Aprendizado Auto Supervisionado

Representação por Componentes Principais

PCA - Principal Component Analysis

- Análise de Componentes Principais

- Transformada de Kahunen-Löeve

via Redes Neurais

$$\underline{x} \quad \Rightarrow \quad \underline{z} \quad \Rightarrow \quad \underline{\tilde{x}} \cong \underline{x}$$

$$\underline{x} \Rightarrow \underline{z} \Rightarrow \underline{\tilde{x}} \cong \underline{x}$$

Critérios e Propriedades da PCA

Redução de dimensionalidade,

Compressão da informação

$$\dim \underline{z} < \dim \underline{x}$$

Mapeamentos Lineares

$$\underline{z} = \underline{W}\,\underline{x} \qquad \underline{\tilde{x}} = \underline{T}\,\underline{z}$$

Representação ótima

$$\underline{W}, \underline{T} = \arg\ Min\ E\left(\left|\underline{x} - \underline{\tilde{x}}\right|^2\right)$$

Base Ortonormal

$$\underline{W}\,\underline{W}^t = \underline{I}$$

Descorrelação das componentes

$$E\left(z_i z_j\right) = 0 \qquad \forall i, j$$

Ordem de importância,

Potência das componentes

$$E\left(z_i^2\right) > E\left(z_j^2\right) \qquad \forall i < j$$

conseqüência:

$$\underline{z} = \underline{W}\,\underline{x} \qquad \underline{\tilde{x}} = \underline{T}\,\underline{z} = \underline{T}\,\underline{W}\,\underline{x}$$

$$se \quad \underline{\tilde{x}} \cong \underline{x} \quad então \quad \underline{T}\,\underline{W} \cong \underline{I} \quad e \quad \underline{T} \cong \underline{W}^{-1} \quad pseudo-inversa$$

$$mas \;\; se \quad \underline{W} \quad é \;\; uma \;\; base \;\; ortonormal$$

$$então \quad \underline{W}^{-1} = \underline{W}^{t} \qquad \log o \qquad \underline{T} \cong \underline{W}^{t}$$

e $$\underline{z} = \underline{W}\,\underline{x} \qquad \underline{\tilde{x}} = \underline{W}^{t}\,\underline{z}$$

Cálculo da PCA ($\underline{W}$) via Redes Neurais

Pré processamento

$$E(\underline{x}) = \underline{0} \qquad \Rightarrow \qquad E(\underline{z}) = E(\underline{\tilde{x}}) = \underline{0}$$

e de preferência $\sigma(x_i) \cong 1 \quad \forall\ i$

Implementação dos mapeamentos

$$\underline{z} = \underline{W}\,\underline{x}$$

$$\dim \underline{x} = n \qquad \dim \underline{z} = m \leq n$$

$$\underline{W} \quad é \quad m\,x\,n$$

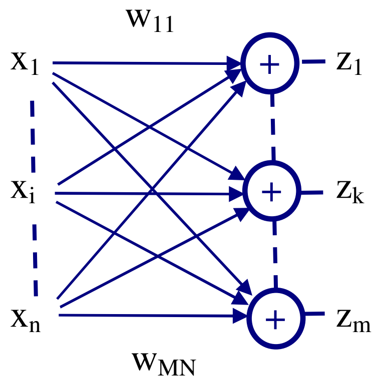
$$\underline{W} = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1n} \\ w_{21} & w_{22} & \dots & \\ \dots & & & \\ w_{m1} & \dots & & w_{mn} \end{bmatrix} = \begin{bmatrix} \underline{w}_1^t \\ \underline{w}_2^t \\ \dots \\ \underline{w}_m^t \end{bmatrix}$$

$$z_i = \underline{w}_i^t\,\underline{x} = \sum_{j=1}^{n} w_{ij} x_j$$

$$\underline{z} = \sum_{j=1}^{n} x_j\,\underline{w}_j$$

$$\underline{z} = \underline{W}\,\underline{x}$$

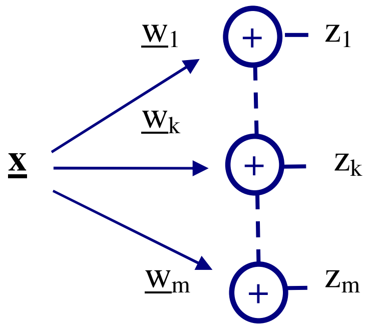
$$z_i = \underline{w}_i^t\,\underline{x}$$



In-star

de

Grossberg



$$\underline{\tilde{x}} = \underline{T}\,\underline{z} = \underline{W}^t\,\underline{z}$$

$$\dim \underline{\tilde{x}} = n \qquad \dim \underline{z} = m \leq n$$

$$\underline{T} = \underline{W}^t = \begin{bmatrix} w_{11} & w_{21} & \dots & w_{m1} \\ w_{12} & w_{22} & \dots & \\ \dots & & & \\ w_{1n} & \dots & & w_{mn} \end{bmatrix}$$

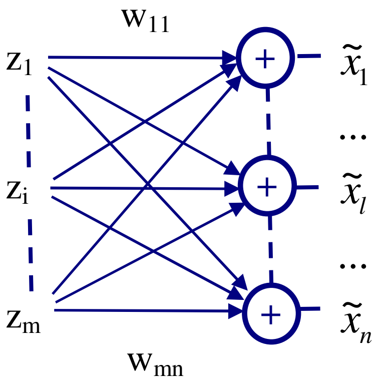
$$= \begin{bmatrix} \underline{w}_1 & \underline{w}_2 & \dots & \underline{w}_m \end{bmatrix}$$

$$\tilde{x}_i = \sum_{j=1}^{m} w_{ji} z_j$$

$$\underline{\tilde{x}} = \sum_{j=1}^{m} z_j\,\underline{w}_j$$

$$\underline{\tilde{x}} = \underline{W}^t \, \underline{z}$$

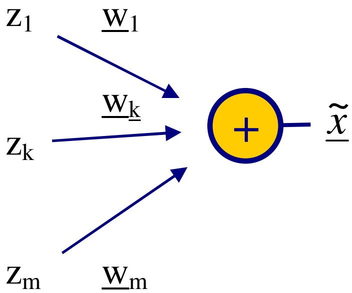
$$\underline{\tilde{x}} = \sum_{j=1}^{m} z_j \, \underline{w}_j$$



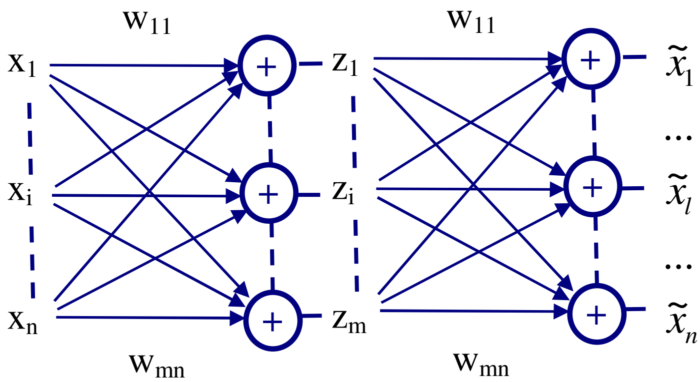
Out-star

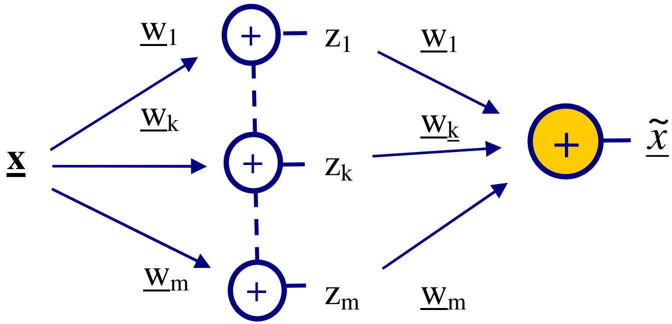
de

Grossberg



Mapeamento completo



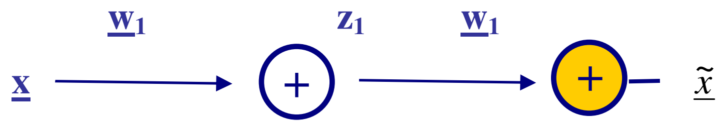


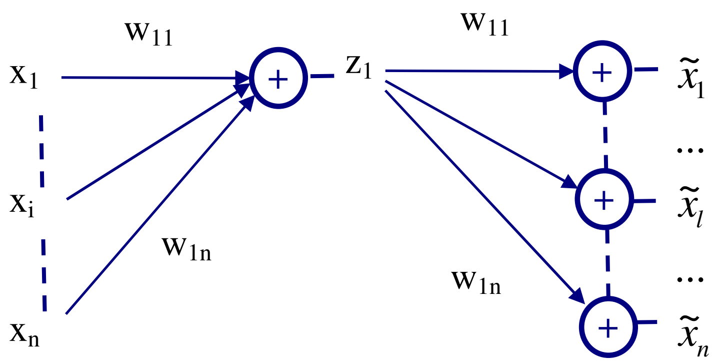
Primeira componente da base, $\underline{w}_1$

$$z_1 = \underline{w}^t \underline{x} = \sum_{j=1}^{n} w_{1j} x_j$$

$$\underline{\tilde{x}} = z_1 \underline{w}_1 = \underline{w}_1^t \underline{x} \underline{w}_1$$

$$\tilde{x}_i = w_{1i} z_1 = w_{1i} \sum_{j=1}^{n} w_{1j} x_j$$





Cálculo da primeira componente, $\underline{w}_1$

Critério

$$\underline{w}_1 = \arg\ Min\ E\left(\left|\underline{x} - \underline{\tilde{x}}\right|^2\right)$$

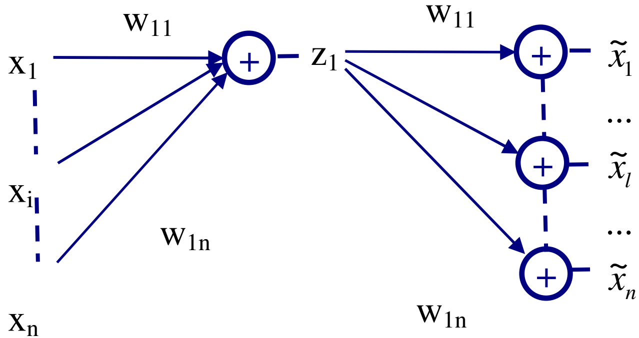
Método : Gradiente descendente

$$\Delta w_{1k} = -\alpha \frac{\partial}{\partial w_{1k}} E\left(\left|\underline{x} - \underline{\tilde{x}}\right|^2\right)$$

$$\left|\underline{x} - \underline{\tilde{x}}\right|^2 = \varepsilon^2 = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} \left(x_i - \tilde{x}_i\right)^2$$

$$\tilde{x}_i = z_1 w_{1i}$$

$$z_1 = \sum_{j=1}^{n} w_{1j} x_j$$



$$donde:$$

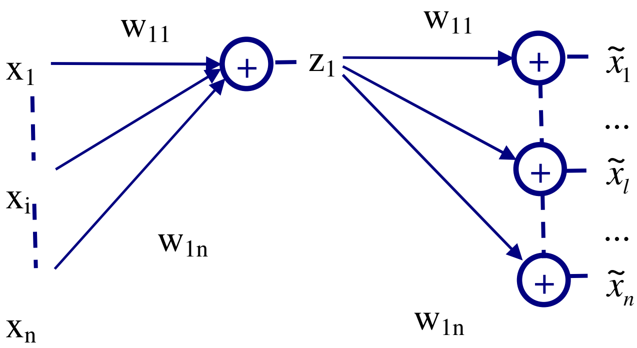
$$\Delta w_{1k} = 2\alpha E\left(z_1 \varepsilon_k + x_k \sum_{i=1}^{n} \varepsilon_i w_{1i} \right)$$

Treinamento da Primeira Componente, $\underline{w}_1$

Inicialização para a primeira componente:

$$\underline{w}_{1\ inicial} = \underline{r} \quad \text{randômico,} \quad |\underline{w}_{inicial}| \approx 1$$

os valores iniciais das sinapses equivalentes da primeira e segunda camadas <u>são obrigatóriamente iguais</u>



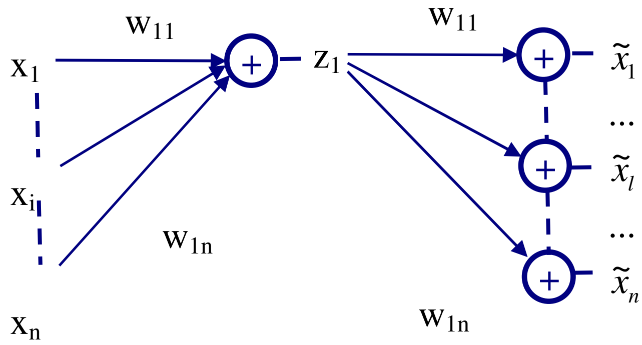
Treinamento da Primeira Componente $\underline{w}_1$

Para cada entrada $\underline{x}$

Signal feedforward,

saída,

erro



Acréscimos nas sinapses

$$\Delta w_{1k} = 2\alpha E\left(z_1 \varepsilon_k + x_k \sum_{i=1}^{n} \varepsilon_i w_{1i} \right)$$

Segunda componente, $\underline{w}_2$

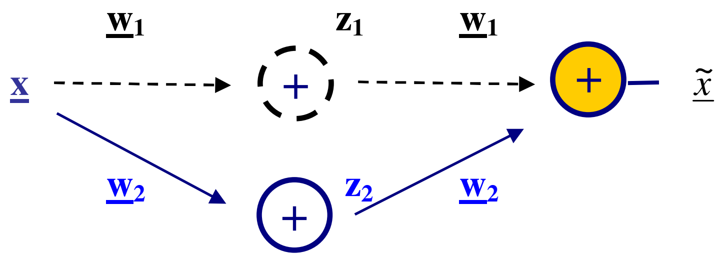
$$z_2 = \underline{w}_2^t \underline{x} = \sum_{j=1}^{n} w_{2j} x_j$$

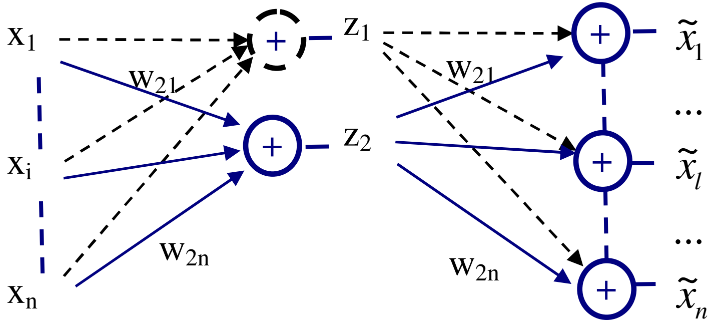
$$\underline{\tilde{x}} = z_1 \underline{w}_1 + z_2 \underline{w}_2$$

$$= z_1 \underline{w}_1 + \underline{w}_2^t \underline{x} \underline{w}_2$$

$$\tilde{x}_i = w_{1i} z_1 + w_{2i} z_2$$

$$= w_{1i} z_1 + w_{2i} \sum_{j=1}^{n} w_{2j} x_j$$





Cálculo da segunda componente, $\underline{w}_2$

Critério

$$\underline{w}_2 = \arg\ Min\ E\left(\left|\underline{x} - \underline{\tilde{x}}\right|^2\right) \quad com \quad \underline{w}_1 = \underline{w}_{1ótimo}$$

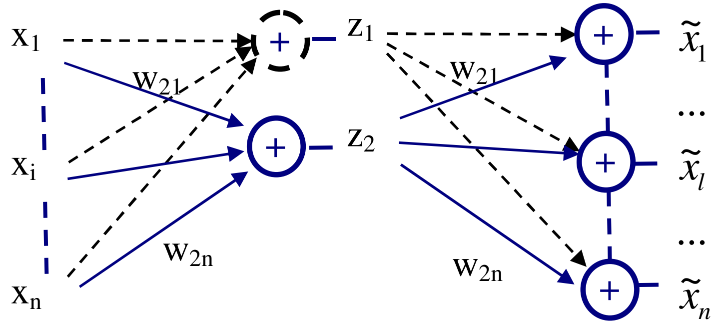
Método : Gradiente descendente

$$\Delta w_{2k} = -\alpha \frac{\partial}{\partial w_{2k}} E\left(\left|\underline{x} - \underline{\tilde{x}}\right|^2\right)$$

$$\left|\underline{x} - \underline{\tilde{x}}\right|^2 = \varepsilon^2 = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} \left(x_i - \tilde{x}_i\right)^2$$

$$\tilde{x}_i = z_1 w_{1i} + z_2 w_{2i}$$

$$z_2 = \sum_{j=1}^{n} w_{2j} x_j$$



donde :

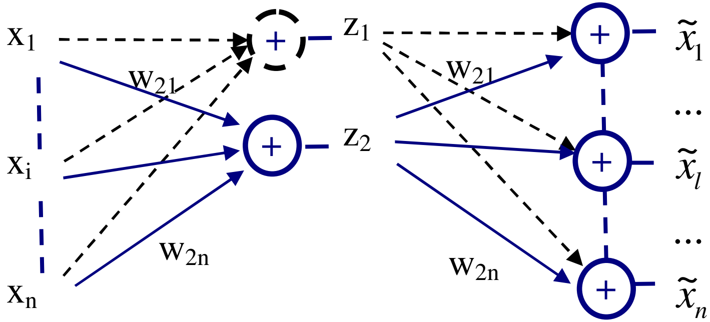
$$\Delta w_{2k} = 2\alpha E\left(z_2 \varepsilon_k + x_k \sum_{i=1}^{n} \varepsilon_i w_{2i} \right)$$

Treinamento da Segunda Componente

Inicialização para a segunda componente:

$$\underline{\mathbf{w}}_{2\ \text{inicial}} = \underline{\mathbf{r}} \quad \text{randômico}, \quad |\,\underline{\mathbf{w}}_{2\ \text{inicial}}\,| \approx 1$$

os valores iniciais das sinapses equivalentes da primeira e segunda camadas <u>são obrigatóriamente iguais</u>



Treinamento da Segunda Componente

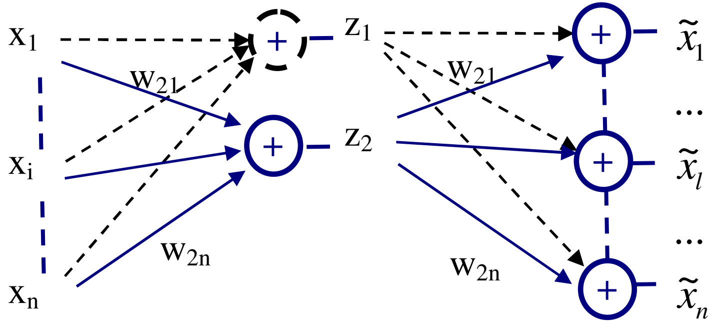
Congelar as sinapses anteriores $\mathbf{w}_1$

Para cada entrada $\underline{x}$

Signal feedforward,

saída,

erro



Acréscimos nas sinapses

$$\Delta w_{2k} = 2\alpha E\left(z_2 \varepsilon_k + x_k \sum_{i=1}^{n} \varepsilon_i w_{2i} \right)$$

Terceira componente $\underline{w}_3$

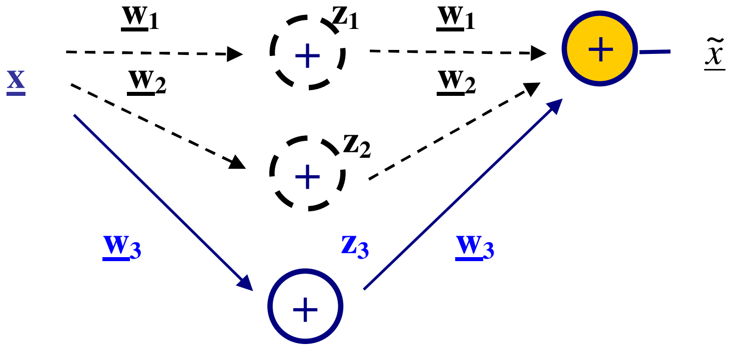
$$z_3 = \underline{w}_3^t \underline{x} = \sum_{j=1}^{n} w_{3j} x_j$$

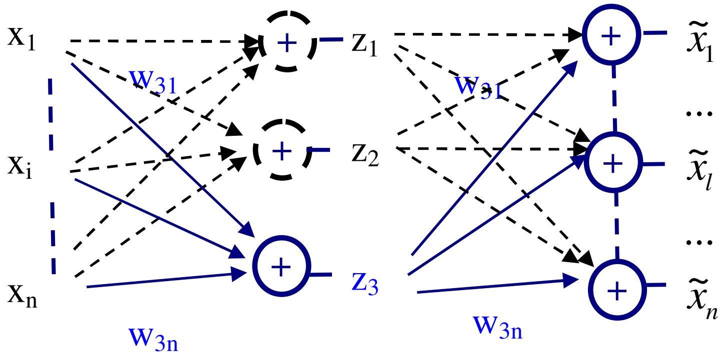
$$\underline{\tilde{x}} = z_1 \underline{w}_1 + z_2 \underline{w}_2 + z_3 \underline{w}_3$$

$$= z_1 \underline{w}_1 + + z_2 \underline{w}_2 + \underline{w}_3^t \underline{x} \underline{w}_3$$

$$\tilde{x}_i = w_{1i} z_1 + w_{2i} z_2 + w_{3i} z_3$$

$$= w_{1i} z_1 + w_{2i} z_2 + w_{3i} \sum_{j=1}^{n} w_{3j} x_j$$





Cálculo da terceira componente $\underline{w}_3$

Critério

$$\underline{w}_3 = \arg\ Min\ E\left(\left|\underline{x} - \underline{\tilde{x}}\right|^2\right)$$

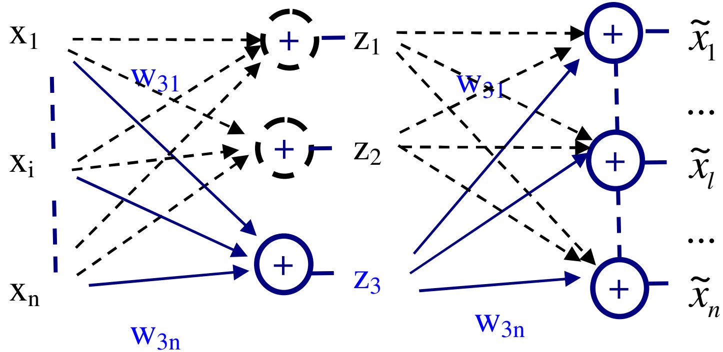
$$com \quad \underline{w}_1 = \underline{w}_{1ótimo} \quad e \quad \underline{w}_2 = \underline{w}_{2ótimo}$$

Método : Gradiente descendente

$$\Delta w_{3k} = -\alpha \frac{\partial}{\partial w_{3k}} E\left(\left|\underline{x} - \underline{\tilde{x}}\right|^2\right)$$

$$\left|\underline{x}-\underline{\tilde{x}}\right|^2 = \varepsilon^2 = \sum_{i=1}^{n}\varepsilon_i^2 = \sum_{i=1}^{n}\left(x_i-\tilde{x}_i\right)^2$$

$$\tilde{x}_i = z_1 w_{1i} + z_2 w_{2i} + z_3 w_{3i}$$



$$z_3 = \sum_{j=1}^{n} w_{3j} x_j$$

donde :

$$\Delta w_{3k} = 2\alpha E\left(z_3\varepsilon_k + x_k\sum_{i=1}^{n}\varepsilon_i w_{3i}\right)$$

Treinamento da Terceira Componente $\underline{w}_3$

Congelar as sinapses anteriores $\underline{w}_1$ e $\underline{w}_2$

Inicialização para a terceira componente $\underline{w}_3$:

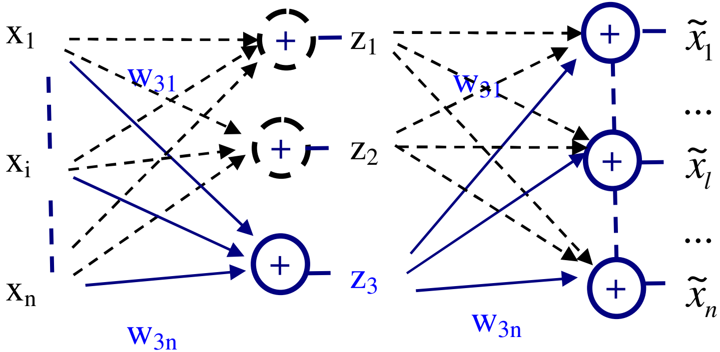
$\underline{w}_{3\,inicial} = \underline{r}$ randômico, $|\,\underline{w}_{3\,inicial}\,| \approx 1$

Para cada entrada $\underline{x}$

Signal feedforward,

saída,

erro

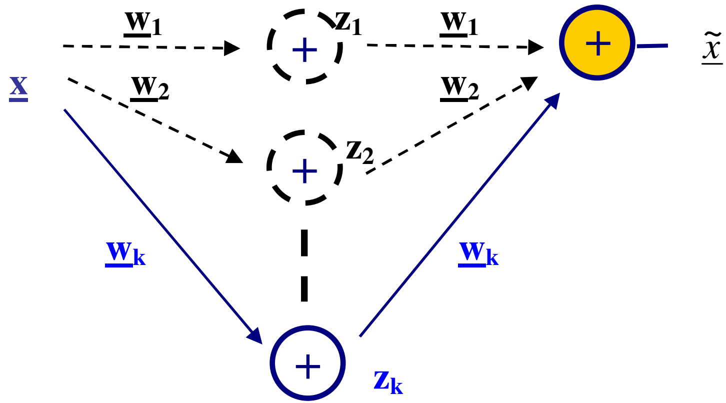


Acréscimos sinapses

$$\Delta w_{3k} = 2\alpha E\left(z_3 \varepsilon_k + x_k \sum_{i=1}^{n} \varepsilon_i w_{3i} \right)$$

k-ésima Componente

– análogo



Observação sobre a ortogonalidade das componentes:

Pelo processo de cálculo as componentes são naturalmente ortogonais.

utilizando apenas a primeira componente

$$\underline{\tilde{x}}_1 = \underline{w}_1^t \underline{x} \underline{w}_1 \qquad \underline{w}_1 = \arg\ Min\ E\left(\left|\underline{x} - \underline{\tilde{x}}\right|^2\right)$$

$$\underline{\varepsilon}_1 = \underline{x} - \underline{\tilde{x}}_1 = \underline{x} - \underline{w}_1^t \underline{x} \underline{w}_1 \qquad \underline{w}_1^t . \underline{\varepsilon}_1 = 0$$

utilizando as duas primeiras componentes

$$\underline{\tilde{x}}_2 = \underline{w}_1^t \underline{x}\underline{w}_1 + \underline{w}_2^t \underline{x}\underline{w}_2 \qquad \underline{w}_2 = \arg\Big|_{\underline{w}_1=\underline{w}_{1\,ótimo}} Min\ E\left(\left|\underline{x}-\underline{\tilde{x}}\right|^2\right)$$

$$\underline{\varepsilon}_2 = \underline{x} - \underline{\tilde{x}}_2 = \underline{x} - \underline{w}_1^t \underline{x}\underline{w}_1 - \underline{w}_2^t \underline{x}\underline{w}_2 = \underline{\varepsilon}_1 - \underline{w}_2^t \underline{x}\underline{w}_2$$

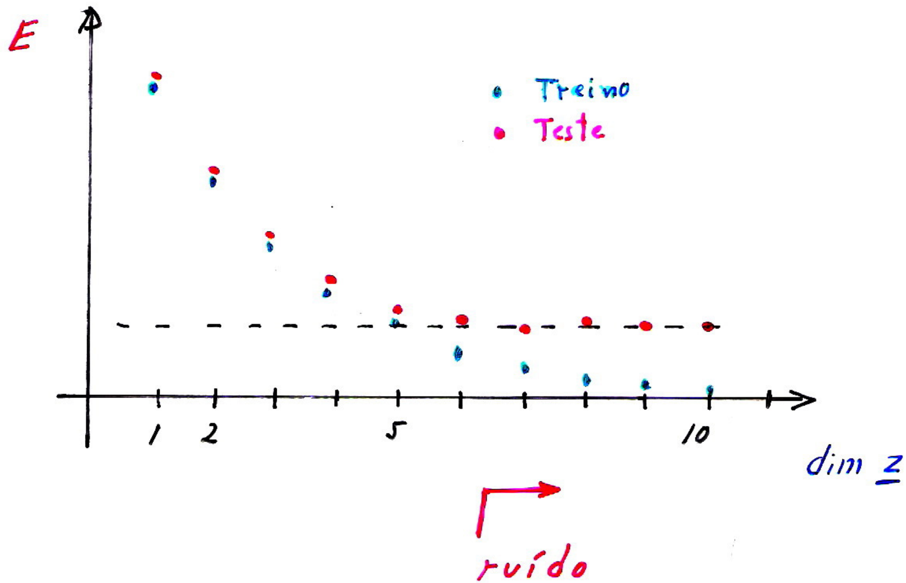
$$\underline{w_2^t}.\underline{\varepsilon}_2 = 0 \qquad \underline{w}_1^t.\underline{\varepsilon}_2 = 0$$

$$\underline{w}_1^t.\underline{\varepsilon}_2 = \underline{w}_1^t\left(\underline{\varepsilon}_1 - \underline{w}_2^t \underline{x}\underline{w}_2\right) = \underline{w}_1^t \underline{\varepsilon}_1 - \underline{w}_1^t\left(\underline{w}_2^t \underline{x}\right)\underline{w}_2 = 0 - \left(\underline{w}_2^t \underline{x}\right)\underline{w}_1^t \underline{w}_2 = 0$$

donde $\qquad \underline{w}_1^t \underline{w}_2 = 0$

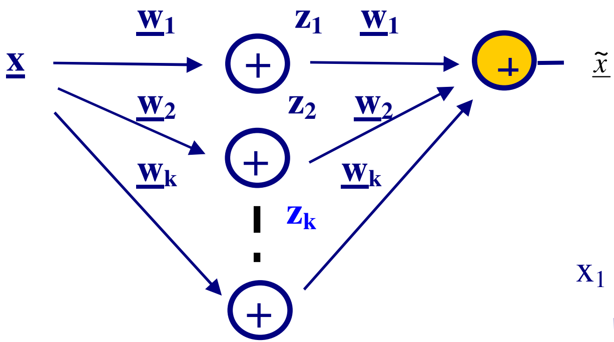
Erro $E\left(\|\underline{x}-\underline{\tilde{x}}\|^2\right)$ **vs.**

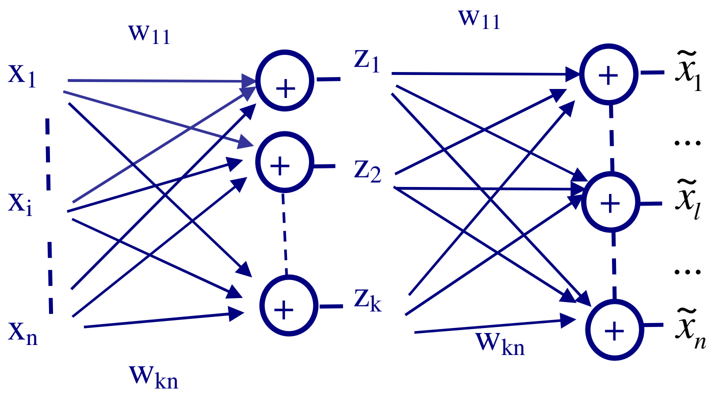
número de componentes = k = dim z



Processo alternativo:

Treinar todas as k componentes simultâneamente





Cálculo das componente, $\underline{w}_1$, ..., $\underline{w}_k$

Critério

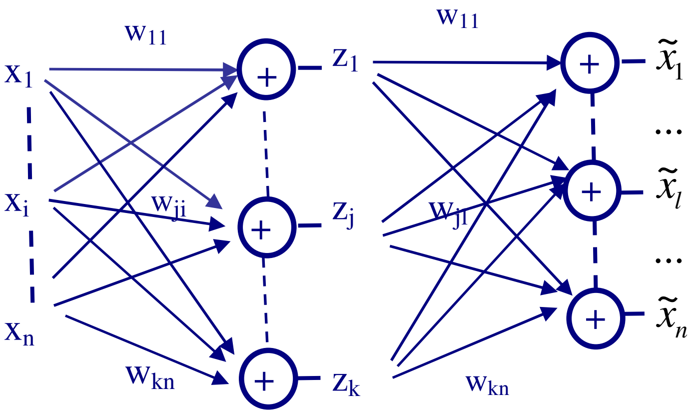
$$\{\underline{w}_1,\ldots,\underline{w}_k\} = \arg\ Min\ E\left(\left|\underline{x}-\underline{\tilde{x}}\right|^2\right)$$

Método : Gradiente descendente

$$\Delta w_{ij} = -\alpha \frac{\partial}{\partial w_{ij}} E\left(\left|\underline{x}-\underline{\tilde{x}}\right|^2\right)$$

$$\left|\underline{x}-\underline{\tilde{x}}\right|^2=\varepsilon^2=\sum_{l=1}^{n}\varepsilon_l^2=\sum_{l=1}^{n}\left(x_l-\tilde{x}_l\right)^2$$

$$\tilde{x}_l=\sum_{j=1}^{k}z_j w_{lj}$$



$$z_j=\sum_{i=1}^{n}w_{ji}x_i$$

donde :

$$\Delta w_{ij}=2\alpha E\sum_{i=1}^{n}\left(z_i\varepsilon_j+x_j\sum_{k=1}^{n}\varepsilon_k w_{ik}\right)$$

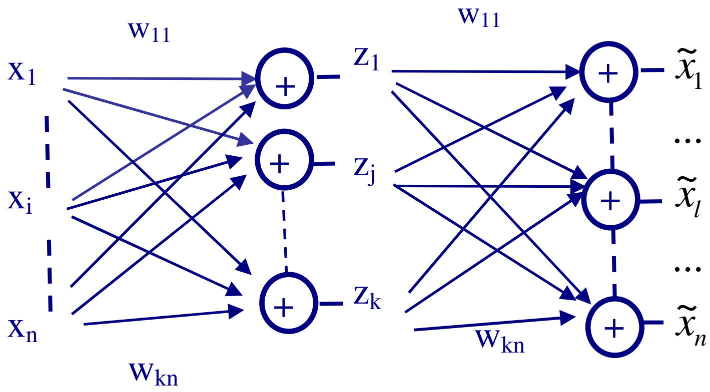
(à verificar)

Treinamento da Rede

Inicialização para cada componente $\underline{\mathbf{w}}_k$

$$\underline{\mathbf{w}}_{k\ \text{inicial}} = \underline{\mathbf{r}} \quad \text{randômico}, \quad |\, \underline{\mathbf{w}}_{k\ \text{inicial}} \,| \approx 1$$

os valores iniciais das sinapses equivalentes da primeira e segunda camadas <u>são obrigatóriamente iguais</u>



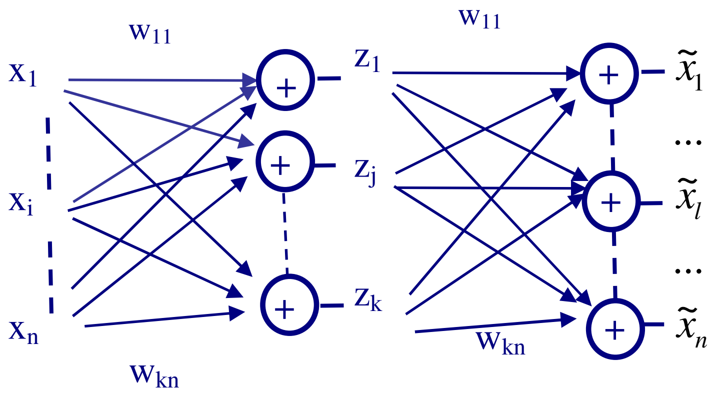
Treinamento das Componentes $\underline{w}_1, \ldots, \underline{w}_k$

Para cada entrada $\underline{x}$

Signal feedforward,

saída,

erro



Acréscimos nas sinapses

$$\Delta w_{ij} = 2\alpha E \sum_{i=1}^{n} \left(z_i \varepsilon_j + x_j \sum_{k=1}^{n} \varepsilon_k w_{ik} \right)$$

(à verificar)

Restrição

$$\left\{ \underline{w}_1, \ldots, \underline{w}_k \right\}$$

obtido é apenas uma base (não ortonormal, não ordenada) para o espaço das k primeiras componentes principais.