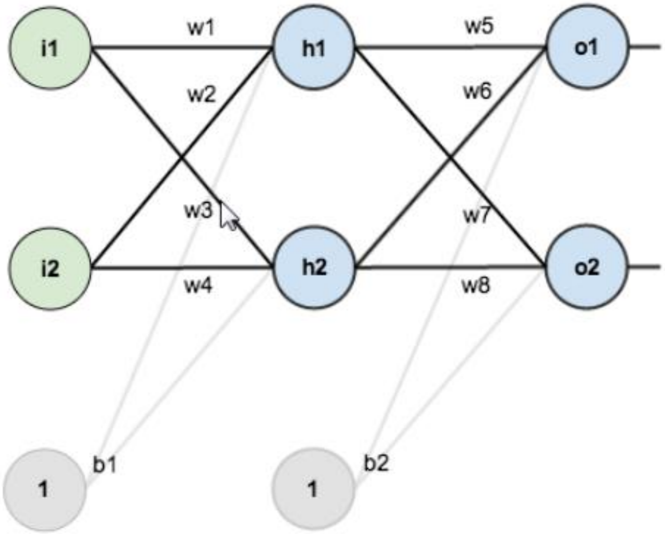
Backpropagation: paso a paso

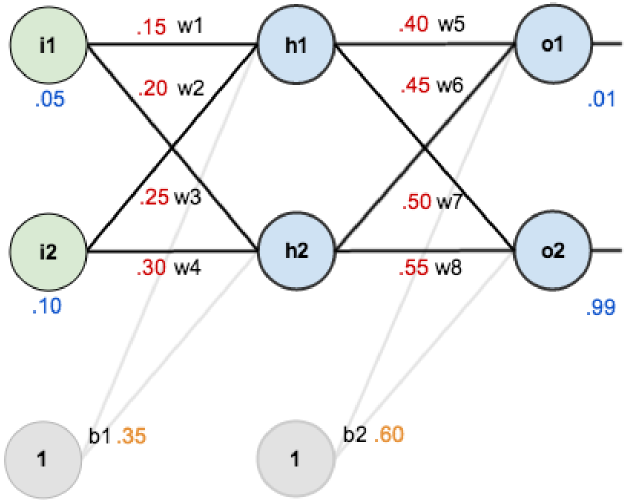
Ing. Juan M. Rodríguez

Perceptrón multicapa (fully connected)



- i : entradas (inputs)
- h : neuronas primera capa
- o : neuronas de salida
- b1 y b2 : Umbrales (bias)

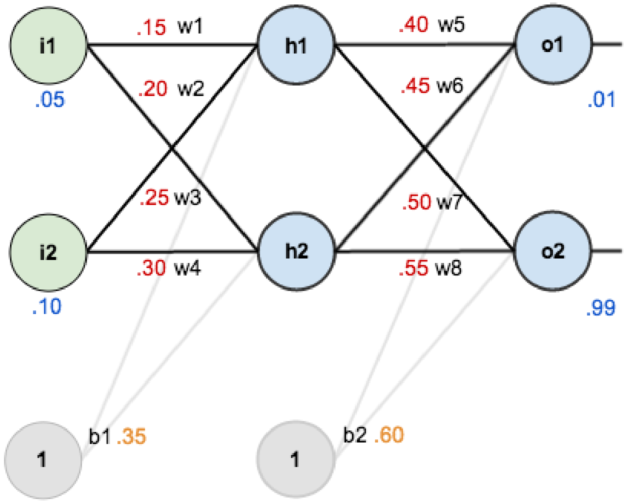
Perceptrón multicapa (fully connected)



- i : entradas (inputs)
- h : neuronas primera capa
- o : neuronas de salida
- b1 y b2 : Umbrales (bias)

Se inicializa con pesos aleatorios

Perceptrón multicapa (fully connected)



Entrada		Salida - esperada	
i1	**i2**	**o1**	**o2**
0.05	0.10	0.01	0.99

Función de activación

$$f(x)=\frac{1}{1+e^{-x}}$$

Calculamos la entrada para **h1**

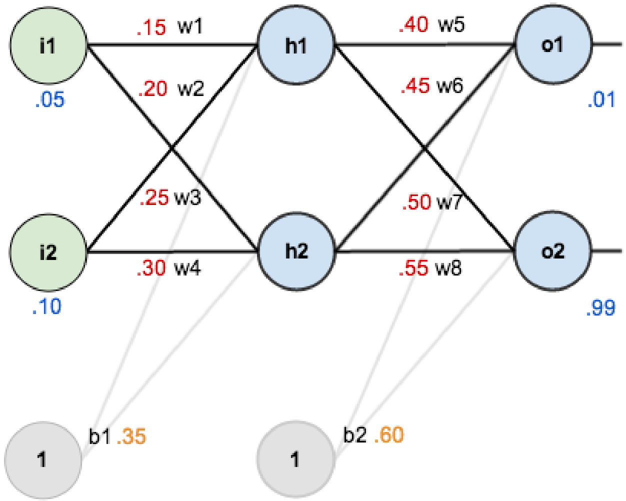
$$net_{h1}=w1*i1+w2*i2+b1*1$$

$$net_{h1}=0{,}15*0{,}05+0{,}2*0{,}1+0{,}35*1=0{,}3775$$

Calculamos la salida para **h1**, usando la activación sigmoide

$$out_{h1}=\frac{1}{1+e^{-0{,}3775}}=0{,}59327$$

Perceptrón multicapa (fully connected)

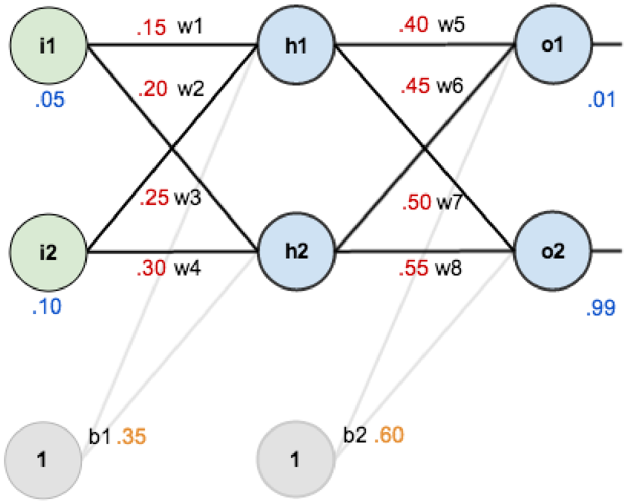


De igual forma se calcula la salida de la neurona h2

$$out_{h2} = 0,59688$$

Entrada		Salida - esperada		Salida - real	
i1	**i2**	**o1**	**o2**	**h1**	**h2**
0.05	0.10	0.01	0.99	0.59327	0.59688

Perceptrón multicapa (fully connected)



Repetimos los mismos pasos para calcular la salida de las neuronas de salida, usando como entrada las salidas de las neuronas **h**.

Calculo la entrada para **o1**:

$$net_{o1} = w5 * out_{h1} + w6 * out_{h2} + b2 * 1$$

$$net_{o1} = 0{,}4 * 0{,}59327 + 0{,}45 * 0{,}59688 + 0{,}6 * 1 = 1{,}1059$$

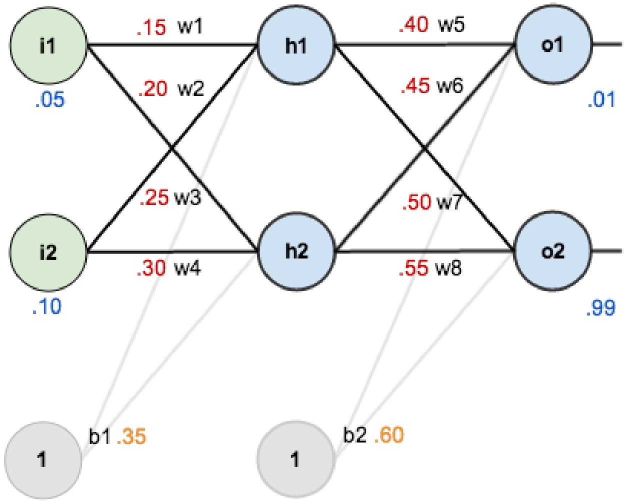
Calculo la salida para **o1**:

$$out_{o1} = \frac{1}{1 + e^{-1{,}1059}} = 0{,}75137$$

$$out_{o2} = 0{,}7729$$

Entrada		Salida - esperada		Salida - real	
i1	**i2**	**o1**	**o2**	**h1**	**h2**
0.05	0.10	0.01	0.99	0.59327	0.59688

Perceptrón multicapa (fully connected)

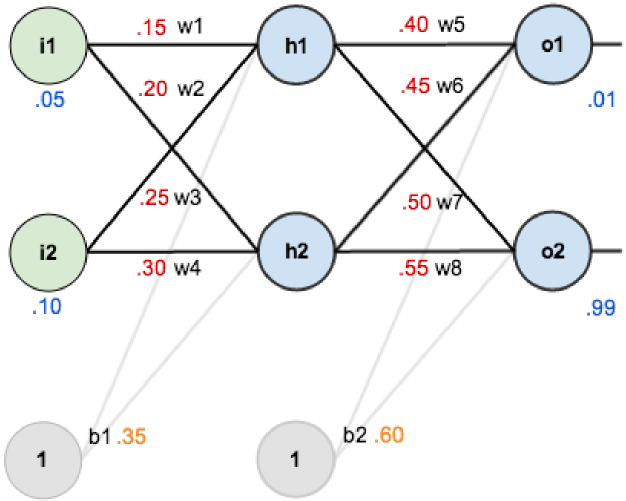


Calculamos la salida para **o2** de igual forma que para o1:

$$out_{o2}=0,7729$$

Entrada		Salida - esperada		Salida - real			
i1	**i2**	**o1**	**o2**	**h1**	**h2**	**o1**	**o2**
0.05	0.10	0.01	0.99	0.59327	0.59688	0.75137	0.7729

Perceptrón multicapa (fully connected)



Calculamos el error cuadrático medio de cada neurona de salida, con la siguiente fórmula:

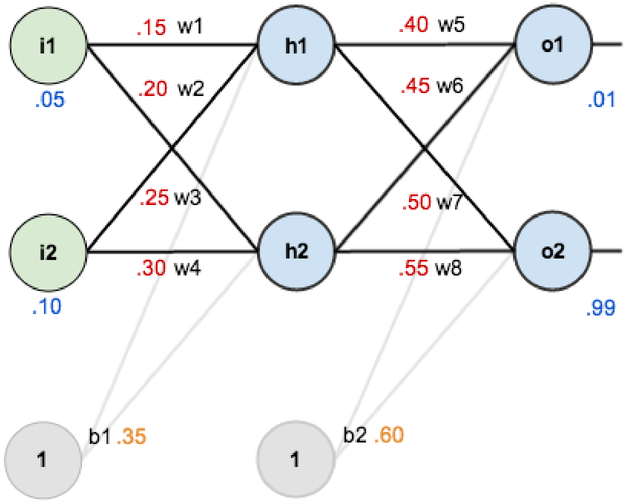
$$E = \sum \frac{1}{2}(target - output)^2$$

$$E_{01} = \sum \frac{1}{2}(target_{o1} - output_{o1})^2 = \frac{1}{2}(0{,}01 - 0{,}75137)^2 = 0{,}274811$$

$$E_{02} = 0{,}02356$$

Entrada		Salida - esperada		Salida - real			
i1	i2	o1	o2	h1	h2	o1	o2
0.05	0.10	0.01	0.99	0.59327	0.59688	0.75137	0.7729

Perceptrón multicapa (fully connected)



Calculamos el error total, como la suma de todos los errores:

$$E_{total} = E_{o1} + E_{o2} = 0{,}274811 + 0{,}02356 = 0{,}29837$$

Entrada		Salida - esperada		Salida - real				Errores	
i1	**i2**	**o1**	**o2**	**h1**	**h2**	**o1**	**o2**	**o1**	**o2**
0.05	0.10	0.01	0.99	0.59327	0.59688	0.75137	0.7729	0.274811	0.02356

Aplicamos Backpropagation

$$\vec{\nabla} E_{total} = \left(\frac{\partial E_{total}}{\partial w1}, \frac{\partial E_{total}}{\partial w2}, \frac{\partial E_{total}}{\partial w3}, \frac{\partial E_{total}}{\partial w4}, \frac{\partial E_{total}}{\partial w5}, \frac{\partial E_{total}}{\partial w6}, \frac{\partial E_{total}}{\partial w7}, \frac{\partial E_{total}}{\partial w8} \right)$$

¿Hacia dónde decrece el error?
Hacia allá vamos

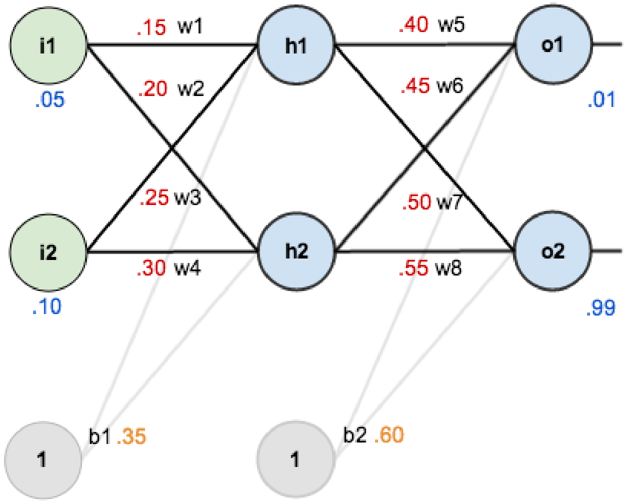
Aplicamos Backpropagation

$$\vec{\nabla} E_{total} = \left(\frac{\partial E_{total}}{\partial w1}, \frac{\partial E_{total}}{\partial w2}, \frac{\partial E_{total}}{\partial w3}, \frac{\partial E_{total}}{\partial w4}, \frac{\partial E_{total}}{\partial w5}, \frac{\partial E_{total}}{\partial w6}, \frac{\partial E_{total}}{\partial w7}, \frac{\partial E_{total}}{\partial w8} \right)$$

Empecemos por calcular uno de estos valores.
Es el primero de los pesos, en la última capa de la red.

$$\frac{\partial E_{total}}{\partial w5} : Se\ lee\ como\ la\ derivada\ parcial\ de\ E_{total}\ respecto\ de\ w5,\ o\ tambien\ como\ el\ gradiente\ respecto\ de\ w5.$$

Backpropagation

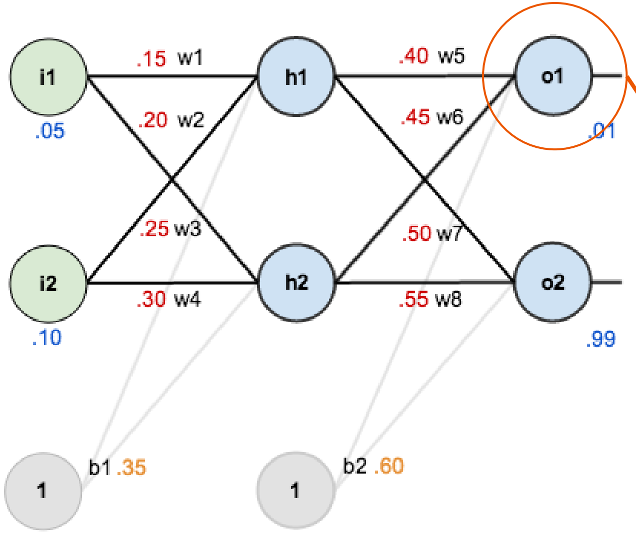


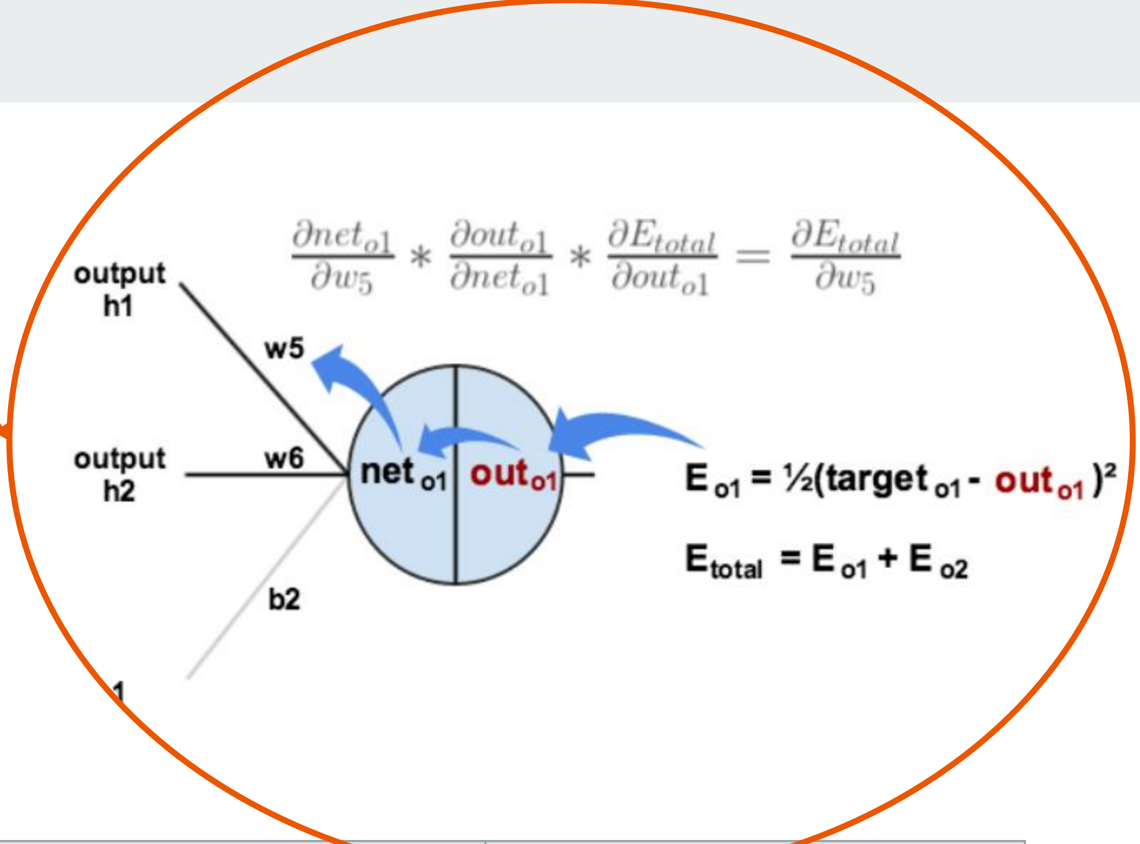
Utilizando la regla de la cadena:

$$\frac{\partial E_{total}}{\partial w5} = \frac{\partial E_{total}}{\partial out_{01}} * \frac{\partial out_{01}}{\partial net_{01}} * \frac{\partial net_{01}}{\partial w5}$$

Entrada		Salida - esperada		Salida - real				Errores		
i1	i2	o1	o2	h1	h2	o1	o2	o1	o2	total
0.05	0.10	0.01	0.99	0.59327	0.59688	0.75137	0.7729	0.274811	0.02356	0.29837

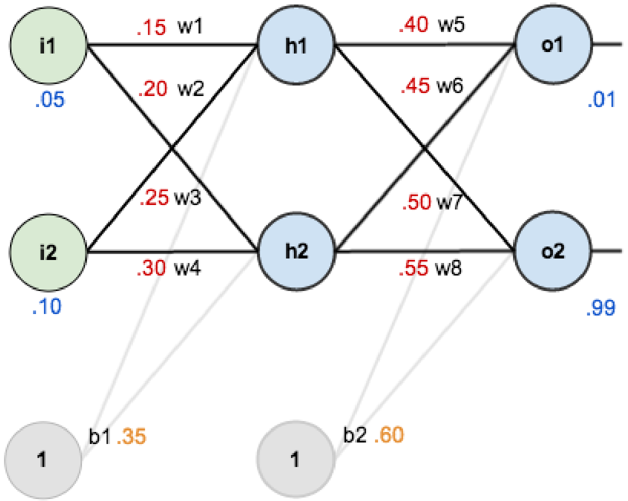
Backpropagation





Entrada		Salida - esperada		Salida - real				Errores		
i1	i2	o1	o2	h1	h2	o1	o2	o1	o2	total
0.05	0.10	0.01	0.99	0.59327	0.59688	0.75137	0.7729	0.274811	0.02356	0.29837

Backpropagation



Utilizando la regla de la cadena:

$$\frac{\partial E_{total}}{\partial w5} = \frac{\partial E_{total}}{\partial out_{01}} * \frac{\partial out_{01}}{\partial net_{01}} * \frac{\partial net_{01}}{\partial w5}$$

Es constante

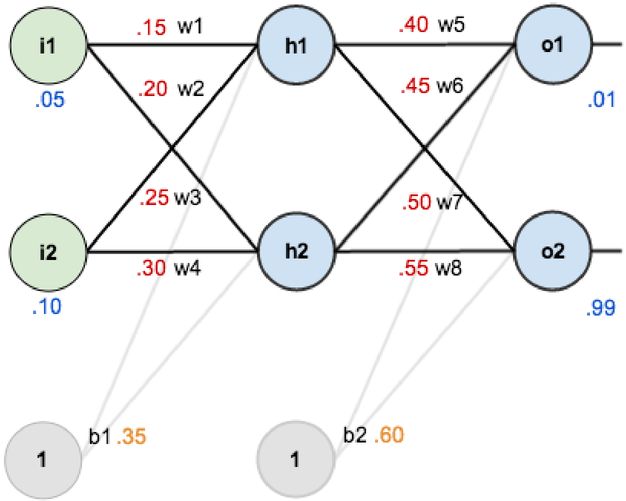
$$E_{total} = \frac{1}{2}(target_{o1} - out_{o1})^2 + \frac{1}{2}(target_{o2} - out_{o2})^2$$

$$\frac{\partial E_{total}}{\partial out_{o1}} = 2 * \frac{1}{2}(target_{o1} - out_{o1})^{2-1} * -1 + 0$$

$$\frac{\partial E_{total}}{\partial out_{o1}} = -(target_{o1} - out_{o1}) = -(0.01 - 0.75136507) = 0.74136507$$

Entrada		Salida - esperada		Salida - real				Errores		
i1	**i2**	**o1**	**o2**	**h1**	**h2**	**o1**	**o2**	**o1**	**o2**	**total**
0.05	0.10	0.01	0.99	0.59327	0.59688	0.75137	0.7729	0.274811	0.02356	0.29837

Backpropagation



Utilizando la regla de la cadena:

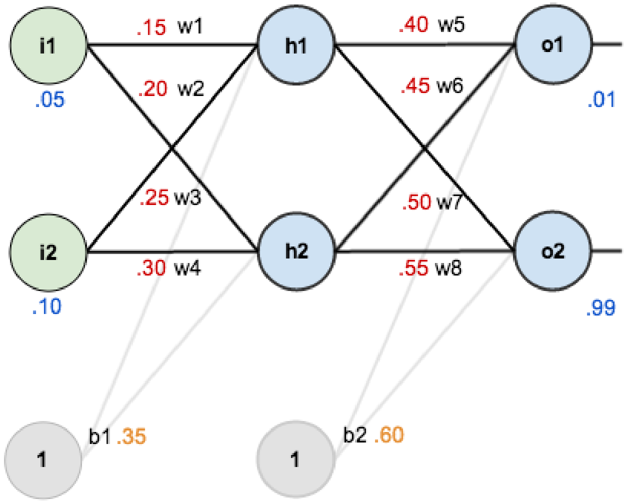
$$\frac{\partial E_{total}}{\partial w5} = \frac{\partial E_{total}}{\partial out_{01}} * \frac{\partial out_{01}}{\partial net_{01}} * \frac{\partial net_{01}}{\partial w5}$$

$$out_{o1} = \frac{1}{1+e^{-net_{o1}}}$$

$$\frac{\partial out_{o1}}{\partial net_{o1}} = out_{o1}(1 - out_{o1}) = 0.75136507(1 - 0.75136507) = 0.186815602$$

Entrada		Salida - esperada		Salida - real				Errores		
i1	**i2**	**o1**	**o2**	**h1**	**h2**	**o1**	**o2**	**o1**	**o2**	**total**
0.05	0.10	0.01	0.99	0.59327	0.59688	0.75137	0.7729	0.274811	0.02356	0.29837

Backpropagation



Utilizando la regla de la cadena:

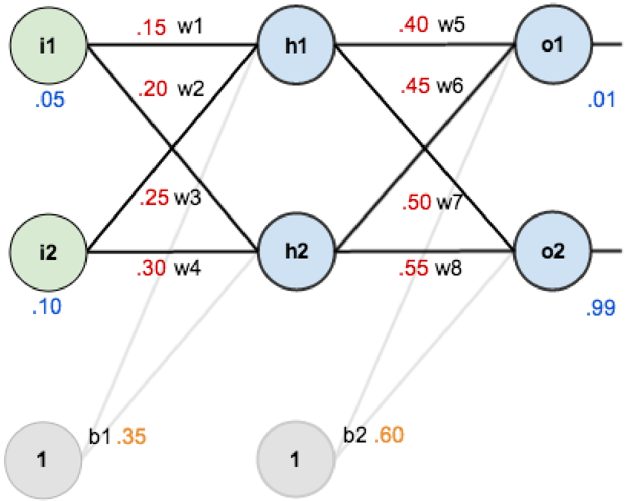
$$\frac{\partial E_{total}}{\partial w5} = \frac{\partial E_{total}}{\partial out_{01}} * \frac{\partial out_{01}}{\partial net_{01}} * \frac{\partial net_{01}}{\partial w5}$$

$$net_{o1} = w_5 * out_{h1} + w_6 * out_{h2} + b_2 * 1$$

$$\frac{\partial net_{o1}}{\partial w_5} = 1 * out_{h1} * w_5^{(1-1)} + 0 + 0 = out_{h1} = 0.593269992$$

Entrada		Salida - esperada		Salida - real				Errores		
i1	i2	o1	o2	h1	h2	o1	o2	o1	o2	total
0.05	0.10	0.01	0.99	0.59327	0.59688	0.75137	0.7729	0.274811	0.02356	0.29837

Backpropagation



Utilizando la regla de la cadena:

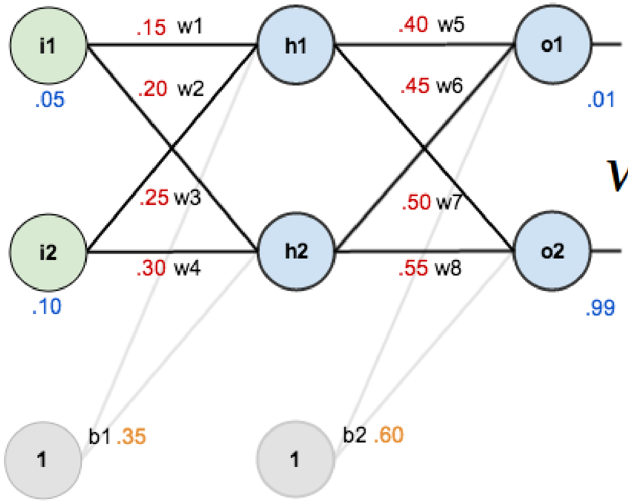
$$\frac{\partial E_{total}}{\partial w5} = \frac{\partial E_{total}}{\partial out_{01}} * \frac{\partial out_{01}}{\partial net_{01}} * \frac{\partial net_{01}}{\partial w5}$$

Todo junto:

$$\frac{\partial E_{total}}{\partial w_5} = 0.74136507 * 0.186815602 * 0.593269992 = 0.082167041$$

Entrada		Salida - esperada		Salida - real				Errores		
i1	i2	o1	o2	h1	h2	o1	o2	o1	o2	total
0.05	0.10	0.01	0.99	0.59327	0.59688	0.75137	0.7729	0.274811	0.02356	0.29837

Backpropagation



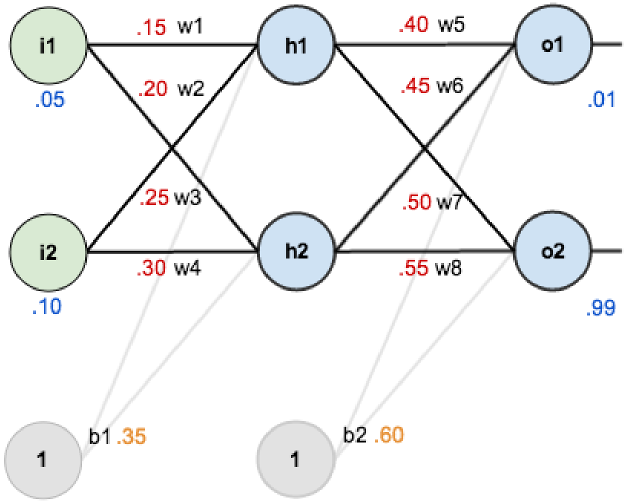
Calculando el nuevo peso w5 o w5’

$$w5' = w5 - \alpha * 0{,}082167041 = 0{,}35891648$$

Tasa de aprendizaje
(learning rate)
= 0.5 en este ejemplo

Entrada		Salida - esperada		Salida - real				Errores		
i1	**i2**	**o1**	**o2**	**h1**	**h2**	**o1**	**o2**	**o1**	**o2**	**total**
0.05	0.10	0.01	0.99	0.59327	0.59688	0.75137	0.7729	0.274811	0.02356	0.29837

Backpropagation

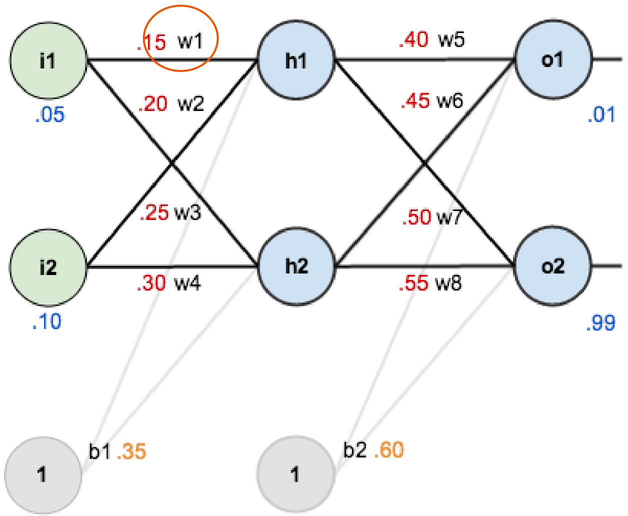


Repetimos para los otros pesos:

$$w_6^+ = 0.408666186$$
$$w_7^+ = 0.511301270$$
$$w_8^+ = 0.561370121$$

Entrada		Salida - esperada		Salida - real				Errores		
i1	i2	o1	o2	h1	h2	o1	o2	o1	o2	total
0.05	0.10	0.01	0.99	0.59327	0.59688	0.75137	0.7729	0.274811	0.02356	0.29837

Backpropagation - capa anterior, oculta



$$\frac{\partial E_{total}}{\partial w1} = \frac{\partial E_{total}}{\partial out_{h1}} * \frac{\partial out_{h1}}{\partial net_{h1}} * \frac{\partial net_{h1}}{\partial w1}$$

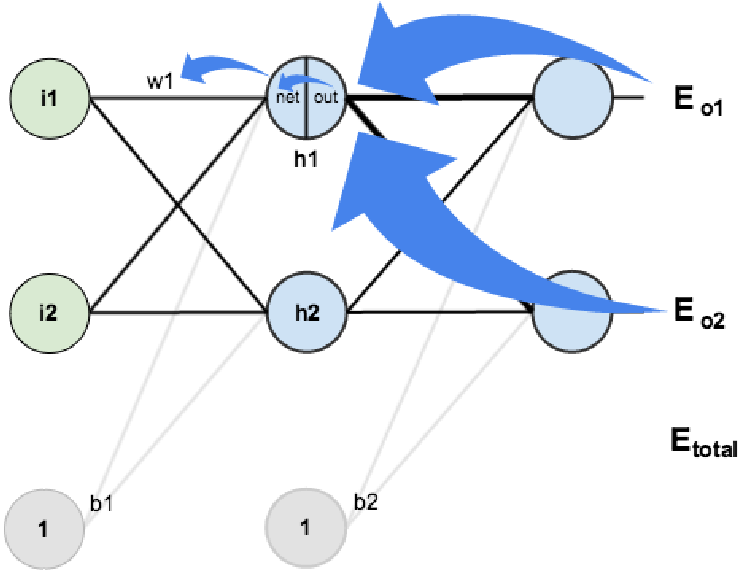
$$\frac{\partial E_{total}}{\partial out_{h1}} = \frac{\partial E_{o1}}{\partial out_{h1}} + \frac{\partial E_{o2}}{\partial out_{h1}}$$

Entrada		Salida - esperada		Salida - real				Errores		
i1	i2	o1	o2	h1	h2	o1	o2	o1	o2	total
0.05	0.10	0.01	0.99	0.59327	0.59688	0.75137	0.7729	0.274811	0.02356	0.29837

Backpropagation - capa anterior, oculta

$$\frac{\partial E_{total}}{\partial w_1} = \frac{\partial E_{total}}{\partial out_{h1}} * \frac{\partial out_{h1}}{\partial net_{h1}} * \frac{\partial net_{h1}}{\partial w_1}$$

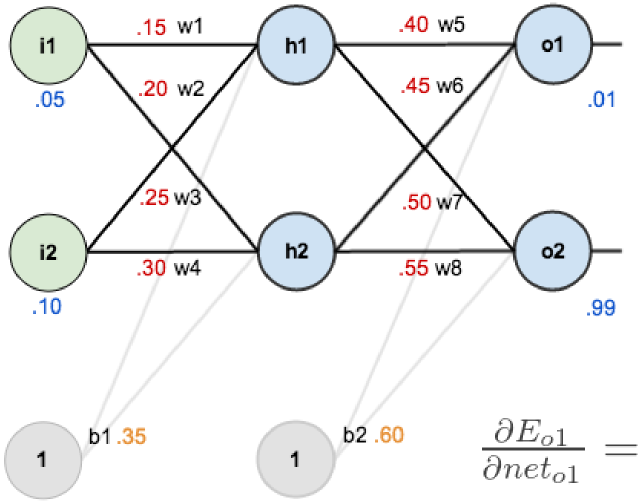
$$\frac{\partial E_{total}}{\partial out_{h1}} = \frac{\partial E_{o1}}{\partial out_{h1}} + \frac{\partial E_{o2}}{\partial out_{h1}}$$



$E_{total} = E_{o1} + E_{o2}$

Backpropagation

$$\frac{\partial E_{total}}{\partial out_{h1}} = \frac{\partial E_{o1}}{\partial out_{h1}} + \frac{\partial E_{o2}}{\partial out_{h1}}$$



Empezaremos por el primer término:

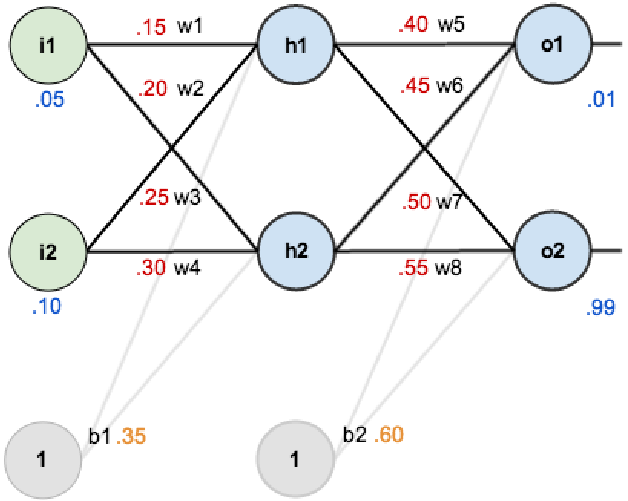
$$\frac{\partial E_{o1}}{\partial out_{h1}} = \frac{\partial E_{o1}}{\partial net_{o1}} * \frac{\partial net_{o1}}{\partial out_{h1}}$$

$\frac{\partial E_{o1}}{\partial net_{o1}}$ Ya fue calculado anteriormente:

$$\frac{\partial E_{o1}}{\partial net_{o1}} = \frac{\partial E_{o1}}{\partial out_{o1}} * \frac{\partial out_{o1}}{\partial net_{o1}} = 0.74136507 * 0.186815602 = 0.138498562$$

Entrada		Salida - esperada		Salida - real				Errores		
i1	i2	o1	o2	h1	h2	o1	o2	o1	o2	total
0.05	0.10	0.01	0.99	0.59327	0.59688	0.75137	0.7729	0.274811	0.02356	0.29837

Backpropagation



$$\frac{\partial E_{total}}{\partial out_{h1}} = \frac{\partial E_{o1}}{\partial out_{h1}} + \frac{\partial E_{o2}}{\partial out_{h1}}$$

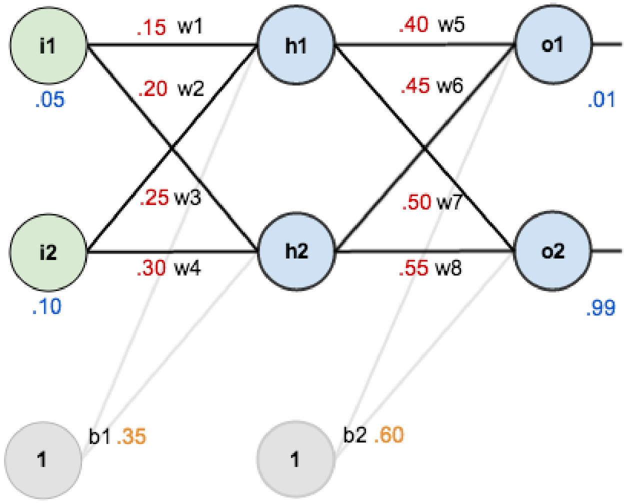
$$\frac{\partial E_{o1}}{\partial out_{h1}} = \frac{\partial E_{o1}}{\partial net_{o1}} * \frac{\partial net_{o1}}{\partial out_{h1}}$$

$$net_{o1} = w_5 * out_{h1} + w_6 * out_{h2} + b_2 * 1$$

$$\frac{\partial net_{o1}}{\partial out_{h1}} = w_5 = 0.40$$

Entrada		Salida - esperada		Salida - real				Errores		
i1	i2	o1	o2	h1	h2	o1	o2	o1	o2	total
0.05	0.10	0.01	0.99	0.59327	0.59688	0.75137	0.7729	0.274811	0.02356	0.29837

Backpropagation



$$\frac{\partial E_{total}}{\partial out_{h1}} = \frac{\partial E_{o1}}{\partial out_{h1}} + \frac{\partial E_{o2}}{\partial out_{h1}}$$

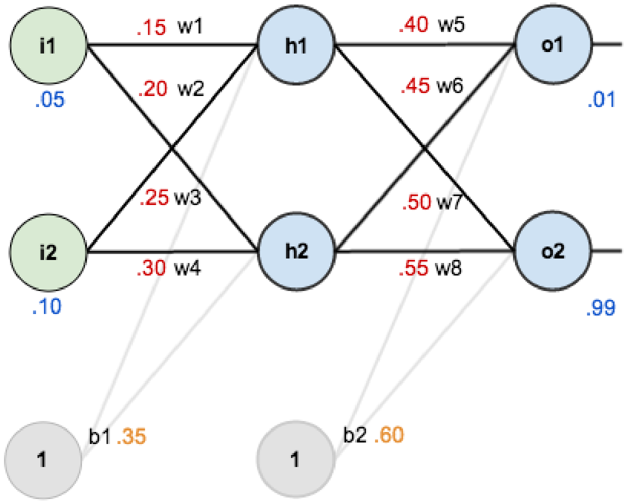
$$\frac{\partial E_{o1}}{\partial out_{h1}} = \frac{\partial E_{o1}}{\partial net_{o1}} * \frac{\partial net_{o1}}{\partial out_{h1}}$$

$$\frac{\partial E_{o1}}{\partial out_{h1}} = \frac{\partial E_{o1}}{\partial net_{o1}} * \frac{\partial net_{o1}}{\partial out_{h1}} = 0.138498562 * 0.40$$

$$= 0.055399425$$

Entrada		Salida - esperada		Salida - real				Errores		
i1	i2	o1	o2	h1	h2	o1	o2	o1	o2	total
0.05	0.10	0.01	0.99	0.59327	0.59688	0.75137	0.7729	0.274811	0.02356	0.29837

Backpropagation



$$\frac{\partial E_{total}}{\partial out_{h1}} = \frac{\partial E_{o1}}{\partial out_{h1}} + \frac{\partial E_{o2}}{\partial out_{h1}}$$

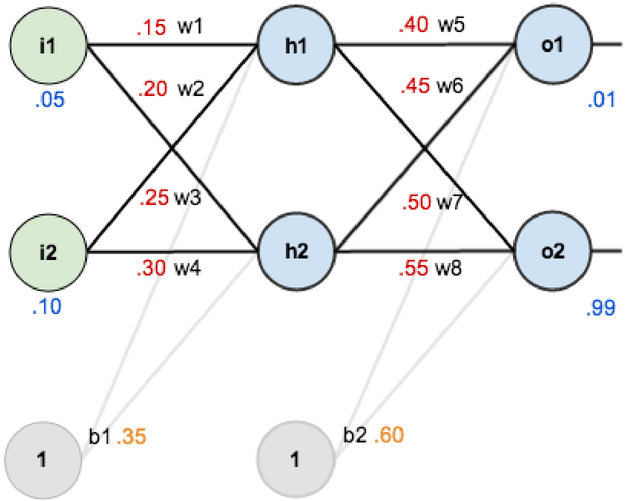
$$\frac{\partial E_{o2}}{\partial out_{h1}} = -0.019049119$$

$$\frac{\partial E_{total}}{\partial out_{h1}} = \frac{\partial E_{o1}}{\partial out_{h1}} + \frac{\partial E_{o2}}{\partial out_{h1}} = 0.055399425 + -0.019049119$$

$$= 0.036350306$$

Entrada		Salida - esperada		Salida - real				Errores		
i1	i2	o1	o2	h1	h2	o1	o2	o1	o2	total
0.05	0.10	0.01	0.99	0.59327	0.59688	0.75137	0.7729	0.274811	0.02356	0.29837

Backpropagation - capa anterior, oculta

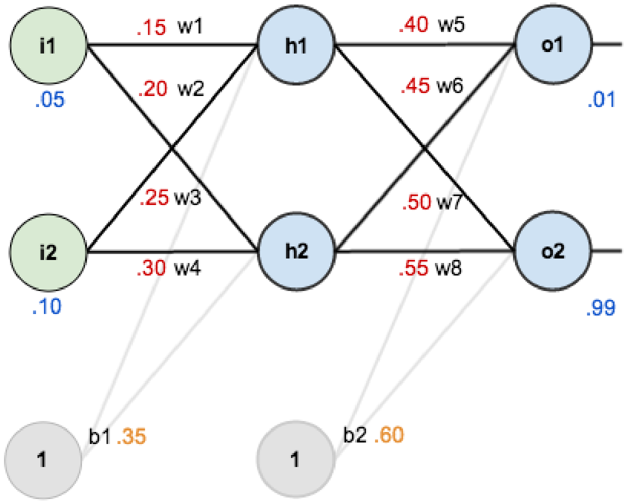


$$\frac{\partial E_{total}}{\partial w1} = \frac{\partial E_{total}}{\partial out_{h1}} * \boxed{\frac{\partial out_{h1}}{\partial net_{h1}} * \frac{\partial net_{h1}}{\partial w1}}$$

Aún hay que calcular estos dos términos

Entrada		Salida - esperada		Salida - real				Errores		
i1	i2	o1	o2	h1	h2	o1	o2	o1	o2	total
0.05	0.10	0.01	0.99	0.59327	0.59688	0.75137	0.7729	0.274811	0.02356	0.29837

Backpropagation - capa anterior, oculta



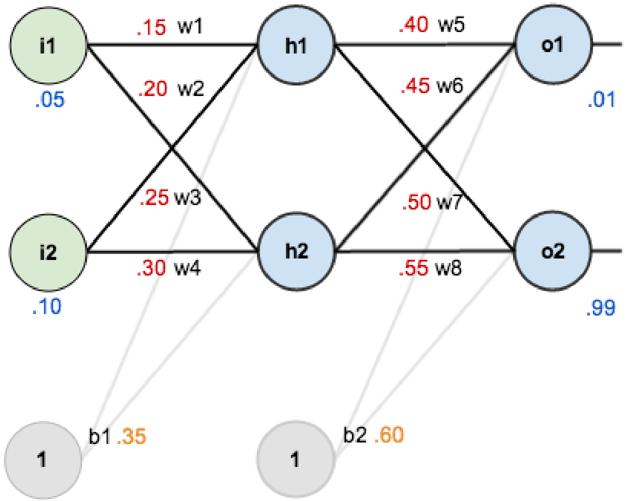
$$out_{h1} = \frac{1}{1+e^{-net_{h1}}}$$

$$\frac{\partial out_{h1}}{\partial net_{h1}} = out_{h1}(1 - out_{h1}) = 0.59326999(1 - 0.59326999)$$

$$= 0.241300709$$

Entrada		Salida - esperada		Salida - real				Errores		
i1	i2	o1	o2	h1	h2	o1	o2	o1	o2	total
0.05	0.10	0.01	0.99	0.59327	0.59688	0.75137	0.7729	0.274811	0.02356	0.29837

Backpropagation - capa anterior, oculta

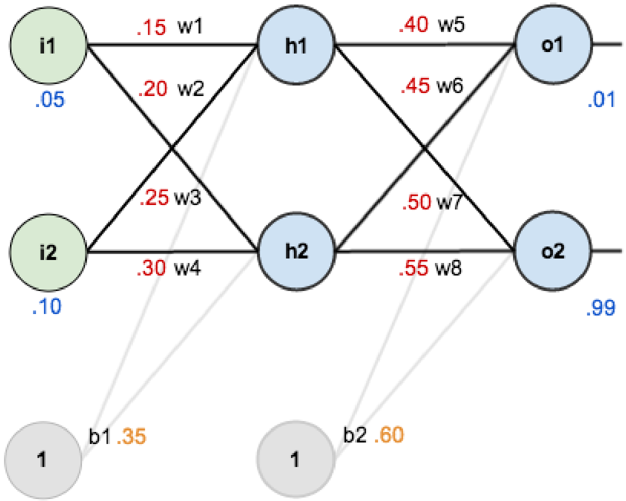


$$net_{h1} = w_1 * i_1 + w_3 * i_2 + b_1 * 1$$

$$\frac{\partial net_{h1}}{\partial w_1} = i_1 = 0.05$$

Entrada		Salida - esperada		Salida - real				Errores		
i1	i2	o1	o2	h1	h2	o1	o2	o1	o2	total
0.05	0.10	0.01	0.99	0.59327	0.59688	0.75137	0.7729	0.274811	0.02356	0.29837

Backpropagation - capa anterior, oculta

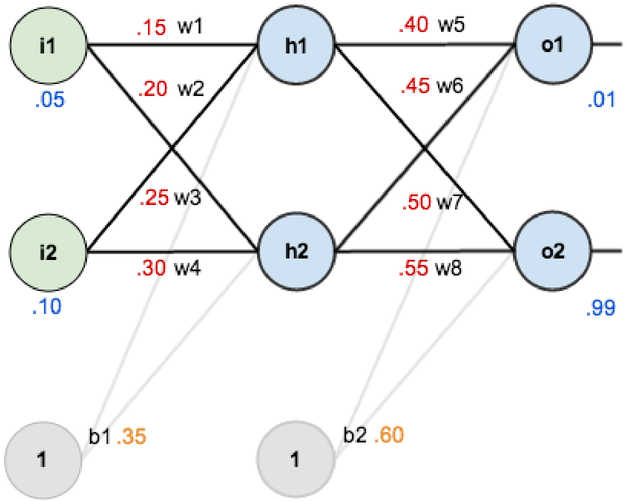


$$\frac{\partial E_{total}}{\partial w1} = \frac{\partial E_{total}}{\partial out_{h1}} * \frac{\partial out_{h1}}{\partial net_{h1}} * \frac{\partial net_{h1}}{\partial w1}$$

$$0.036350306 * 0.241300709 * 0.05 = 0.000438568$$

Entrada		Salida - esperada		Salida - real				Errores		
i1	i2	o1	o2	h1	h2	o1	o2	o1	o2	total
0.05	0.10	0.01	0.99	0.59327	0.59688	0.75137	0.7729	0.274811	0.02356	0.29837

Backpropagation - capa anterior, oculta

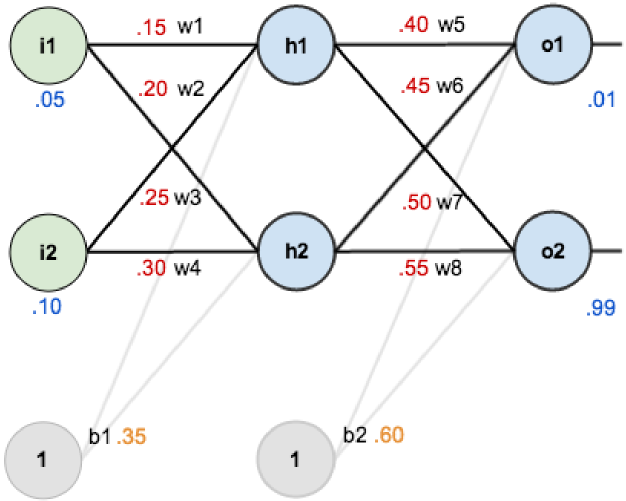


$$w1' = w1 - \alpha * \frac{\partial E_{total}}{\partial w1}$$

$$= 0.15 - 0.5 * 0.000438568 = 0.149780716$$

Entrada		Salida - esperada		Salida - real				Errores		
i1	i2	o1	o2	h1	h2	o1	o2	o1	o2	total
0.05	0.10	0.01	0.99	0.59327	0.59688	0.75137	0.7729	0.274811	0.02356	0.29837

Backpropagation - capa anterior, oculta



Repetimos para w2, w3 y w4

$$w_2^+ = 0.19956143$$

$$w_3^+ = 0.24975114$$

$$w_4^+ = 0.29950229$$

Entrada		Salida - esperada		Salida - real				Errores		
i1	i2	o1	o2	h1	h2	o1	o2	o1	o2	total
0.05	0.10	0.01	0.99	0.59327	0.59688	0.75137	0.7729	0.274811	0.02356	0.29837