Codificación</i>
	<i>Mapeo de características</i>

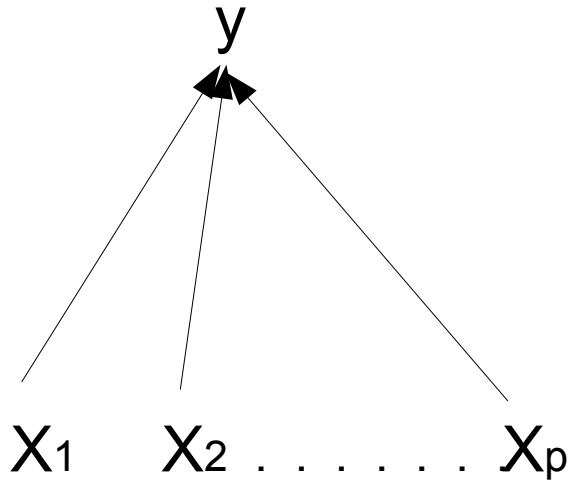
Principios básicos:

- Para el aprendizaje no supervisado: ***los datos de entrada deben contener redundancia.***
- Para la autoorganización: ***el orden global es un emergente de múltiples interacciones locales***

Válido tanto para las RN biológicas (= sistema nervioso) como para las RNA.

En general, aproximan a estructuras biológicamente plausibles en mucho mayor medida que los usados para aprendizaje supervisado

Aprendizaje hebbiano no supervisado



$$y = \sum w_i x_i$$

modelo lineal con aprendizaje hebbiano y una unidad de salida.

$X \approx \mathcal{P}(x) \rightarrow$ la red entrenada deberá indicar correspondencia de la entrada a \mathcal{P} .

Hechos:

(1) Si el aprendizaje es plano: $\Delta w_i = \eta y x_i$

\rightarrow incremento no acotado de los pesos (= sólo equilibrios inestables)

(2) Si se normalizan los pesos después de cada iteración: $w_i' = \alpha w_i, \alpha / |w'| = 1$

\rightarrow Regla de Oja (82'): $w_i(n+1) = w_i(n) + \eta y(n) [x_i(n) - y(n) w_i(n)]$

a primer orden en η (η pequeño)

(3) Teorema de Oja: la dinámica de (2) evoluciona a w tal que:

i) $|w| = 1$

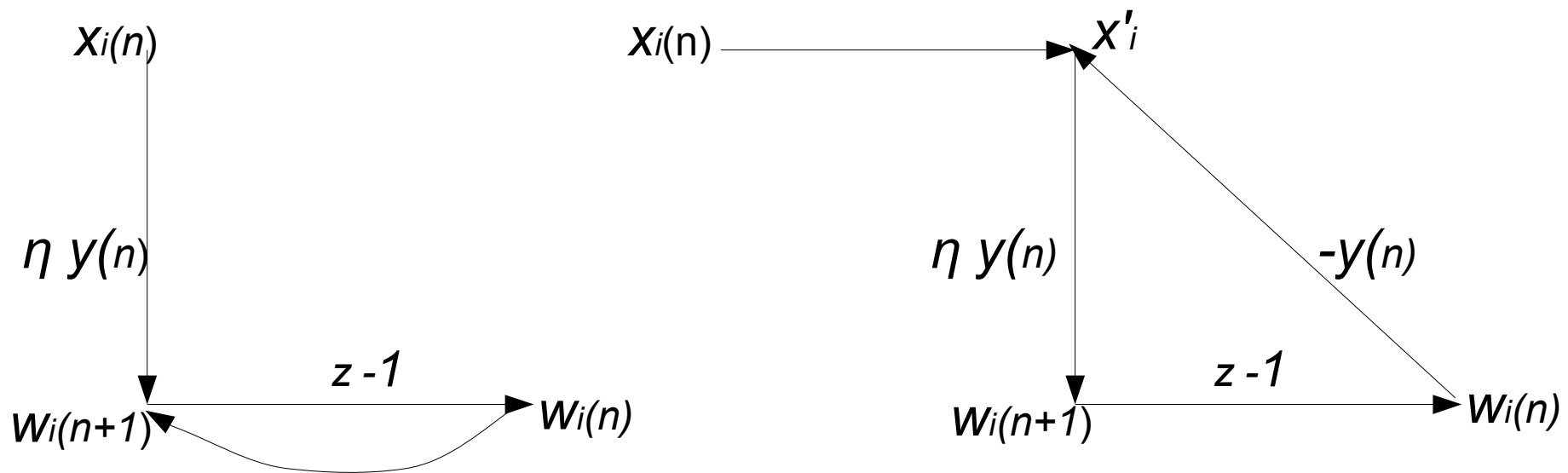
ii) w está en la dirección de un autovector de λ máximo de R , matriz de correlación: $R = \langle x x' \rangle$

iii) w está en una dirección que maximiza $\langle y^2 \rangle$

→ *filtro para la primera componente principal de la entrada*

PLANO → sólo feedback positivo

OJA → feedback positivo y negativo



z^{-1} operador de retardo de una unidad de tiempo

Análisis de componentes principales

Problema: *dada una distribución en R^n , encontrar m vectores ortogonales ($m < n$) con la mayor varianza de los datos al proyectar sobre ellos \rightarrow reducción de la dimensión.*

Definición (inductiva): la k -ésima componente principal (cp) es la dirección que maximiza la varianza dentro del subespacio ortogonal a las primeras $k-1$ cp's.

PROPOSICION: la k -ésima componente principal es en la dirección de un autovector asociado al k -ésimo mayor autovalor de la matriz de covarianza:

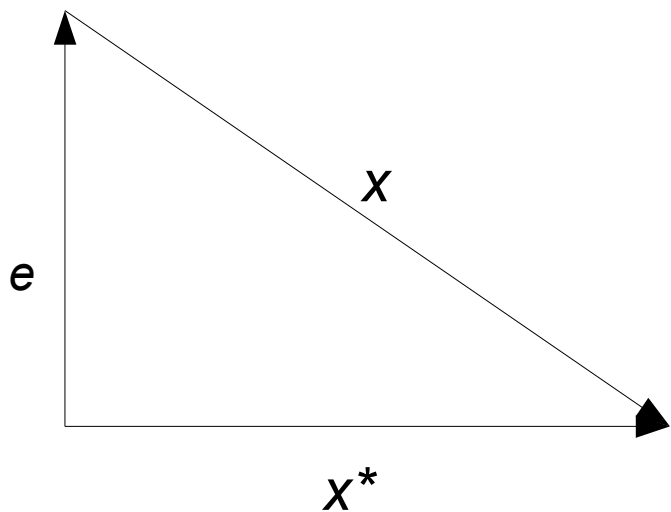
$$\langle (x - \langle x \rangle)(x - \langle x \rangle)' \rangle$$

$$\text{Si } x \in R \rightarrow x = \sum_{k=1}^n a_k u_k \quad a_k = u'_k x$$

$\{u_k\}_{k=1}^n$ *base de autovectores ortonormales*

$$x^* = \sum_{k=1}^m a_k u_k \sim x \quad \text{aproximación truncando a } m \text{ primeras cp.}$$

$$\rightarrow e = x - x^* = \sum_{k=m+1}^n a_k u_k \quad \text{e.c.m., ortogonal a } x^*$$



*Cuanto más cercanos a 0 estén los λ_k con $m+1 \leq k \leq n$, más efectiva la reducción de dimensionalidad.
La proyección sobre cp es la transformación lineal inversible que minimiza el e.c.m.*

Procesamiento de una distribución de entrada

Selección de características

Espacio de datos → espacio de características
Sin pérdida de información
→ misma dimensión (e.g. wavelets)

Extracción de características

la transformación permite extraer las características más relevantes
→ reducción de la dimensión

ACP o Transformación de Karhunen-Loeve

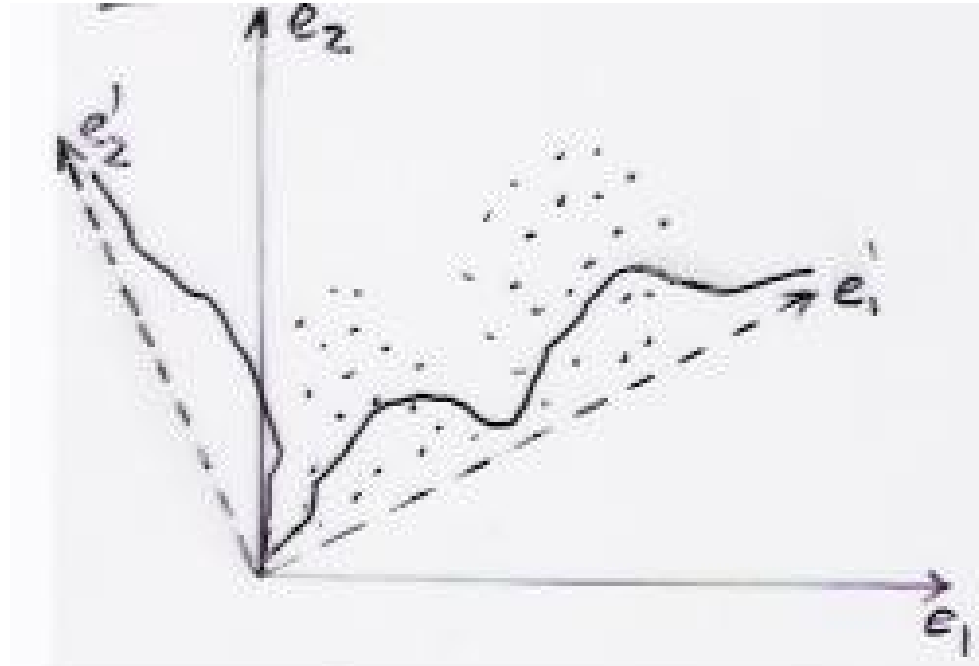
Pearson, 1901: contexto biológico

Hotelling, 1933: en psicometría

Karhunen, 1947: en teoría de probabilidades

Loeve, 1963: en teoría de las comunicaciones

Ejemplo:



Las proyecciones sobre los ejes canónicos enmascaran la naturaleza bimodal de la distribución.

La primera componente principal maximiza la varianza de la distribución (su proyección). La segunda, siendo ortogonal, no distingue.

→ $\sigma_{e'_1}^2 \gg \sigma_{e'_2}^2$ → reducción 2 → 1 dimensión da máxima eficiencia

$$\sigma_{e_1}^2 + \sigma_{e_2}^2 = \sigma_{e'_1}^2 + \sigma_{e'_2}^2 \quad (\text{varianza total es invariante})$$

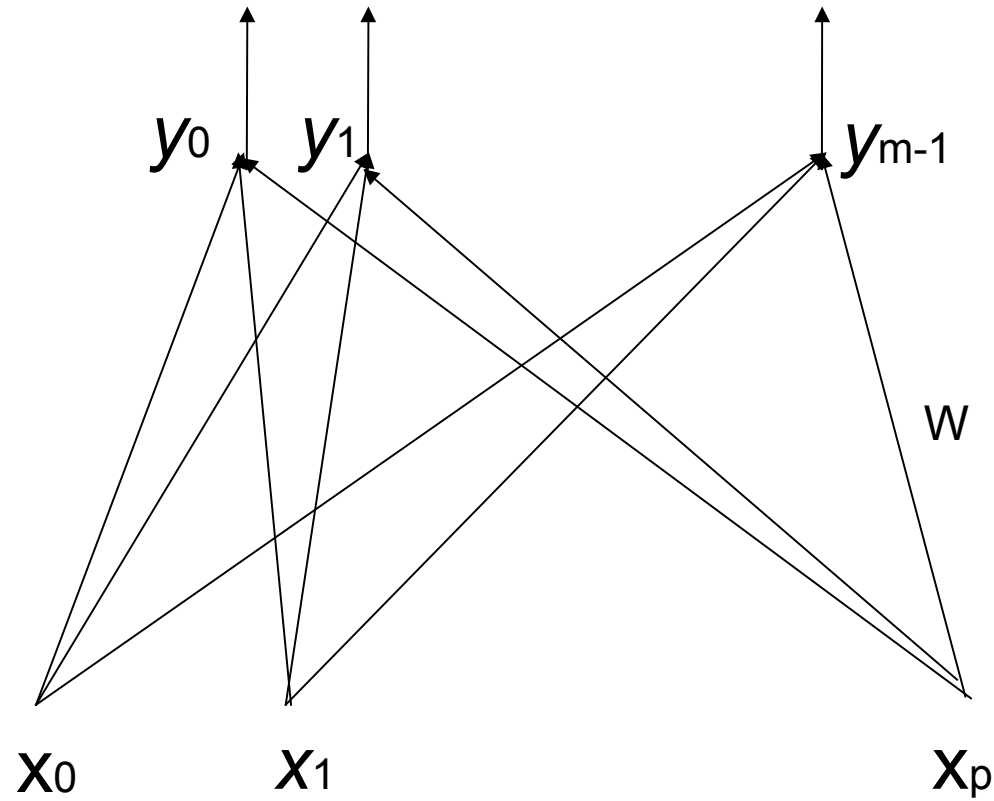
$$\langle (x'e_1)^2 \rangle + \langle (x'e_2)^2 \rangle = \langle (x'e'_1)^2 \rangle + \langle (x'e'_2)^2 \rangle$$

$$\langle x_1^2 \rangle + \langle x_2^2 \rangle = e'_1{}' R e'_1 + e'_2{}' R e'_2$$

$$\langle \|x\|^2 \rangle = \lambda_1 + \lambda_2$$

Si $\lambda_1 \gg \lambda_2 \rightarrow$ la proyección unidimensional sobre e'_1 maximiza la varianza

Aprendizaje hebbiano generalizado



Aprendizaje hebbiano generalizado

$$y_j(n) = \sum_{i=0}^p w_{ji}(n) x_i(n) \quad j = 0, 1, \dots, m-1$$

Reglas de aprendizaje:

- Oja ('89): $\Delta w_{ji}(n) = \eta y_j(n) [x_i(n) - \sum_{k=0}^{m-1} w_{ki}(n) y_k(n)]$

- Sanger ('89): $\Delta w_{ji}(n) = \eta y_j(n) [x_i(n) - \sum_{k=0}^j w_{ki}(n) y_k(n)] =$

$\eta y_j(n) [x_i'(n) - w_{ji}(n) y_j(n)]$ con $x_i'(n) = x_i(n) - \sum_{k=0}^{j-1} w_{ki}(n) y_k(n)$

$m - 1$

En común: $\{w_j\}_{j=0}^{m-1}$ límite son base ortonormal: $w_i' w_j = \delta_{ij}$

Generan el mismo subespacio: $\text{span}\{w_j(\text{oja})\} = \text{span}\{w_j(\text{sanger})\}$

Diferencia: Sanger extrae cp, Oja no.

Sup. w_{ji} ya convergieron a las k primeras cp ($j = 0, 1, \dots, k-1$)

→ entrada a la unidad k : $x_i'(n) = x_i(n) - \sum_{j=0}^{k-1} w_{ji}(n) y_j(n)$

o bien: $x'(n) = x(n) - \sum_{j=0}^{k-1} W_j(n) y_j(n) = x(n) - \sum_{j=0}^{k-1} W_j(n) W_j'(n) x$

= entrada x menos las primeras k cp de la entrada original en la base ortonormal de cp.

→ $x'(n)$ vive en el subespacio de las $m-k$ últimas cp.

→ la $k+1$ -ésima neurona tenderá a extraer la $k+1$ -ésima cp.

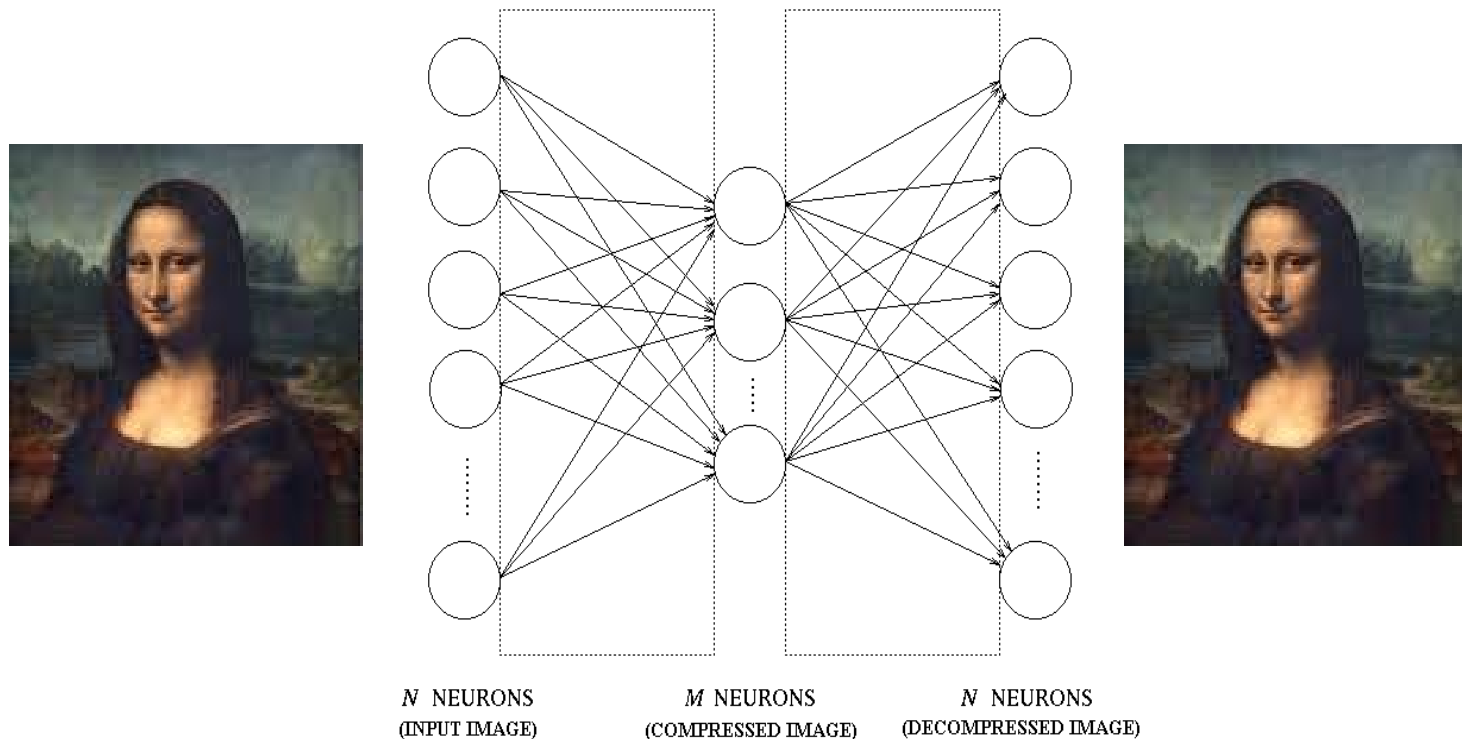
Luego:

Sanger: *resultados reproducibles (salvo diferencias de signo)*

Oja: *-resultados dependientes de los valores iniciales.
-preferible si se necesita información
uniformemente repartida*

Hertz: desde un punto de vista biológico, ¿qué ventaja tendría la información ordenada según componentes principales?

Compresión de imágenes por encoder/decoder



Entrada=salida (*perceptrón multicapa autoasociador*)

$M \ll N$

Codifica en m dimensiones → *mismo resultado que Oja m dimensional, con igual varianza promedio en cada unidad oculta.*