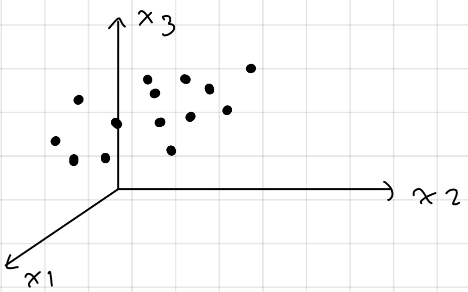
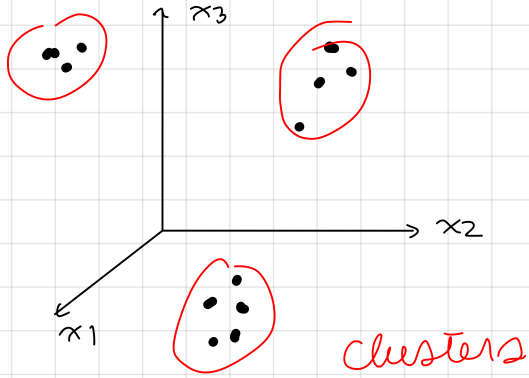
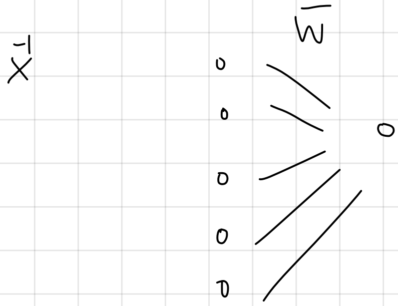
Aprendizaje no supervisado

- los datos de entrada tienen una estructura estadística no trivial

- se impone alguna regla de modificación de las conexiones sinápticas







$$V = \sum_{j=1}^{N} w_j x_j = \vec{w} \cdot \vec{x}$$

Postulo $\boxed{\Delta w_i = \eta V x_i}$ ⇒ ¿Los pesos convergen?

$$0 = \underbrace{\langle \Delta w_i \rangle}_{\text{los pesos no se modifican más}} = \eta \langle V x_i \rangle = \eta \left\langle \sum_{j=1}^{N} w_j x_i x_j \right\rangle$$

$$= \eta \sum_{j=1}^{N} w_j \underbrace{\langle x_i x_j \rangle}_{C_{ij}}$$

C_{ij}
↓
es simétrica
↓
los elementos diagonales son 0
⇒ C_i

$$\boxed{\langle \Delta \vec{w} \rangle = \bar{\bar{C}} \, \vec{w} \, \eta}$$

(notación matricial)

$\bar{V}_\alpha$: autovectores de $\bar{\bar{C}}$ / $\bar{\bar{C}}\,\bar{V}_\alpha = \lambda_\alpha \bar{V}_\alpha$

λ_α : autovalores de $\bar{\bar{C}}$

si $\lambda_\alpha \neq \lambda_\beta \Rightarrow \bar{V}_\alpha \cdot \bar{V}_\beta = 0$

$$\underset{\substack{\downarrow \\ \text{expansión de } W \text{ en autovectores de } \bar{V}_\alpha}}{\bar{W}} = \sum_{\alpha=1}^{N} W_\alpha \bar{V}_\alpha$$

$$\Rightarrow \bar{\bar{C}}\,\bar{W} = \sum_{\alpha=1}^{N} W_\alpha \bar{\bar{C}}\,\bar{V}_\alpha = \sum_{\alpha=1}^{N} W_\alpha \lambda_\alpha \bar{V}_\alpha$$

$$\Rightarrow \langle \Delta \bar{W} \rangle = \underbrace{\eta \sum_{\alpha=1}^{N} W_\alpha \lambda_\alpha \bar{V}_\alpha}_{\text{el cambio de } \Delta\bar{W} \text{ es en dirección del autovalor } \lambda_\alpha \text{ más grande}} \overset{?}{=} 0$$

no! (no converge)

$\downarrow$

N veces : $\langle \Delta \bar{W} \rangle = \eta \sum_{\alpha=1}^{N} W_\alpha \lambda_\alpha^{N} \bar{V}_\alpha$

a tiempos largos W apunta en dirección de máximo autovalor

Postulemos ahora $\boxed{\Delta W_i = \eta V (x_i - V W_i)}$ regla de Oja

converge!

Calculemos
$$\langle \Delta W_i \rangle = \eta \left[\sum_j W_j \langle x_i x_j \rangle - \left\langle \left(\sum_j W_j x_j\right)^2 W_i \right\rangle \right]$$
$$= \eta \left[\sum_j C_j W_j - \sum_{j,j'} W_j W_{j'} \underbrace{\langle x_j, x_{j'} \rangle}_{C_{jj'}} W_i \right]$$
$$= \eta \left[\sum_j C_{ij} W_j - \left[\sum_{j,j'} W_j C_{jj'} W_{j'} \right] W_i \right]$$

$$\Rightarrow \langle \Delta \bar{W} \rangle_i := \eta \left[(\bar{\bar{C}}\,\bar{W})_i - [(\bar{W}^T \bar{\bar{C}}\,\bar{W})\,\bar{W}_i] \right]$$

$\Rightarrow$ se tiene convergencia si $\bar{\bar{C}}\,\bar{W} - [\bar{W}^T \bar{\bar{C}}\,\bar{W}]\,\bar{W} = 0$

si $\bar{W}$ / $\bar{\bar{C}}\,\bar{W} = \lambda \bar{W}$ $\Rightarrow$ $\lambda \bar{W} - [\lambda |\bar{W}|^2]\,\bar{W} = 0$

$$d\bar{w}\left[1-|\bar{w}|^2\right]=0$$

$\Rightarrow$ hay convergencia si $\bar{w}$ es autovector de $\bar{\bar{C}}$ con norma 1. $(|\bar{w}|^2=1)$

Se puede pensar a el punto donde converge el sist como un punto fijo. Evaluemos su estabilidad

$$\bar{W}=\bar{V}_\alpha+\bar{\varepsilon} \qquad \bar{\bar{C}}\bar{V}_\alpha=d\bar{V}_\alpha \qquad (|\bar{V}_\alpha|^2=1) \qquad |\bar{\varepsilon}|\ll 1$$

$$\langle\Delta\bar{\varepsilon}\rangle=\langle\Delta\bar{w}\rangle=\eta\left[(\bar{\bar{C}}\bar{w})-\left[(\bar{w}^T\bar{\bar{C}}\bar{w})\bar{w}\right]\right]$$

$$=\eta\left[\bar{\bar{C}}(\bar{V}_\alpha+\bar{\varepsilon})-\left[(\bar{V}_\alpha+\bar{\varepsilon})^T\bar{\bar{C}}(\bar{V}_\alpha+\bar{\varepsilon})\right](\bar{V}_\alpha+\bar{\varepsilon})\right]$$

$$=\eta\left[\bar{\bar{C}}\bar{V}_\alpha+\bar{\bar{C}}\bar{\varepsilon}-(\bar{w}^T\bar{\bar{C}}\bar{V}_\alpha)(\bar{V}_\alpha+\bar{\varepsilon})(\bar{V}_\alpha+\bar{\varepsilon})\right]$$

$$=\eta\left[\bar{\bar{C}}\bar{V}_\alpha-(\bar{w}^T\bar{\bar{C}}\bar{V}_\alpha)\bar{V}_\alpha\right]$$

$$+\eta\left[\bar{\bar{C}}\bar{\varepsilon}-\underbrace{\left[\bar{V}_\alpha^T\underbrace{\bar{\bar{C}}\bar{V}_\alpha}_{d_\alpha\bar{V}_\alpha}\right]}_{d_\alpha}\bar{\varepsilon}-2\left[\bar{\varepsilon}^T\underbrace{\bar{\bar{C}}\bar{V}_\alpha}_{d_\alpha\bar{V}_\alpha}\right]\bar{V}_\alpha\right]$$

$$+\mathcal{O}(\varepsilon^2)$$

$$=\eta\left[\bar{\bar{C}}\bar{\varepsilon}-d_\alpha\bar{\varepsilon}-2d_\alpha(\bar{\varepsilon}^T\bar{V}_\alpha)\bar{V}_\alpha\right]$$

$$\bar{V}_\beta^T\langle\Delta\bar{\varepsilon}\rangle=\eta\left[\bar{V}_\beta^T\bar{\bar{C}}\bar{\varepsilon}-2d_\alpha(\bar{\varepsilon}^T\bar{V}_\alpha)\bar{V}_\beta^T\bar{V}_\alpha-d_\alpha\bar{V}_\beta^T\bar{\varepsilon}\right]$$

$$=\eta\left[(\bar{V}_\beta^T\bar{\bar{C}}\bar{\varepsilon})-2d_\beta(\bar{\varepsilon}^T\bar{V}_\beta)\delta_{\alpha\beta}-d_\alpha(\bar{V}_\beta^T\bar{\varepsilon})\right]$$

$$=\eta\underbrace{\left[d_\beta-d_\alpha-2d_\beta\delta_{\alpha\beta}\right]}_{\substack{<0 \text{ estable}\\ >0 \text{ inestable}}}(\bar{V}_\beta^T\bar{\varepsilon})$$

si $\beta=\alpha \Rightarrow -2d_\alpha<0 \rightarrow$ estable

$\beta\neq\alpha \qquad d_\beta-d_\alpha$

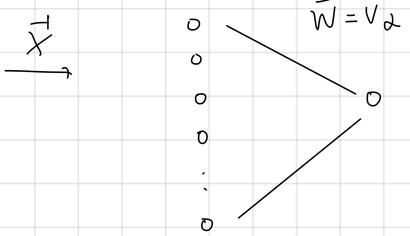
$\downarrow$

si $d_\alpha\geq d_\beta \ \forall\beta$

$\Rightarrow d_\beta-d_\alpha<0$

$\Rightarrow$ el eq. es estable

los $\bar{w}$ convergen allí $\Leftarrow$ hay un solo pto fijo estable en el autovector con máximo autovalor $\leftarrow$ la única dirección en la que la perturbación es estable es la del máximo autovalor



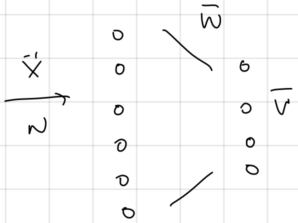
$\bar{V} = \bar{W} \cdot \bar{X}$
↓
proyección

$$\langle V^2 \rangle = \sum_{j,j'} W_j W_{j'} \langle x_j, x_{j'} \rangle$$
$$= \bar{W}^T \bar{\bar{C}}\, \bar{W} \qquad \left(\bar{W} = \sum_\alpha w_\alpha \bar{v}_\alpha\right)$$
$$= \sum_\alpha w_\alpha^2 \lambda_\alpha$$

⇒ la red estable es la que maximiza $\langle V^2 \rangle$
↓
máxima varianza ($\langle V \rangle = 0$)

función de costo $E = \frac{1}{2} \sum_{j,j'} W_j C_{ij} W_j + \frac{1}{4} \left(\sum_i W_i^2 \right)^2$

Generalización de la red de Oja



$\bar{V} = \bar{\bar{W}}\, \bar{X}$

$V_i = \sum_{j=1}^{N} W_{ij} x_j$

$i = 1, \dots, n$

la salida de la primera unidad va a dar la proyección de la entrada en dirección del autovalor más grande
la salida de la seg. unidad va a dar la proyección de la entrada en dirección de seg autovalor más grande

⋮ así siguiendo

PCA: principal component analysis