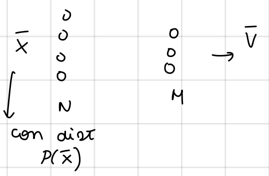
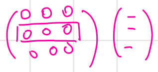
aprendizaje no supervisado competitivo



$$V_i = \sum_{j=1}^{N} w_{ij}\, x_j = \bar{w}_i \cdot \bar{x}$$



se determina el indice i^* / $\bar{w}_{i^*} \cdot \bar{x} \geq \bar{w}_i \bar{x} \quad \forall x$

(tal que la salida es máximo

$|\bar{w}_i| = 1$, $|\bar{w}_{i^*} - \bar{x}| \leq |\bar{w}_i - \bar{x}| \quad \forall i$

$$|\bar{w}_{i^*} - \bar{x}| \leq |\bar{w}_i - \bar{x}| \quad \forall i$$

$$\Delta w_{i^*} = \eta \left(\frac{x}{\sum x_j} - \bar{w}_{i^*} \right)$$ (normalización

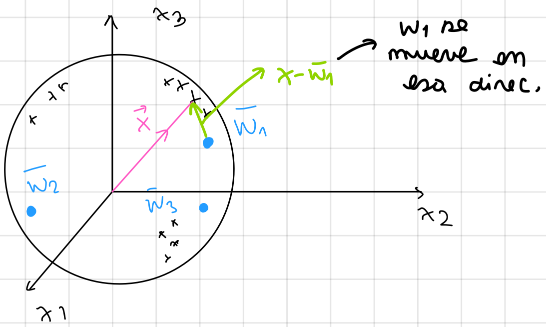
o $\Delta \bar{w}_{i^*} = \eta\,(x - w_{i^*})$

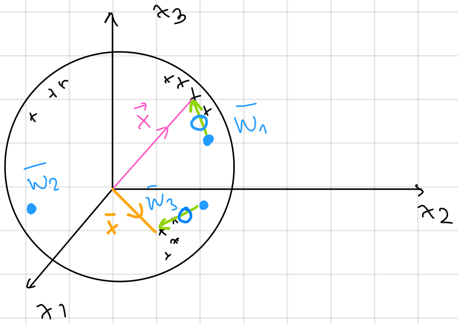
modifico w en la direccion i^*

Suponemos que $|x| = 1$ ⇒ los ptos caen en una esfera centrada en el (0,0,0)

$M = 3$

\# de clusters





tomamos $\Delta \bar{w}_i = M_i\, \eta\, (\bar{x} - \bar{w}_i)$ $\quad M_i = \begin{cases} 0 & \text{si } i \neq i^* \\ 1 & i = i^* \end{cases}$

Supongamos que generamos con $P(\bar{x})$ p datos : $\bar{x}^\mu \quad \mu = 1, \dots, p$

cada set nos da un ganador diferente (i^*) ⇒ un dist M_i^μ

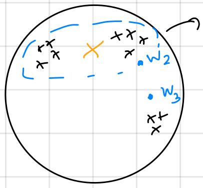
Definimos la función de costo $E(\{\bar{w}\}) = \frac{1}{2} \sum_{i=1}^{n} \sum_{\mu=1}^{p} \underbrace{M_i^\mu}_{\text{la salida es no nula si } M_i^\mu = 1} |\bar{x}^\mu - \bar{w}_i|^2$

$$\Rightarrow \frac{\partial E}{\partial w_{ij}} = - \sum_{\mu=1}^{p} M_i^\mu \,(x_j^\mu - w_{ij})$$

$$\Rightarrow \Delta \bar{w} = -\eta \frac{\partial E}{\partial \bar{w}}$$

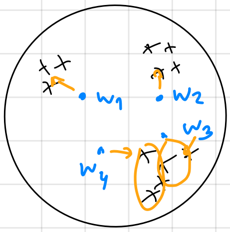
Si tenemos $\Delta W_{ij} = \begin{cases} \eta \, (x - \bar{W}_{i^*}) & i = i^* \\ -\eta' \, (\bar{x} - \bar{W}_{i^*}) & i \neq i^* \end{cases}$

Si se piensa que se tienen dos clusters (pero en realidad se tienen 3)



→ un gran cluster por más que en el CM no haya nodo

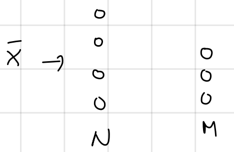
Si ponemos un W de más → un cluster se puede dividir en dos



ayuda a entender dist de prob en muchos dim

Forma de resolución → propongo muchas unidades (M grande)
si se ve que alguna hace algo raro, se descarta
↓
Adaptive Resonance Theory → hace esto

ADAPTIVE RESONANCE THEORY



1) inicializan pesos aleatorios y pequeños
2) elijo entrada x de $P(\bar{x})$
3) elijo ganador i^* (neurona de salida con máxima actividad)
↓ maximiza $\bar{W}_i \bar{x}$

4) Calculo $R = \dfrac{\bar{W}_{i^*} \cdot \bar{x}}{\sum_{j=1}^{N} x_j}$ con $\underbrace{\bar{W}_{i^*} = \dfrac{W_{i^*}}{\varepsilon + \sum_{j=1}^{N} w_{ij}}}_{W \text{ normalizado}}$

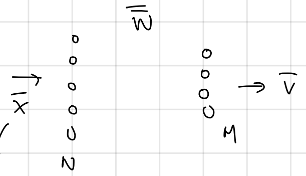
5) si $R < \rho$: elimino la unidad i^* → voy a (2)
si $R > \rho$ → voy a (2)

6)

Puede pasar que : → se quede sin unidades

→ se tienen x unidades → converge

Feature mapping



vienen de $P(\bar{x})$

los neuronas no están en realidad en una línea

Supongamos que se tienen entradas $\bar{x}$ con dim N y dist $P(\bar{x})$ y se quiere hacer un mapeo

$$\bar{x}, N \to \bar{y}, M$$

Se quiere que

- ptos vecinos en el espacio de entrada $(\bar{x})$
- ptos vecinos en el esp. de salida $(\bar{y})$

(En sist sensoriales: se tienen estímulos, cada uno actúa de manera dist en la corteza (estructura en capas))

Algoritmo de Kohonen → modelo de Feature mapping (ej 2 - P5)

$V_i = \bar{w_i} \bar{x}$

1) minimizan $\bar{x}$
2) eligen $\bar{x}$ de $P(\bar{x})$
3) $i^* \;/\; |\bar{w}_{i^*} - \bar{x}| \leq |w_i - \bar{x}|$
4) $\Delta \bar{w}_i = \eta \, \Omega(i, i^*) \, (\bar{x} - \bar{w}_i)$

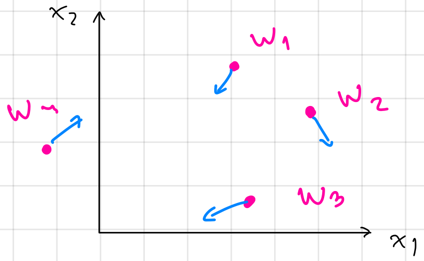
función de vecindad ↓ dice como están dist. los ptos en el espacio de salida

Cada i tiene asociado un $r_i = (r_{ix}, r_{iy})$ ↓ como están dist en el espacio

$$r(i, i^*) = e^{-|r_i - r_{i^*}|^2 / 2\sigma^2}$$

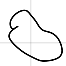
$$\Omega(i, i^*) = e^{-|r_i - r_{i^*}|^2 / 2\sigma^2}$$

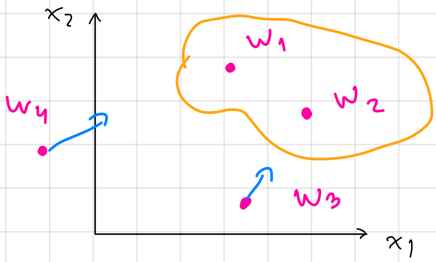
σ → varía dep. del tiempo, inicialmente grande



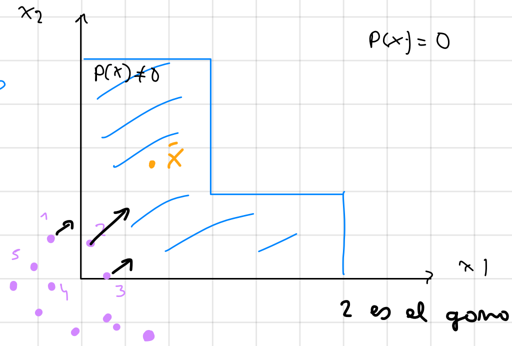
N - vectores $\bar{w}$

idealmente los vectores $\bar{w}$ apuntan hacia los $\bar{x}$

Supongamos que se tienen datos en 



los perdedores se mueven en la misma direc. pero mas lento



pueden de los datos de entrada

$P(x) = 0$

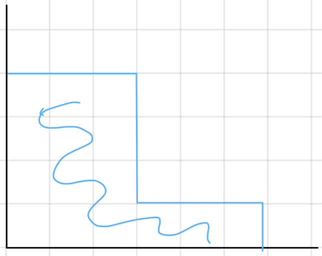
$N = 2$

$M = 10$

$d_M = 1$

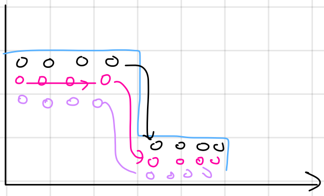
(cerca en el sentido del vector de salida)

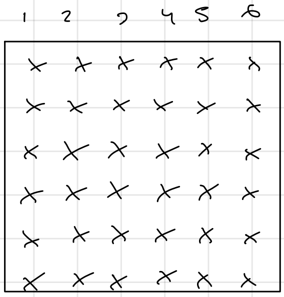
2 es el ganador, los vecinos también se mueven en esa direc



curvas de Peano
↓
intentan llenar una región de dim mayor

Los pesos en el espacio de salida están en 2D pero en una grilla cuad.





M = 36

noción de vecindad en este espacio

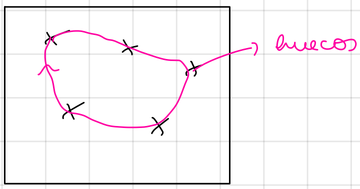
distr. de los pesos

se tiene que $P(\bar{w}) \neq P(\bar{x})$

$P(\bar{w}) \propto P(x)^{3/2}$ → calculo en el Hertz

Aplicación: soluciones a problemas de opt. combinatoria

ej: problema del viajante de comercio



se quiere encontrar el camino cerrado más corto que pase por todos los huecos

ver que ocurre si θ = cte, grande

los datos se mueven como en bloque