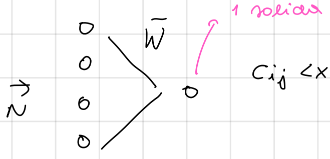
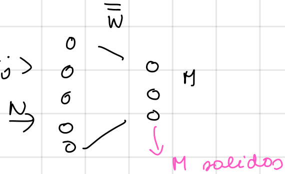


$C_{ij}\ \langle x_i, x_j\rangle$



Matriz de covarianza

$C_{ij} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle)\rangle$

↓

simétrica

$\boxed{\Rightarrow \text{tiene autovectores ortogonales}}$

Datos independientes - salida lineal

<u>Si no restamos la media</u>:

Típicamente el primer autovector va a apuntar en direc del valor medio y no en la direc donde la covarianza es mayor

autovector de $\bar{\bar C}$

$$\bar{\bar C} \cdot \bar V_\alpha = \lambda_\alpha \bar V_\alpha$$

$$\Rightarrow \sigma^2_{\hat v} \equiv \langle (\bar x^T \cdot \hat v)^2 \rangle = \langle \hat v^T (\bar x \bar x^T) \hat v \rangle = \langle \hat v^T \bar{\bar C} \hat v \rangle$$

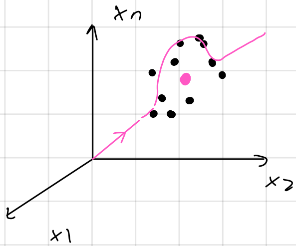
a $\bar x^T$ ya se le restó la media

$$\hat v = \sum_\alpha \bar V_\alpha c_\alpha \qquad \bar{\bar C}\hat v = \sum_\alpha \lambda_\alpha \bar V_\alpha e_\alpha$$

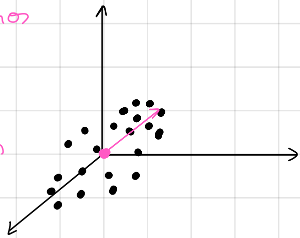
expandimos en la base de autovectores

$$\hat v^T = \sum_\alpha \bar V_\alpha^T e_\beta \quad \Rightarrow \quad \sigma^2_{\hat v} = \langle \hat v^T \bar{\bar C} \hat v \rangle = \sum \lambda_\alpha e_\alpha^2 \qquad \text{con } \sum_\alpha e_\alpha^2 = 1$$

$$\lambda_1 \geq \lambda_2 \geq \lambda_3 \dots \geq \lambda_N \qquad \lambda_1 = \lambda_{max}$$



corremos la nube al origen

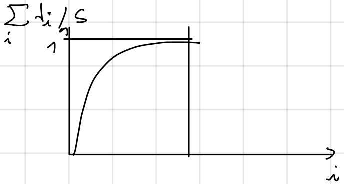


$$S = \sum_{i=1}^{N} \lambda_i$$

si $\frac{\lambda_1}{S} \to 1$ los datos tienen una forma extendida

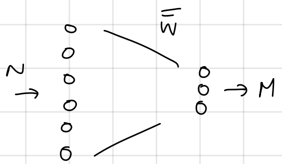
fracción de la varianza del primer autovector

$\frac{\lambda_1 + \lambda_2}{S}$: fracción de los primeros dos autovectores



Esto es completo si $P(\vec{x})$ es una dist de prob gaussiana pues todos los momentos se pueden calcular a partir de los dos primeros momentos

Principal problema en la memoria $|\underline{C_{ij}}|$ $\Rightarrow$ si $\overline{\overline{C}}$ es grande, calcular los autovectores es costoso, y no es necesario calcularlos todos

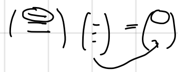


$$\overline{\overline{W}} \in N \times M$$

$$V_i = \sum_{j=1}^{N} W_{ij} x_j \qquad i = 1, \dots, M$$

$$= \overline{W}_i^{T} \, \overline{x}$$

prod pto a pto



$$\overbrace{\phantom{w_{11}\ w_{21}}}^{W^T}$$

$$\begin{pmatrix} w_{11} & w_{21} & \dots & \\ w_{12} & w_{22} & \dots & \\ \vdots & & & \\ w_{1i} & w_{2i} & \dots & w_{ni} \\ \vdots & \vdots & & \vdots \end{pmatrix} \to \overline{W}^{T}_{i}$$

regla de sanger

$$\Delta W_{ij} = \eta \, V_i \left(x_j - \sum_{k=1}^{i} V_k W_{kj} \right) \quad \boxed{i = 1, \dots, M}$$

Regla de oja $i = 1$ $\quad \eta V_1 (x_j - V_1 W_{1j})$

$$\langle \Delta W_{ij} \rangle = \eta \left\langle \left(\sum_{p=1}^{N} W_{ip} x_p \right) x_j - \left(\sum_{p=1}^{N} w_{ip} x_p \right) \sum_{k=1}^{i} \left(\sum_{q=1}^{N} W_{kq} x_q \right) W_{kj} \right\rangle$$

$$= \eta \left\{ \sum_{p=1}^{N} W_{ip} C_{pj} - \sum_{k=1}^{i} \sum_{p,q=1}^{N} W_{ip} W_{kq} C_{pq} W_{kj} \right\}$$

$$\langle \Delta \overline{W}_i \rangle = \eta \left\{ \overline{\overline{C}} \, \overline{W}_i - \sum_{k=1}^{i} \left(\overline{W}_k^{T} \overline{\overline{C}} W_i \right) \overline{W}_k \right\}$$

$$= \eta \left\{ \overline{\overline{C}} \, \overline{W}_i - \sum_{k=1}^{i-1} \left[\overline{W}_k^{T} \overline{\overline{C}} \, \overline{W}_i \right] \overline{W}_k - \left[\overline{W}_i^{T} \overline{\overline{C}} \, \overline{W}_i \right] \overline{W}_i \right\}$$

si $i=1 \rightarrow$ regla de oja $\bar{w}_1 \propto \bar{v}_1$: autovector de $\bar{\bar{C}}$ con máximo autovalor

si $i=2$ $\langle \Delta \bar{w}_2 \rangle = \eta \{ \bar{\bar{C}} \bar{w}_2 - [\bar{w}_1^T \bar{\bar{C}} \bar{w}_2] \bar{w}_1 - [\bar{w}_2^T \bar{\bar{C}} \bar{w}_2] \bar{w}_2 \}$

$|\bar{w}_1|^2 = 1$ $= \eta \{ (\bar{\bar{C}} \bar{w}_2)^\perp - [\bar{w}_2^T \bar{\bar{C}} \bar{w}_2] \bar{w}_2 \}$

regla de oja en el subespacio ortogonal a $\bar{w}_1$

$\bar{\bar{C}}$ mantiene componentes de esp. ortogonales en dichos espacios (exponsición w_2 en la base de autovectores de $\bar{\bar{C}}$ y ficonsegue se verifica)

$w_2 = \bar{w}_2^\perp$ $\bar{\bar{C}} w_2^\perp = \lambda_2 \bar{w}_2^\perp \Rightarrow \langle \Delta \bar{w}_2 \rangle = \eta \{ \lambda \bar{w}_1^\perp - \lambda_2 \underbrace{[\bar{w}_2^{T\perp} \bar{w}_2^\perp]}_{=1} \bar{w}_2^\perp \} = 0$

λ_2 : autovalor más grande en el esp ortogonal a $\bar{v}_1$
$\Rightarrow$ es el seg. autovalor más grande

$$\begin{array}{ccccccc} \bar{w}_1 & \rightarrow & \bar{w}_2 & \rightarrow & \bar{w}_3 & \rightarrow & \dots \\ \downarrow & & \downarrow & & \downarrow & & \\ \lambda_1 & > & \lambda_2 & > & \lambda_3 & & \dots \end{array}$$

$\Rightarrow$ restamos de manera ordenada
{ ORTOGONALIZACIÓN ITERATIVO

- falta prueba estabilidad

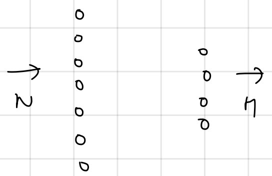
regla de sanger $\Delta w_{ij} = \eta V_i \left(x_j - \sum_{k=1}^{i} V_k w_{kj} \right)$

regla M de oja $\Delta w_{ij} = \eta V_i \left(x_i - \sum_{k=1}^{M} V_k w_{kj} \right)$

			SANGER	n-OJA
λ_1	$\bar{v}_1$	=	$\bar{w}_1$	$\bar{w}_1(\text{gen}(\bar{v}_1, \dots, \bar{v}_n))$
V				
λ_2	$\bar{v}_2$	=	$\bar{w}_2$	$\bar{w}_2(\text{gen}(\bar{v}_1, \dots, \bar{v}_n))$
V				
$\vdots$	$\vdots$		$\vdots$	$\vdots$
λ_n	$\bar{v}_n$	=	$\bar{w}_n$	$\bar{w}_n(\text{gen}(\bar{v}_1, \dots, \bar{v}_n))$

no ordenado

regla de cinsker → ver libro de Hertz



$$V_i = \sum_{j=1}^{N} w_{ij} x_j \quad (*)$$

$$\Delta w_{ij} = \eta \left(V_i x_j + b x_j + c V_i + d \right)$$

$b, c, d \rightarrow$ parámetros a elegir

de (*)

$$\langle \Delta w_{ij} \rangle = \eta \left\langle \sum_{l=1}^{N} w_{il} x_l x_j + b x_i + c \sum_{l=1}^{N} w_{il} x_l + d \right\rangle$$

$$= \eta \sum_{l=1}^{N} w_{il} C_{lj} + b \langle x_i \rangle + c \sum_{l=1}^{N} w_{il} \langle x_l \rangle + d$$

llamo $d_y = b \langle x_i \rangle + d$

el valor medio se convierte en un valor cte

↓

no depende del índice del vector $(\langle x_l \rangle = \langle x \rangle)$

$$= \eta \sum_{l=1}^{N} w_{il} C_{lj} + c \sum_{l=1}^{N} w_{il} \langle x_l \rangle$$

llamo $\lambda = -c \sum_{l} \langle x_l \rangle$

$$= \eta \left\{ \sum_{l} w_{il} C_{lj} + \lambda \Big[\underbrace{\mu - \sum_{l} w_{il}} \Big] \right\} \quad - \lambda \sum_{l} w_{il}$$

suma de las conexiones que van a la neurona i tienden al valor μ (equilibrio)

↓

tamaño global de las conexiones

$\bar{z}_i$: va gaussianas ind $\langle z_i z_j \rangle = \sigma^2 \delta_{ij}$ → matriz covarianza diagonal

tomamos el cambio de variables $\bar{x} = \bar{\bar{C}}^{1/2} \bar{z}$

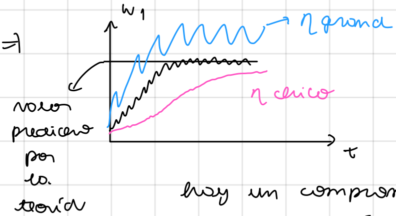
para obtener $\langle x_i x_j \rangle = C_{ij}$

comparar pesos con los autovectores de $\bar{\bar{C}}$

converge en media → cada vez que se toma un λ

↓ genera perturbaciones

⇒



hay un compromiso entre precisión y tiempo de cálculo