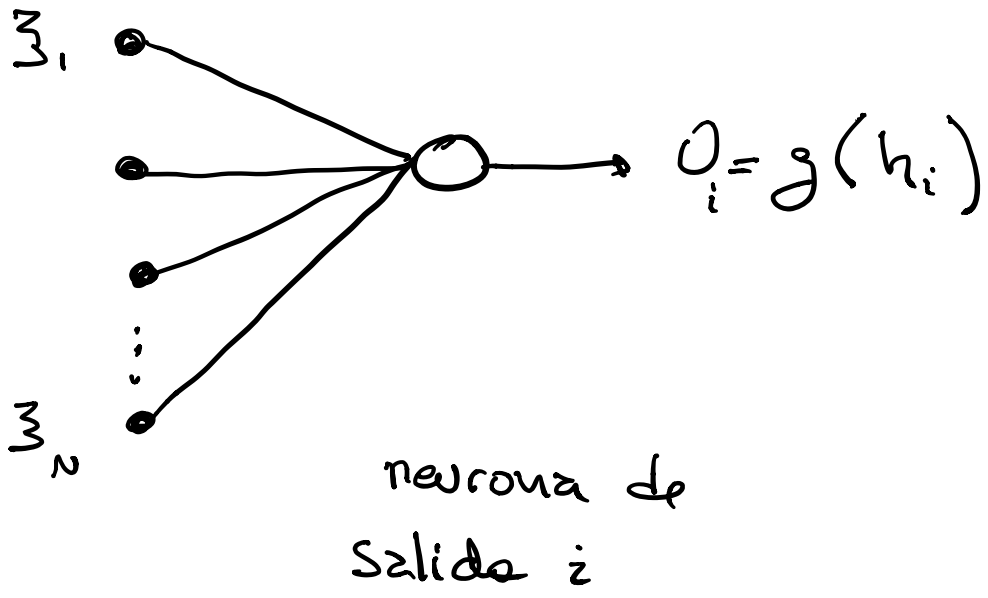
Seguimos con perceptron simple



Ayer vimos el caso $g(h) = \text{signo}(h)$

Si tenemos un conjunto de entrenamiento de P relaciones input-output correctas

$$\xi_i^\mu \longrightarrow \sigma_i^\mu \qquad \mu = 1, 2, \ldots P$$

Si el problema es linealmente separable, tenemos un algoritmo para $g(h) = \text{signo}(h)$

$$w_{ik}^{new} = w_{ik}^{old} + \eta \left(d_i^\mu - O_i^\mu \right) \xi_k^\mu$$

Neuronas de salida lineales

$$g(x) = x$$

$$O_i = \sum_k w_{ik} \xi_k$$

Otra vez asumimos que tenemos un conjunto de entrenamiento

$$\vec{\xi}^\mu \rightarrow d_i^\mu$$

$$d_i^\mu = \sum_k w_{ik} \xi_k^\mu$$

Dado un vector cualquiera $\bar{w}$ puedo calcular el error que comete el aprender

$$E(\bar{w}) = \frac{1}{2} \sum_{i\mu} (d_i^\mu - O_i^\mu)^2$$

$$= \frac{1}{2} \sum_{\mu} \sum_{i} (d_i^\mu - O_i^\mu)^2$$

$$E(\bar{w}) = \frac{1}{2} \sum_{\mu} \sum_{i} \left(d_i^\mu - \sum_k w_{ik} \xi_k^\mu\right)^2$$

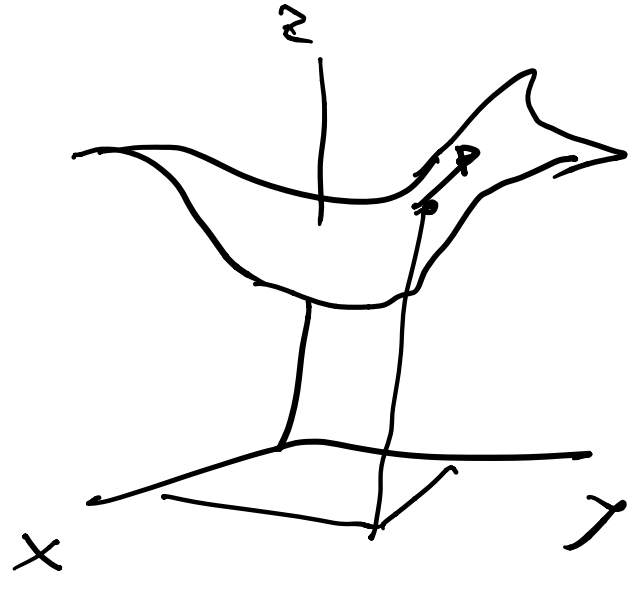
Agregué muchos renglones de salida etiquetados con i

Gradiente

Sea $f : \mathbb{R}^2 \to \mathbb{R}$

$$f(x_1, x_2) = x_1 + \operatorname{sen}(x_2)$$

$$\nabla f_{\bar{x}} = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2} \right)_{\bar{x}}$$



El gradiente define la dirección en un punto en $\mathbb{R}^2$ en la cual f crece más.

f decrece más rápido en

la dirección $-\vec{\nabla} f$

En el más general

$$f(x_1, x_2, \ldots x_N)$$

En nuestro caso

$$E(\tilde{w}) = E(w_{11}, w_{12} \ldots w_{21}, w_{22} \ldots \ldots w_{Nm})$$

m: nº de neuronas de salida

$$\nabla E_{ik} = \frac{\partial E}{\partial w_{ik}}$$

$$w_{ik}^{new} = w_{ik}^{old} + \Delta w_{ik}$$

$$\Delta w_{ik} = -\eta \frac{\partial E(\vec{w})}{\partial w_{ik}}$$

$$\Delta w_{ik} = +\eta \sum_{\mu} \left(\zeta_i^{\mu} - \sum_k w_{ik} \xi_k^{\mu} \right) \xi_k^{\mu}$$

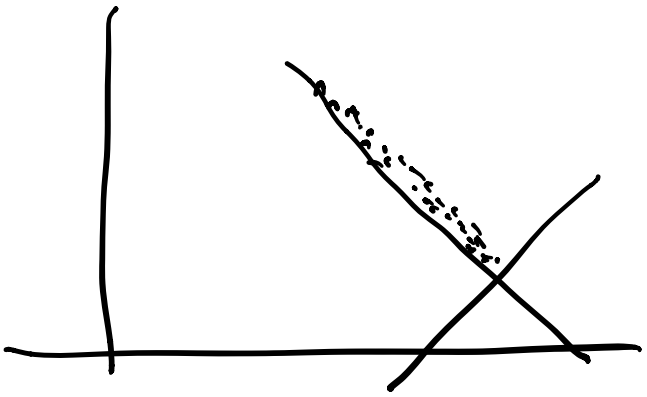
$$= \eta \sum_{\mu} \left(\zeta_i^{\mu} - O_i^{\mu} \right) \xi_k^{\mu}$$

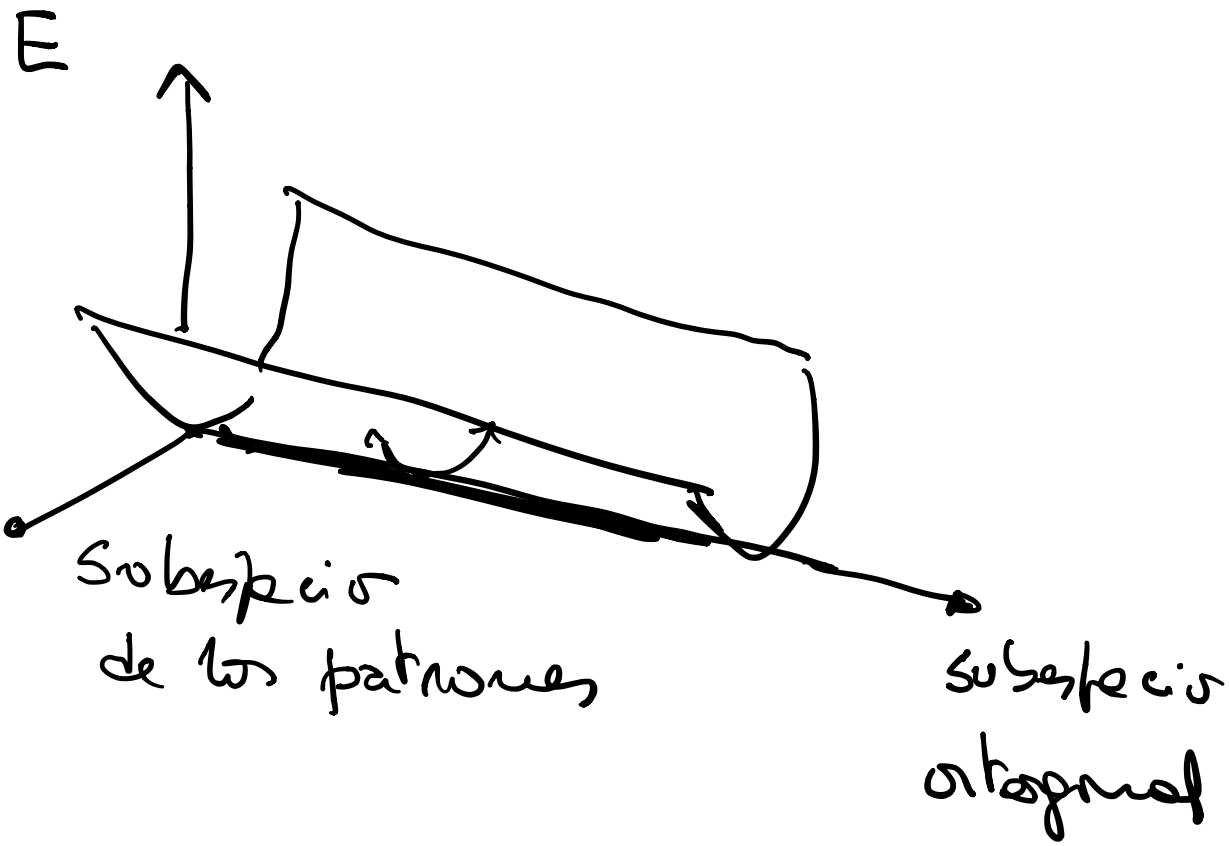
$$\Delta w_{ik} = \eta \sum_{\mu} \delta_i^{\mu} \xi_k^{\mu}$$

$$\delta_i^{\mu} = \left(\zeta_i^{\mu} - O_i^{\mu} \right)$$

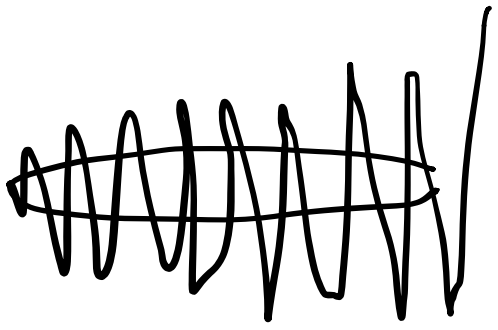
Se la conoce como

- regla delta
- regla adeline
- Widrow - Hoff
- regla LMS





En general los mínimos no
son puntos en $\mathbb{R}^n$ sino subespacios



El libro fuerza la convergencia

———— o ————

Unidades de salida no lineales

$g(h)$ no lineal

- $g(h) = \dfrac{1}{1 + e^{-\beta h}}$

- $g(h) = \tanh(\beta h)$

la topología es la misma

Tenemos ejemplos correctos

$$E(\vec{w}) = \frac{1}{2} \sum_{\mu} \sum_{i} \left(\zeta_i^{\mu} - O_i^{\mu}\right)^2$$

$$= \frac{1}{2} \sum_{\mu} \sum_{i} \left[\zeta_i^{\mu} - g\left(\sum_k w_{ik} \xi_k^{\mu}\right)\right]^2$$

$$w_{ik}^{new} = w_{ik}^{old} + \Delta w_{ik}$$

$$\Delta w_{ik} = -\eta \frac{\partial E(\vec{w})}{\partial w_{ik}}$$

$$\frac{\partial E(\bar{w})}{\partial w_{ik}} = -\frac{1}{2} \sum_{\mu} 2 \left[d_i^{\mu} - O_i^{\mu} \right] g'(h_i^{\mu}) \xi_k^{\mu}$$

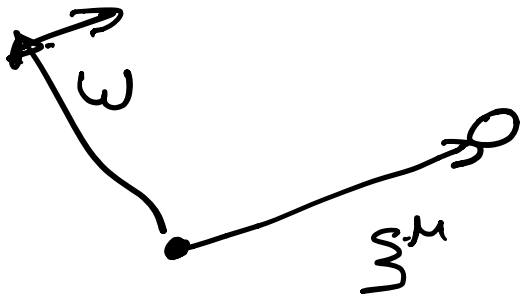
$$\Delta w_{ik} = -\eta \frac{\partial E(\bar{w})}{\partial w_{ik}}$$

$$\Delta w_{ik} = \eta \sum_{\mu} \overbrace{\left[d_i^{\mu} - O_i^{\mu} \right] g'(h_i^{\mu})}^{\delta_i^{\mu}} \xi_k^{\mu}$$

$$O_i^{\mu} = g\left(\sum_k w_{ik} \xi_k^{\mu} \right)$$

$$\Delta w_{ik} = \eta \sum_{\mu} \delta_i^{\mu} \xi_k^{\mu}$$

Δw me acerca a ξ^μ



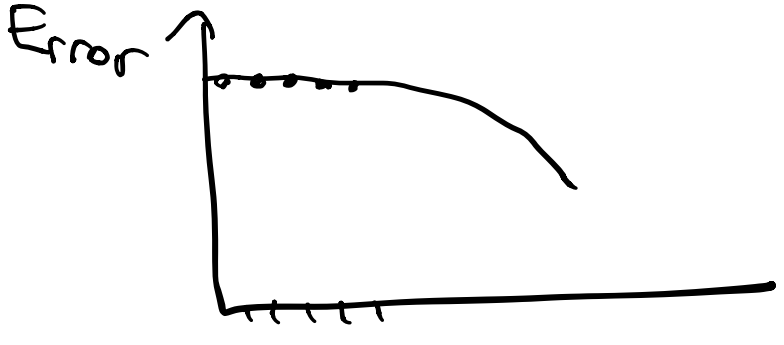
$$\Delta \vec{w}_i = \eta \sum_\mu \delta_i^\mu \vec{\xi}^\mu$$

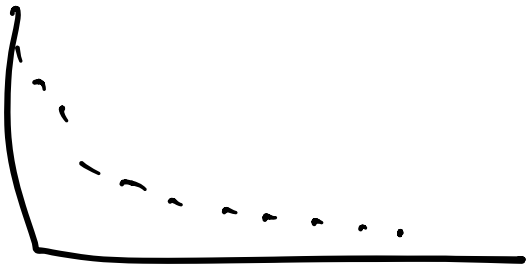
Si $g = \tanh(\beta h)$

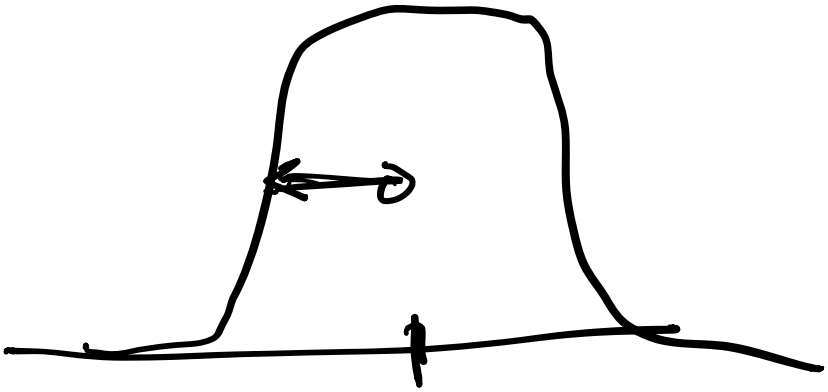
$$g' = \beta(1 - g^2)$$

Si $g = \frac{1}{1 + e^{-2\beta h}}$

$$g' = 2\beta g(1 - g)$$







Una medida mejor del error
la de la función

$$E(\bar{w}) = \sum_{i\mu} \left[\frac{1}{2}(1+d_i^\mu) \log\left(\frac{1+d_i^\mu}{1+O_i^\mu}\right) + \right.$$

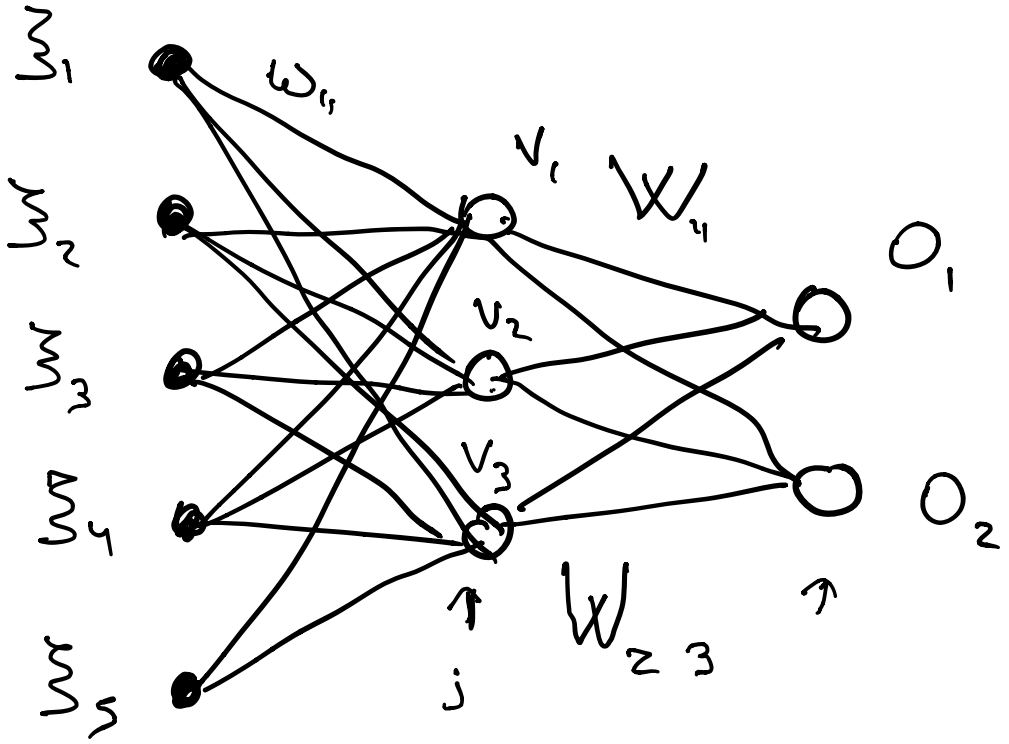
$$\left. + \frac{1}{2}(1-d_i^\mu) \log\left(\frac{1-d_i^\mu}{1-O_i^\mu}\right) \right]$$

Es bueno usar números aleatorios

Calculamos el h y fabricamos una probabilidad p

Perceptron multi capas

Supongamos una capa oculta



$$\xi_k \quad \omega_{jk} \quad V_j \quad W_{ij} \quad O_i$$

Tenemos p ejemplos correctos

Presentamos un ejemplo μ

Calculamos todos los V_j

$$h_j^\mu = \sum_k w_{jk} \xi_k^\mu$$

$$V_j^\mu = g(h_j^\mu) = g\left(\sum_k w_{jk} \xi_k^\mu\right)$$

Calculamos los salidas

$$h_i^\mu = \sum_j W_{ij} V_j$$

$$= \sum_j W_{ij}\, g\left(\sum_k w_{jk} \xi_k^\mu\right)$$

La salida i será

$$O_i^\mu = g(h_i^\mu) = g\left(\sum_j W_{ij} V_j\right)$$

$$O_i^\mu = g\left(\sum_j W_{ij}\, g\left(\sum_k w_{jk}\, \xi_k^\mu\right)\right)$$

Acá suponemos que las 2 capas tienen función g

Definimos una función error

$$E(\bar{w}) = \frac{1}{2}\sum_\mu \sum_i \left(d_i^\mu - O_i^\mu\right)^2$$

$$= \frac{1}{2}\sum_\mu \sum_i \left(d_i^\mu - g\left(\sum_j W_{ij}\, g\left(\sum_k w_{jk}\, \xi_k^\mu\right)\right)\right)^2$$

$$\Delta W_{ij} = -\eta \frac{\partial E}{\partial W_{ij}}$$

$$= \eta \sum_{\mu} [d_i^{\mu} - O_i^{\mu}] g'(h_i^{\mu}) V_j^{\mu}$$

$$= \eta \sum_{\mu} \delta_i^{\mu} V_j^{\mu}$$

$$\delta_i^{\mu} = [d_i^{\mu} - O_i^{\mu}] g'(h_i^{\mu})$$

Vamos a la otra capa de acoplamientos

$$\Delta w_{jk} = -\eta \frac{\partial E}{\partial w_{jk}}$$

$$= -\eta \sum_{\mu} \frac{\partial E}{\partial V_j} \cdot \frac{\partial V_j}{\partial w_{jk}}$$

$$= \eta \sum_{\mu i} (d_i^{\mu} - O_i^{\mu}) g'(h_i^{\mu}) \times$$

$$\times W_{ij} g'(h_j^{\mu}) \xi_k^{\mu}$$

$$= \eta \sum_{\mu} \sum_{i} S_i^{\mu} W_{ij} g'(h_j^{\mu}) \xi_k^{\mu}$$

$$= \eta \sum_{\mu} S_j^{\mu} \xi_k^{\mu}$$

donde $S_j^{\mu} = g'(h_j^{\mu}) \sum_i W_{ij} S_i^{\mu}$

la información va de entrada a salida

la corrección va de la salida a la entrada

BACK PROPAGATION

Ejemplo

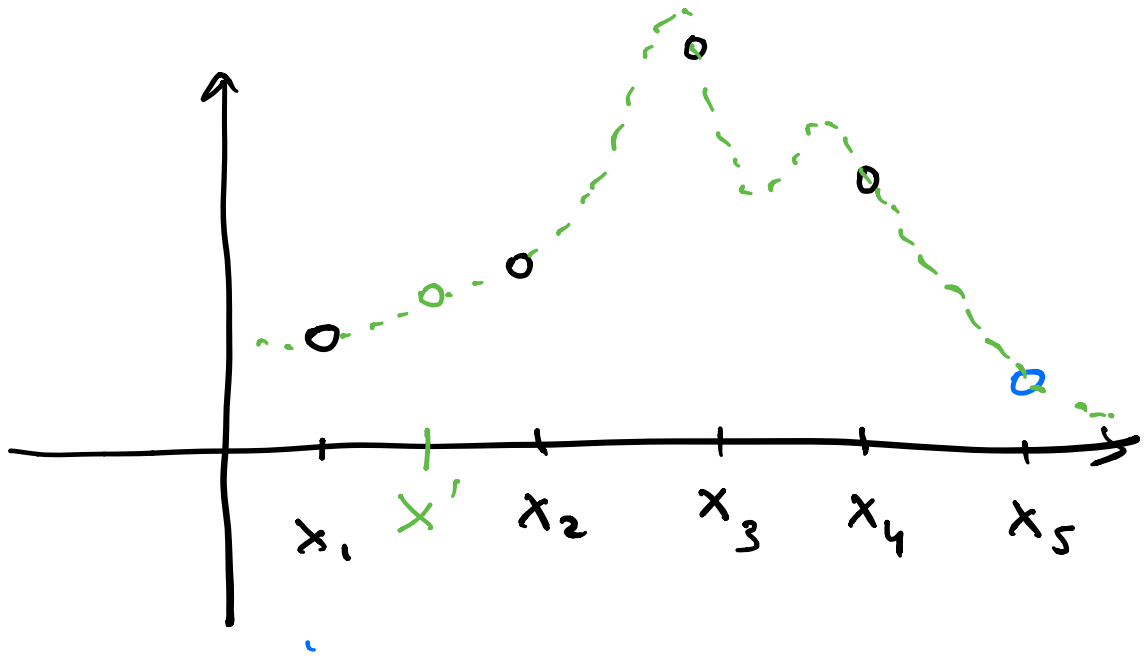
$x_1 \longrightarrow y_1$

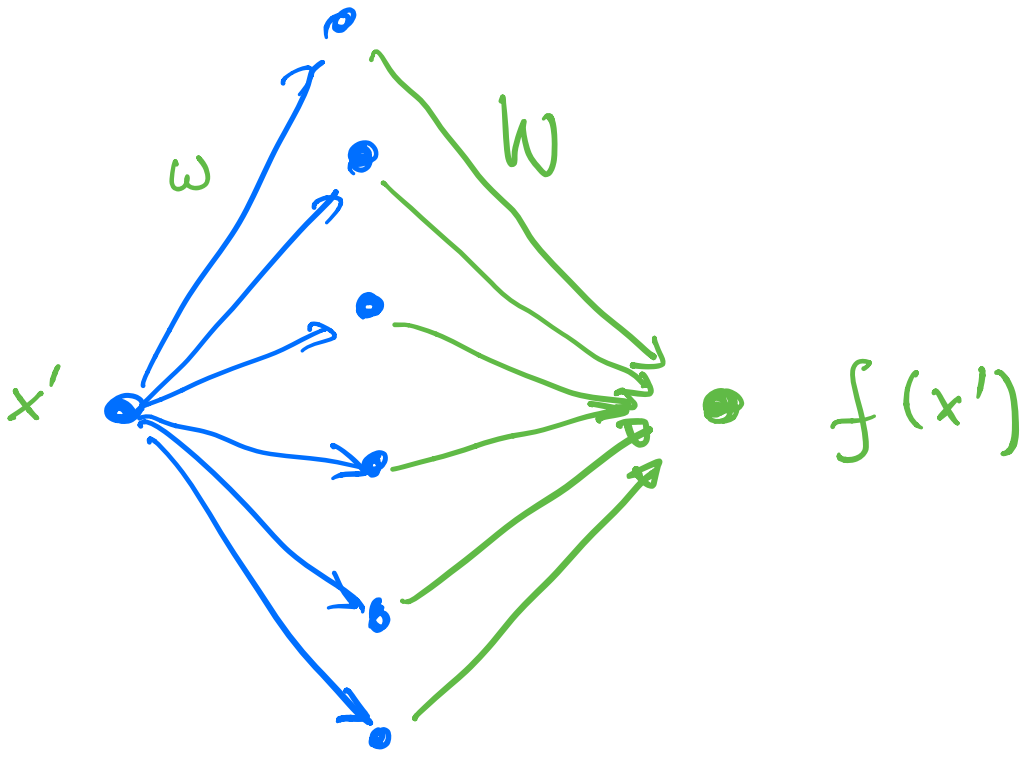
$x_2 \longrightarrow y_2$

$x_3 \longrightarrow y_3$

$x_4 \longrightarrow y_4$

$x_5 \longrightarrow y_5$





AUTO ENCODER