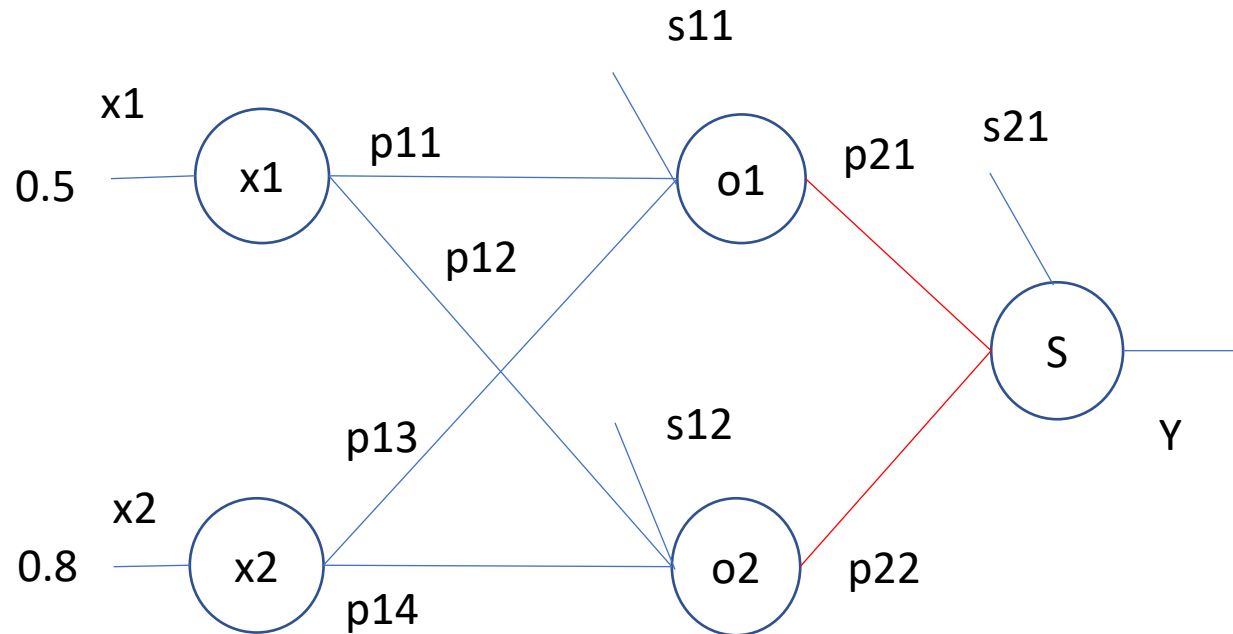


Retro propagación en una red de 3 capas

jafh

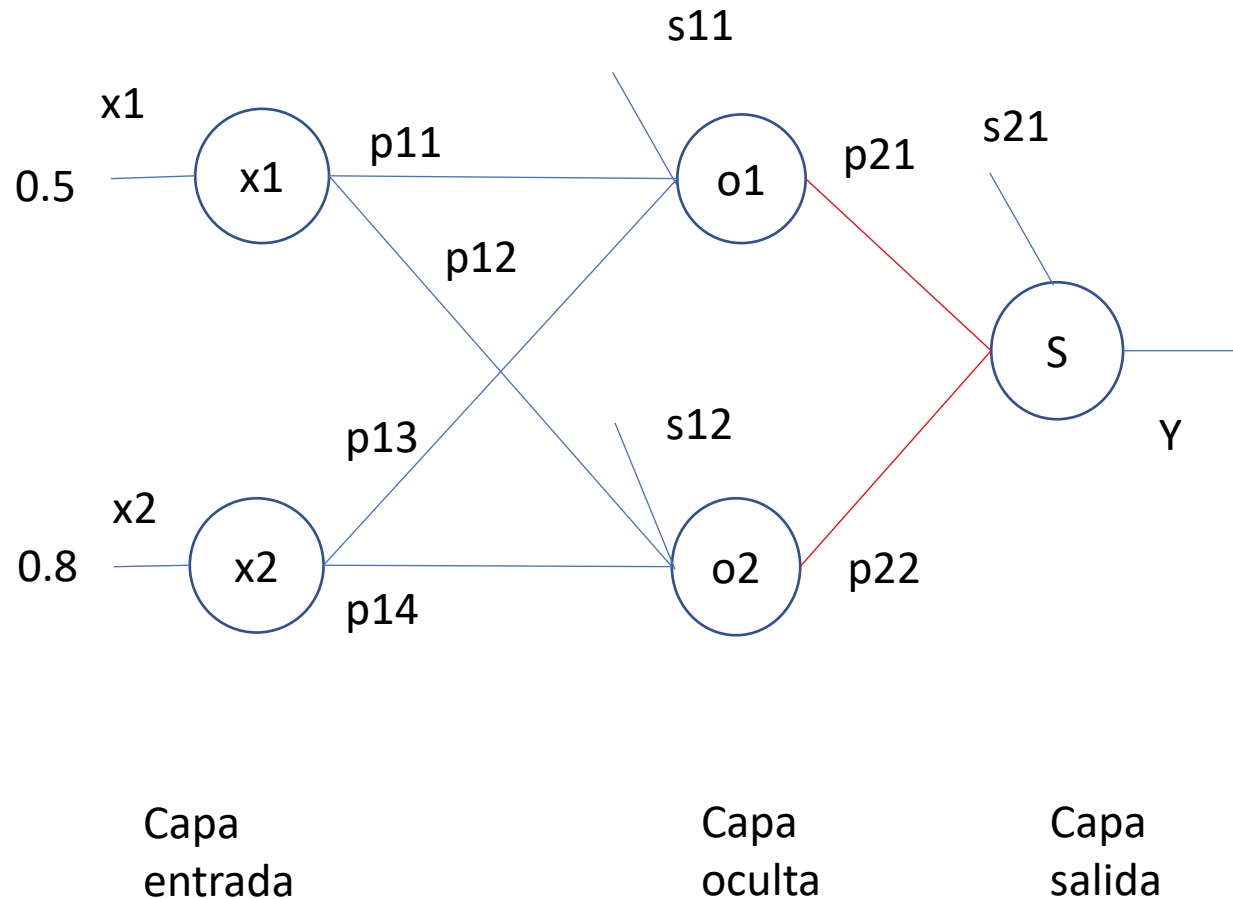
Red de Neuronas sigmoides de 2*2*1



Usaremos neuronas con función de activación sigmoide:

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

Capa de entrada



La capa de entrada solo sirve para recibir los datos y enviarlos a la capa oculta.

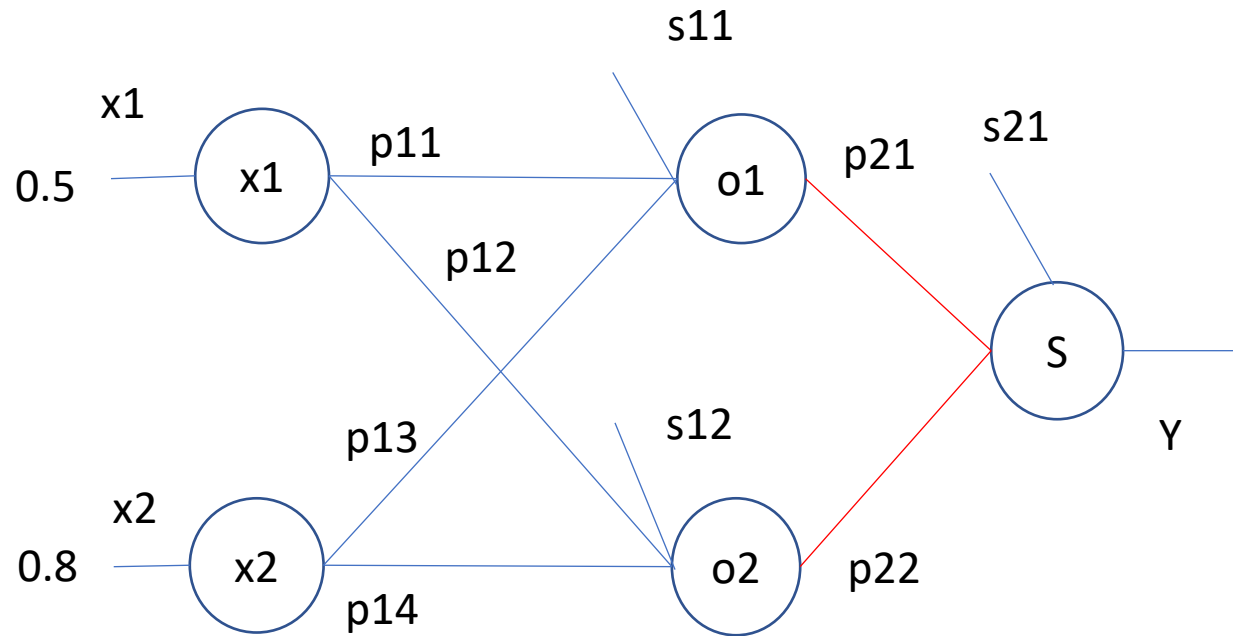
Pesos capa oculta:
 $p11, p12, p13, p14$

Pesos capa de salida:
 $p21, p22$

Forward pass

Propagación hacia adelante

Entradas co



Entradas de la capa de entrada

$$\text{Suma } o1 = p11 * x1 + p13 * x2 + s11$$

$$\text{Suma } o2 = p12 * x1 + p14 * x2 + s12$$

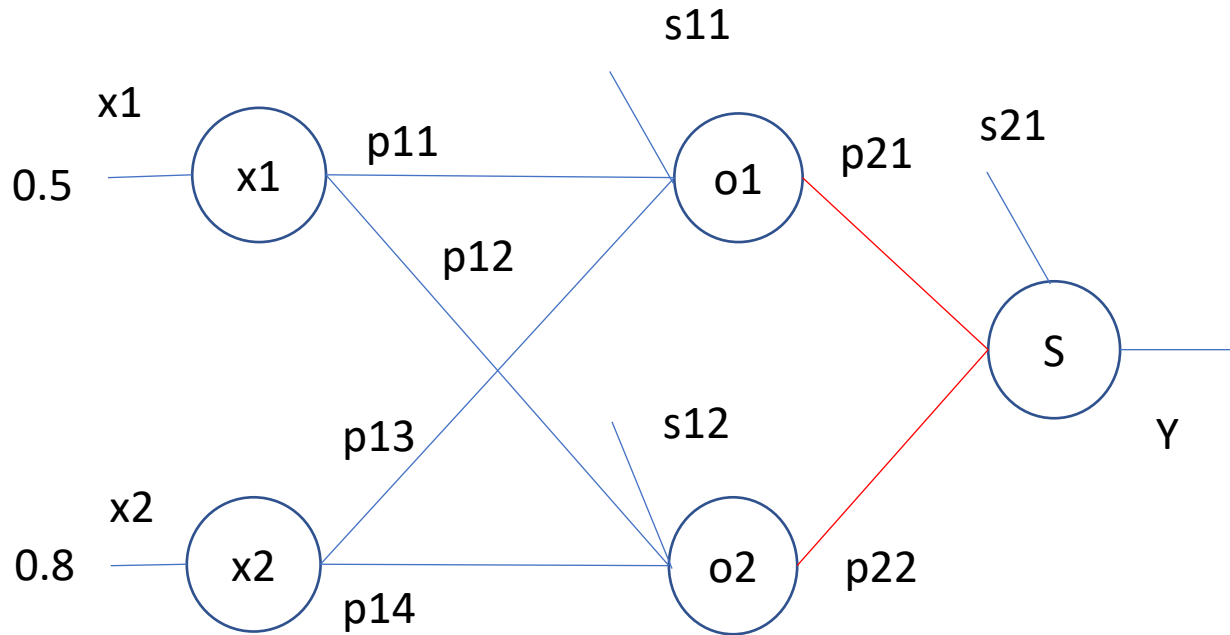
$$\begin{bmatrix} p11 & p13 \\ p12 & p14 \end{bmatrix} * \begin{bmatrix} x1 \\ x2 \end{bmatrix} = \begin{bmatrix} \text{suma } o1 \\ \text{suma } o2 \end{bmatrix}$$

Salidas de la capa oculta $o1, o2$

$$\sigma(\text{suma } o1) = \frac{1}{1 + e^{-\text{suma } o1}}$$

$$\sigma(\text{suma } o2) = \frac{1}{1 + e^{-\text{suma } o2}}$$

Salidas cs



Cálculo de la salida capa de salida:

Entradas de la capa de salida:

- $h1 = \sigma(\text{suma } o1)$

- $h2 = \sigma(\text{suma } o2)$

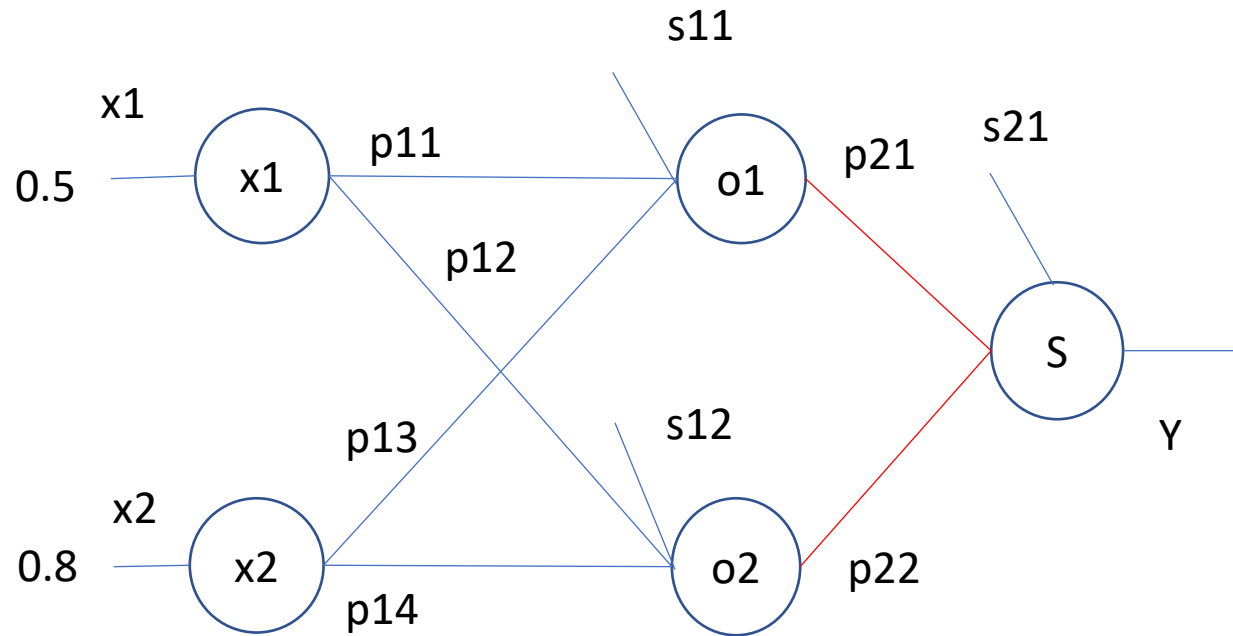
Entrada ponderada de la capa de salida

- $zy = p21h1 + p22h2 + s21$

Cálculo de salida de la capa de salida:

- $y = \sigma(zy) = \frac{1}{1 + e^{-zy}}$

Error



Cálculo del error:

Para cálculo del error usaremos el error cuadrático medio. Con y salida predicha y La salida esperada t :

El error cuadrático medio es:

$$E = 1/2 (t - y)^2$$

En nuestro caso:

$$E = 1/2 (t - \sigma(zy))^2$$

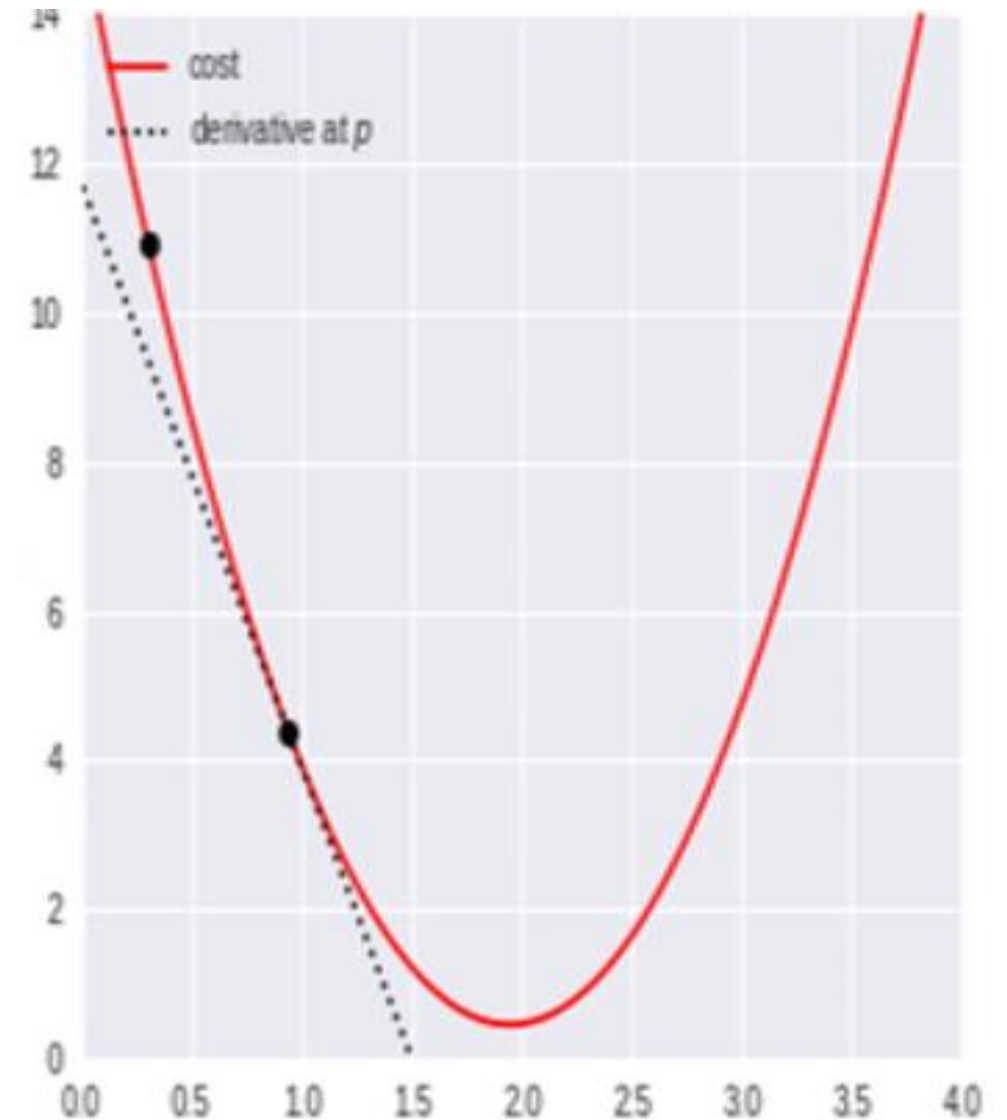
Minimizar el error

Usaremos descenso de gradiente para reducir el error, por lo que necesitaremos calcular el gradiente de la función de calculo de error = $\frac{1}{2}(t-y)^2$

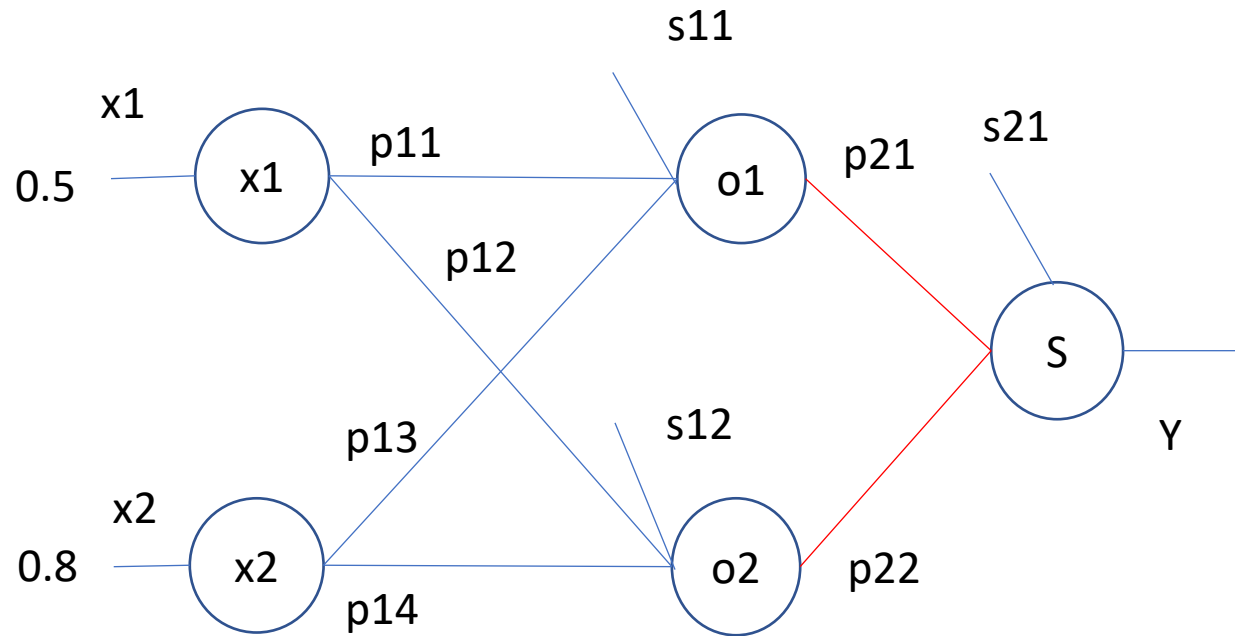
La derivada de la función:

$$(\frac{1}{2}(t-y)^2)' = -(t-y) = y-t$$

La derivada me dice hacia donde crece la función.

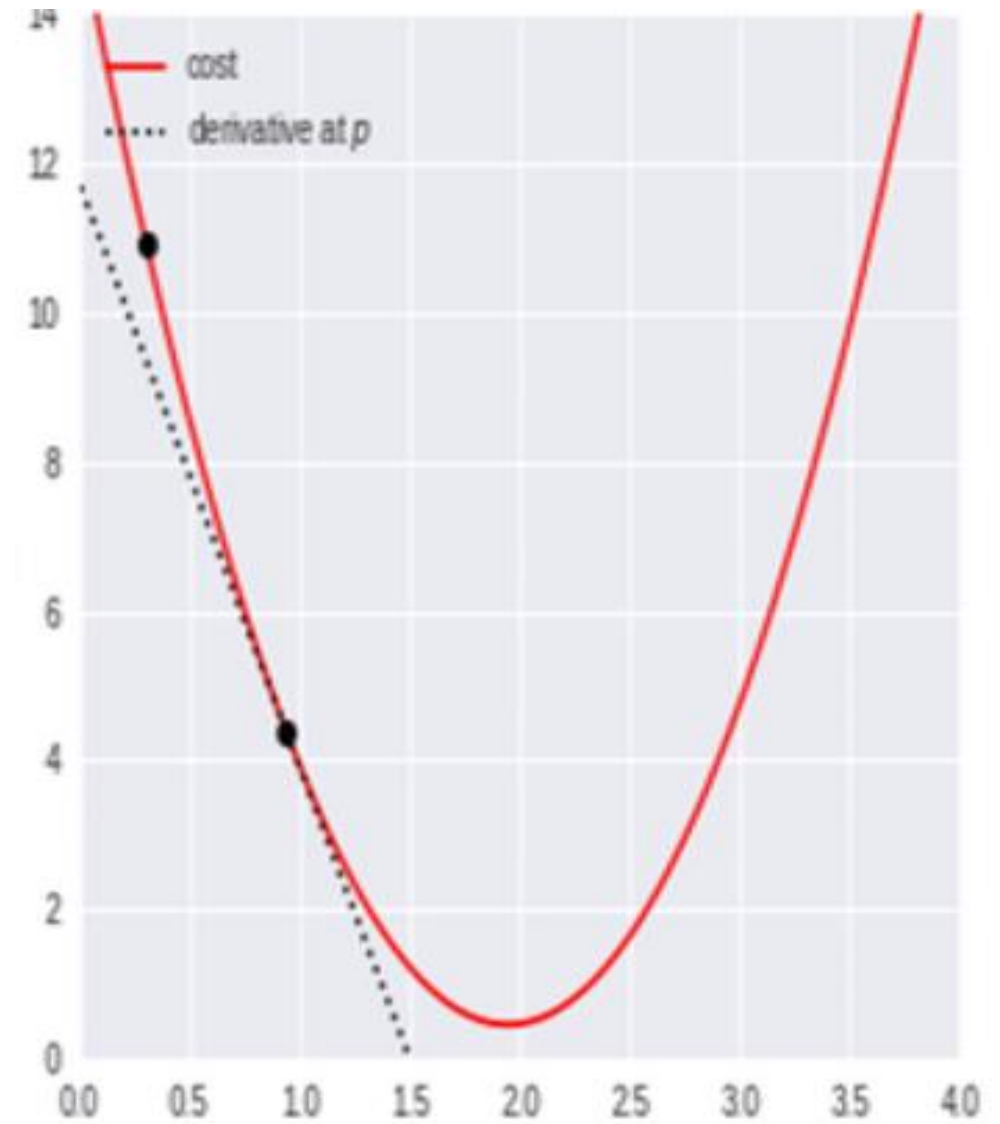


Retro propagación

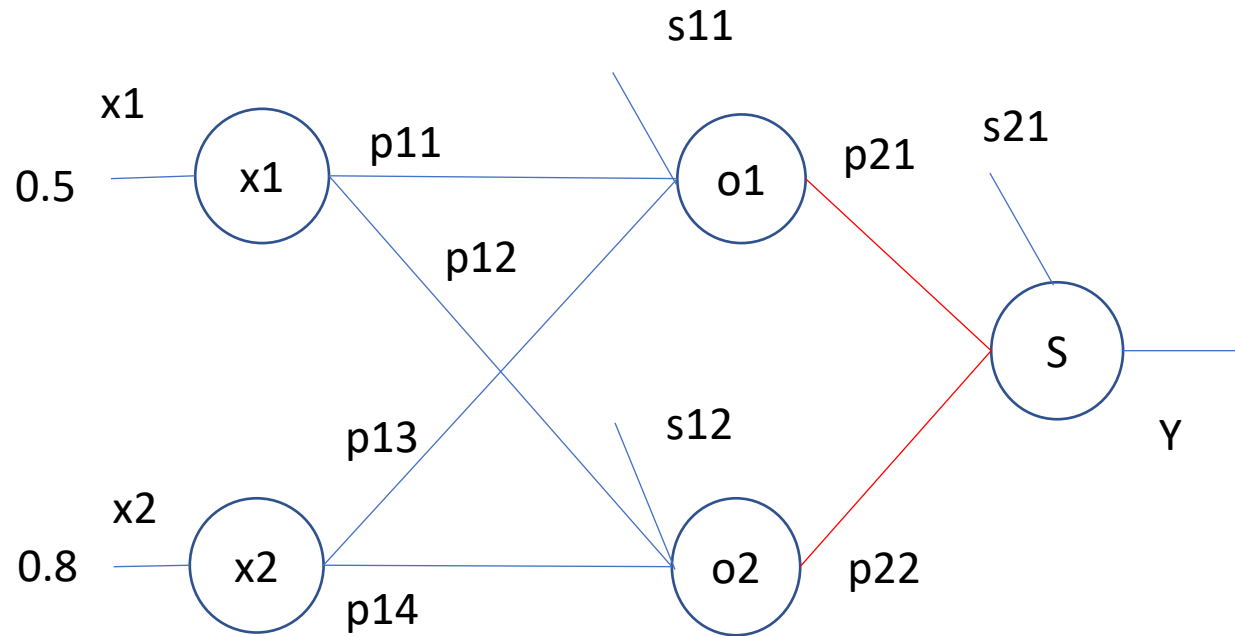


La actualización de pesos se hace del final hacia el inicio por eso se llama retro propagación, usaremos descenso de gradiente para reducir el error, por lo que necesitaremos calcular el gradiente de la función de calculo de error $\frac{1}{2}(t-y)^2$.

jafh



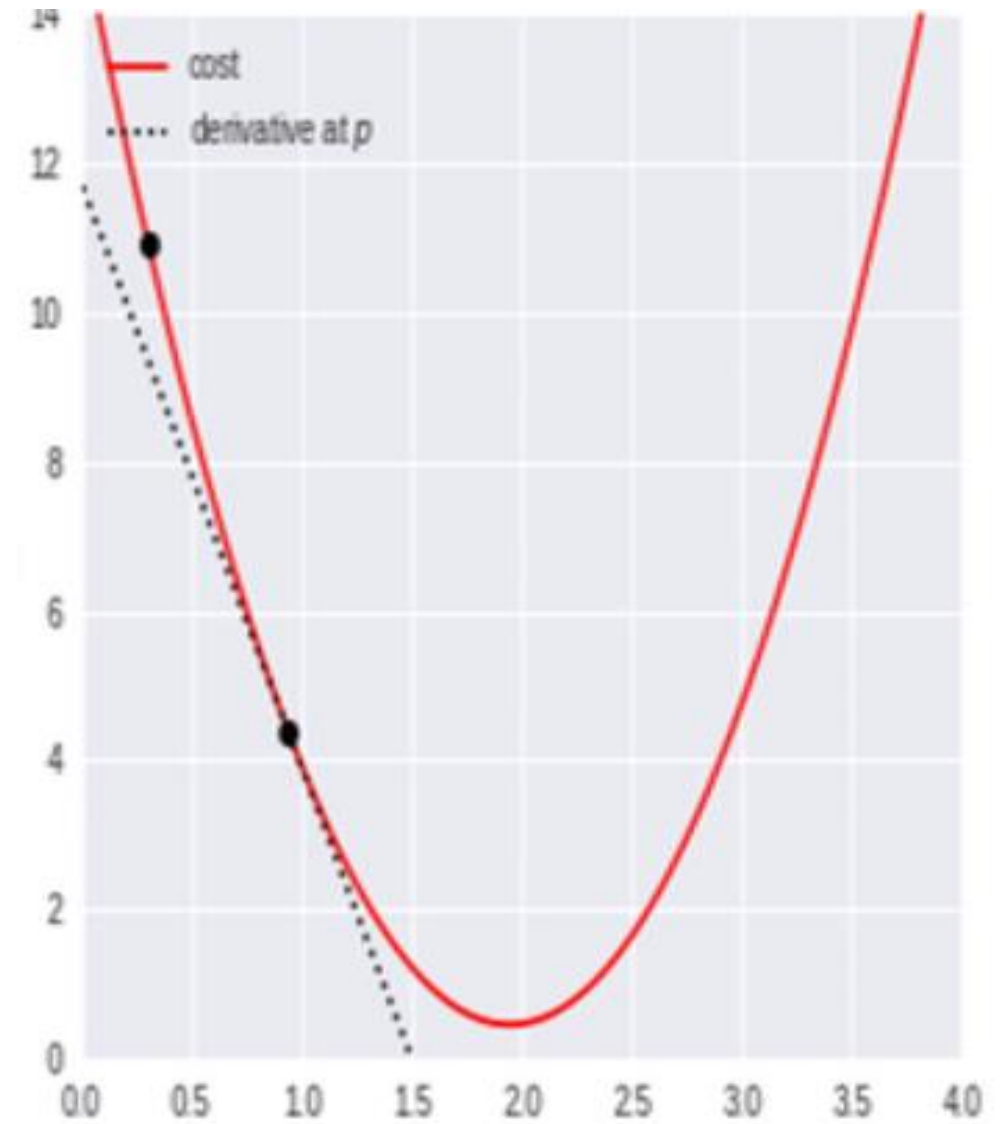
Gradiente



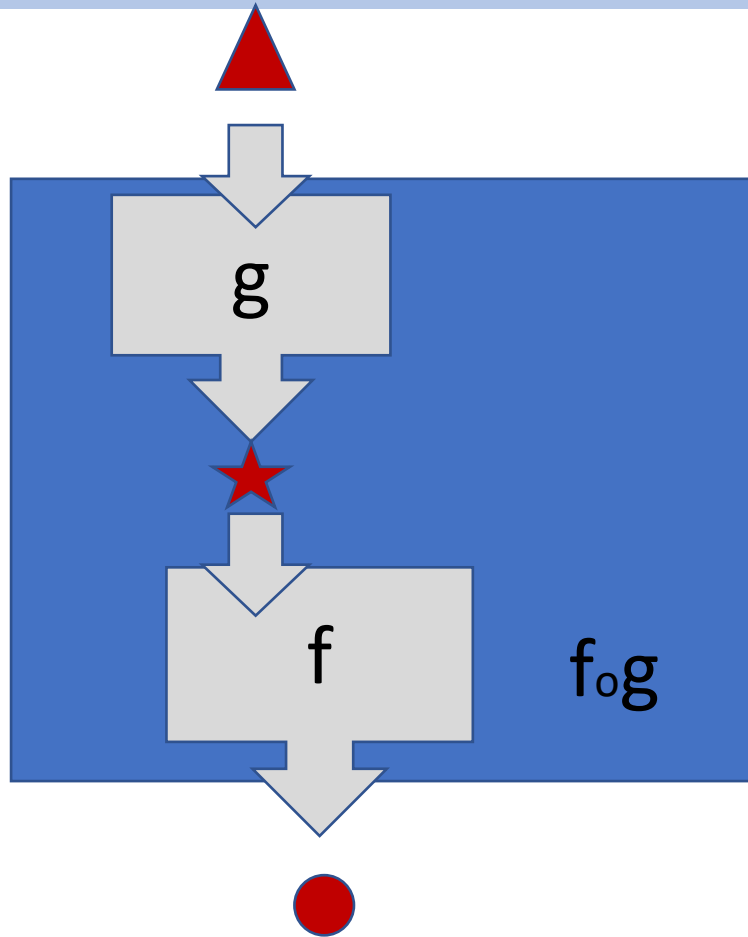
Como en la función $1/2(t-y)^2$

$$y = \sigma(z) = \frac{1}{1 + e^{-zy}}$$

Necesitamos usar la regla de la cadena para la derivada

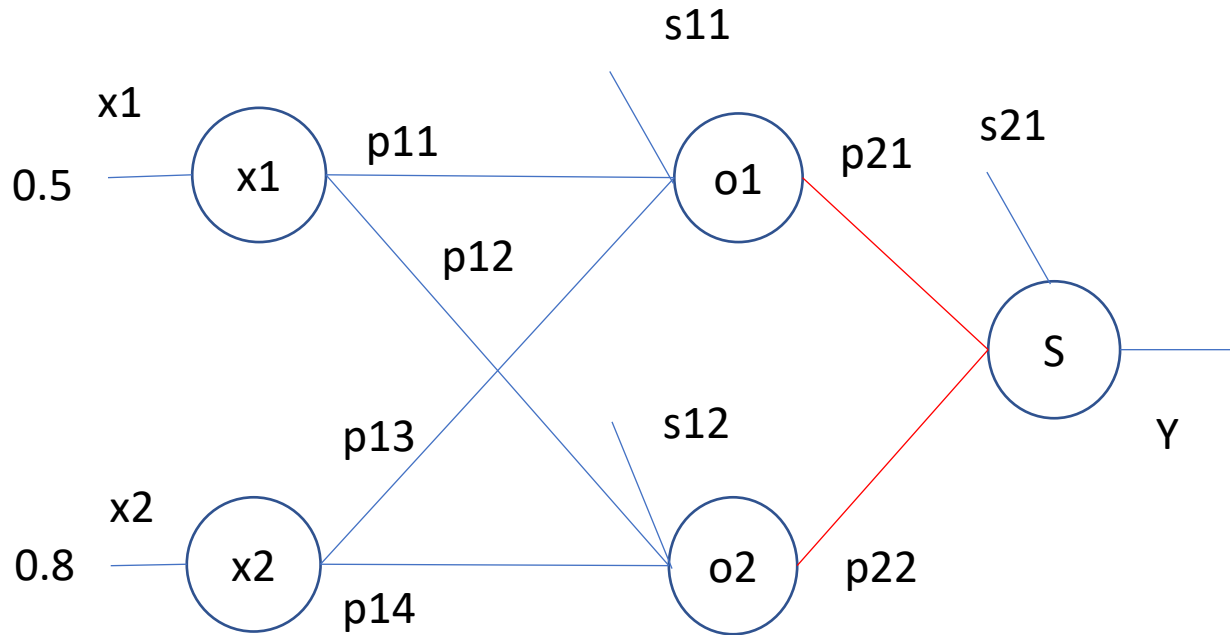


Regla de la cadena



$$\frac{d \bullet}{d \blacktriangle} = \frac{d \bullet}{d \star} * \frac{d \star}{d \blacktriangle}$$

Aplicando rc



$$\frac{d \bullet}{d \blacktriangle} = \frac{d \bullet}{d \star} * \frac{d \star}{d \blacktriangle}$$

El error con respecto al peso 21 se calcula:

El error depende de la salida y

La salida depende de suma_s

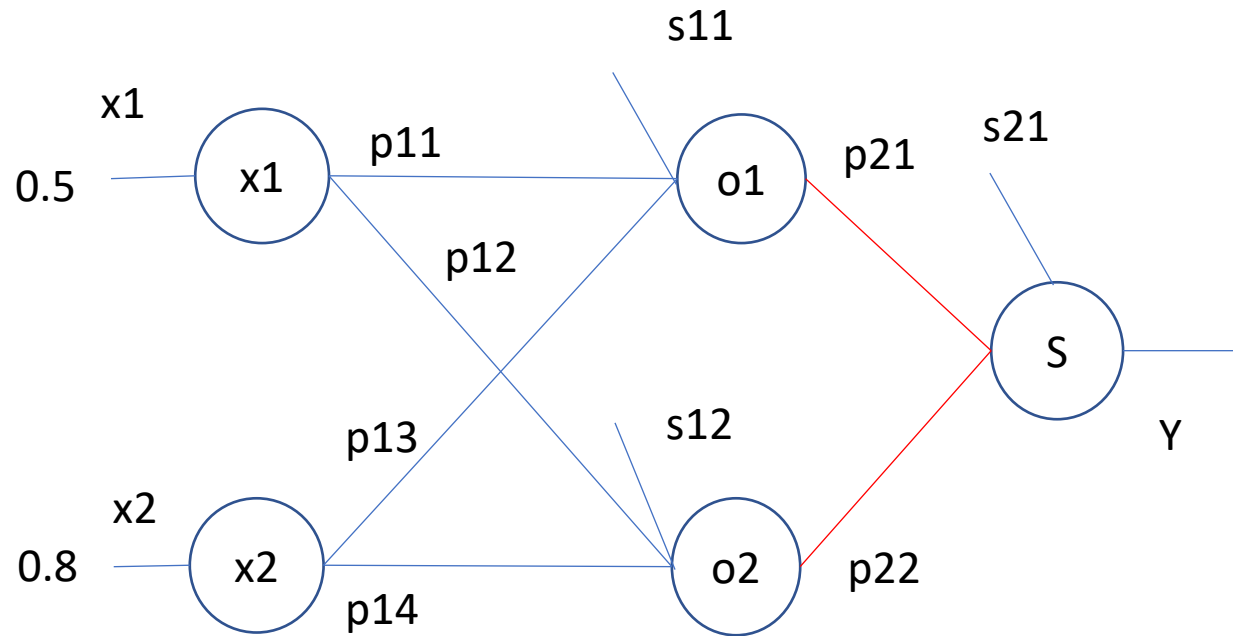
Y suma_s depende de p21

Aplicando regla de la cadena (2 veces)

$$\frac{\partial E}{\partial p21} = \frac{\partial E}{\partial y} * \frac{\partial y}{\partial \text{suma}_s} * \frac{\partial \text{suma}_s}{\partial p21}$$

∂ es el símbolo de la derivada parcia

Sacando derivadas parciales

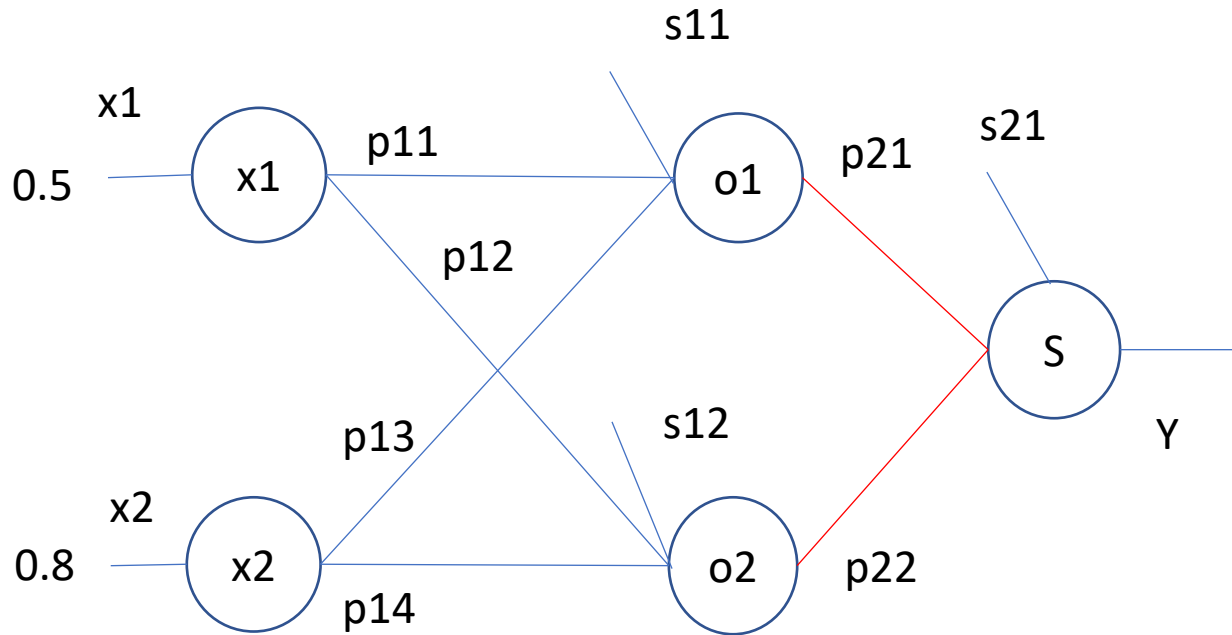


El error con respecto al peso 21

$$\frac{\partial E}{\partial p_{21}} = \frac{\partial E}{\partial y} * \frac{\partial y}{\partial suma_s} * \frac{\partial suma_s}{\partial p_{21}}$$

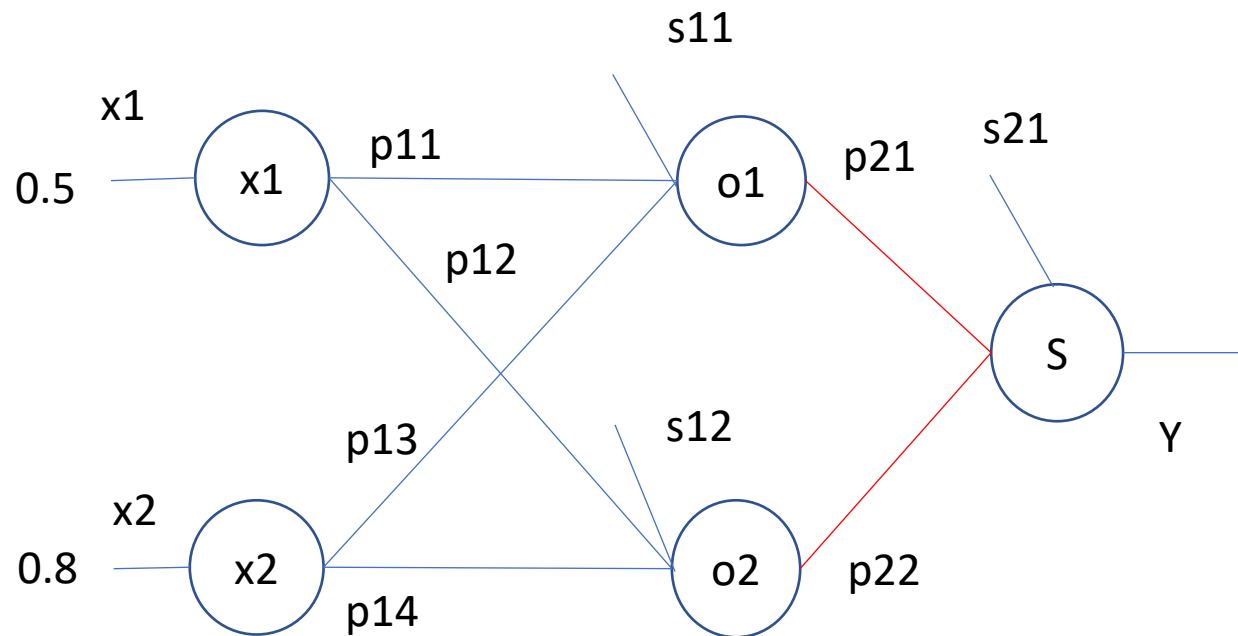
Primera derivada parcial

El error con respecto al peso 21



$$\frac{\partial \text{suma}_s}{\partial p_{21}} = \frac{\partial \text{salida}_{o_1} p_{21} + \text{salida}_{o_2} p_{22} + \text{sesgo}_{21}}{\partial p_{21}} = \text{salida}_{o_1}$$

2da y 3ra



$$\frac{\partial E}{\partial y} = \frac{\partial 1/2(t-y)^2}{\partial y} = y - t$$

$$\frac{\partial suma_s}{\partial p_{21}} = \frac{\partial salida_{o_1} p_{21} + salida_{o_2} p_{22} + sesgo_{21}}{\partial p_{21}} = salida_{o_1}$$

Simplificando

Por tanto

$$\frac{\partial E}{\partial p_{21}} = (y - t) * (y(1 - y)) * \text{salidao1}$$

$$\text{Si: } \delta s = y - t * y * (1 - y)$$

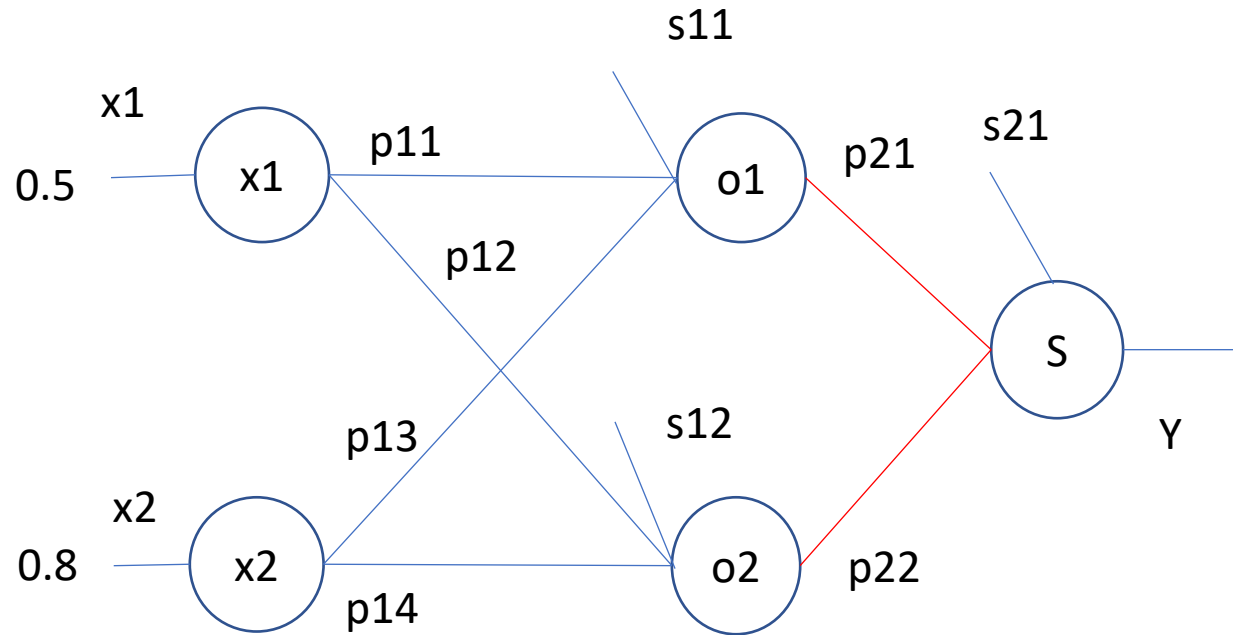
$$\frac{dE}{dp_{21}} = \delta s * \text{salidao1}$$

De forma similar:

$$\frac{dE}{dp_{21}} = \delta s * \text{salidao2}$$

$$\frac{dE}{dsesgo_{21}} = \delta s * 1$$

Gradiente capa salida



Gradiente capa se salida:

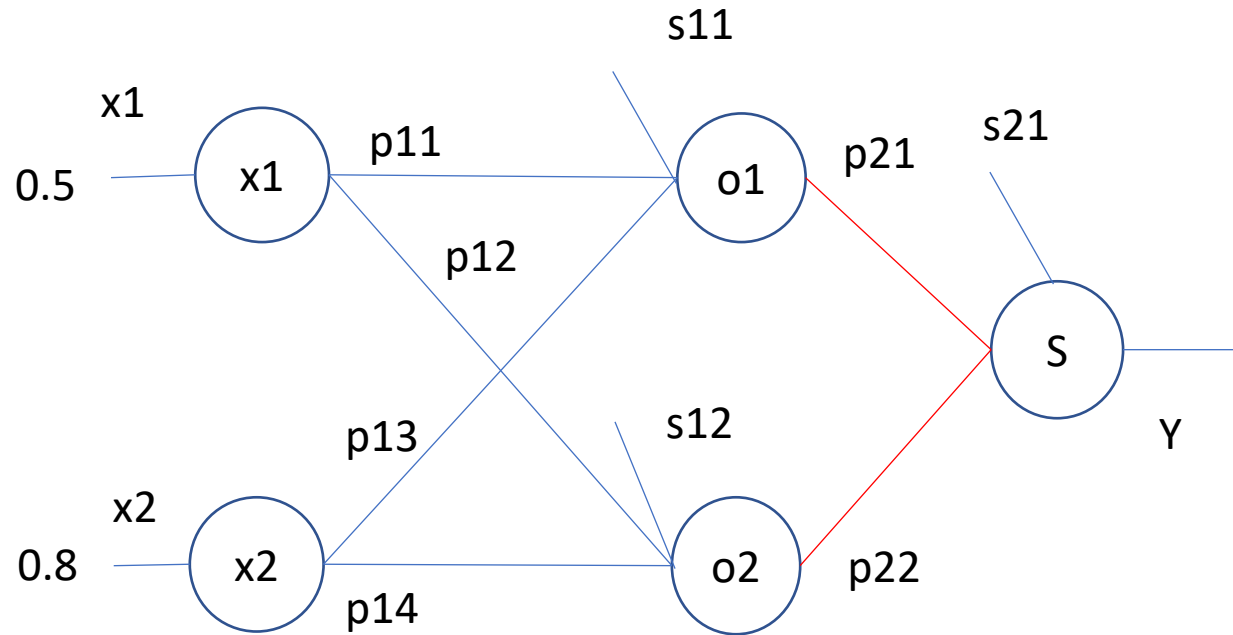
$$\frac{dE}{dp_{21}} = \delta_s * \text{salida}_{o1}$$

$$\frac{dE}{dp_{22}} = \delta_s * \text{salida}_{o2}$$

$$\frac{dE}{ds_{ego21}} = \delta_s * 1$$

Gradiente capa oculta

Calculamos:



$$\frac{\partial E}{\partial p11} = \frac{\partial E}{\partial y} * \frac{\partial y}{\partial suma_s} * \frac{\partial suma_s}{\partial salido1} * \frac{\partial salido1}{\partial sumao1} * \frac{\partial sumao1}{\partial p11}$$

jafh

Gradiente capa oculta

$$\frac{\partial E}{\partial p_{11}} = y - t * y(1 - y) * p_{21} * \text{salidao1}(1 - \text{salidao1}) * x_1$$

$$\frac{\partial E}{\partial p_{13}} = y - t * y(1 - y) * p_{21} * \text{salidao1}(1 - \text{salidao1}) * x_2$$

$$\frac{\partial E}{\partial \text{sesgo11}} = y - t * y(1 - y) * p_{21} * \text{salidao1}(1 - \text{salidao1}) * 1$$

$$\frac{\partial E}{\partial p_{12}} = y - t * y(1 - y) * p_{22} * \text{salidao2}(1 - \text{salidao2}) * x_1$$

$$\frac{\partial E}{\partial p_{14}} = y - t * y(1 - y) * p_{22} * \text{salidao2}(1 - \text{salidao2}) * x_2$$

$$\frac{\partial E}{\partial \text{psesgo12}} = y - t * y(1 - y) * p_{22} * \text{salidao2}(1 - \text{salidao2}) * 1$$

Gradiente capa oculta

$$\frac{\partial E}{\partial p_{11}} = \delta_s * p_{21} * \text{salidao1}(1 - \text{salidao1}) * x_1$$

$$\frac{\partial E}{\partial p_{13}} = \delta_s * p_{21} * \text{salidao1}(1 - \text{salidao1}) * x_2$$

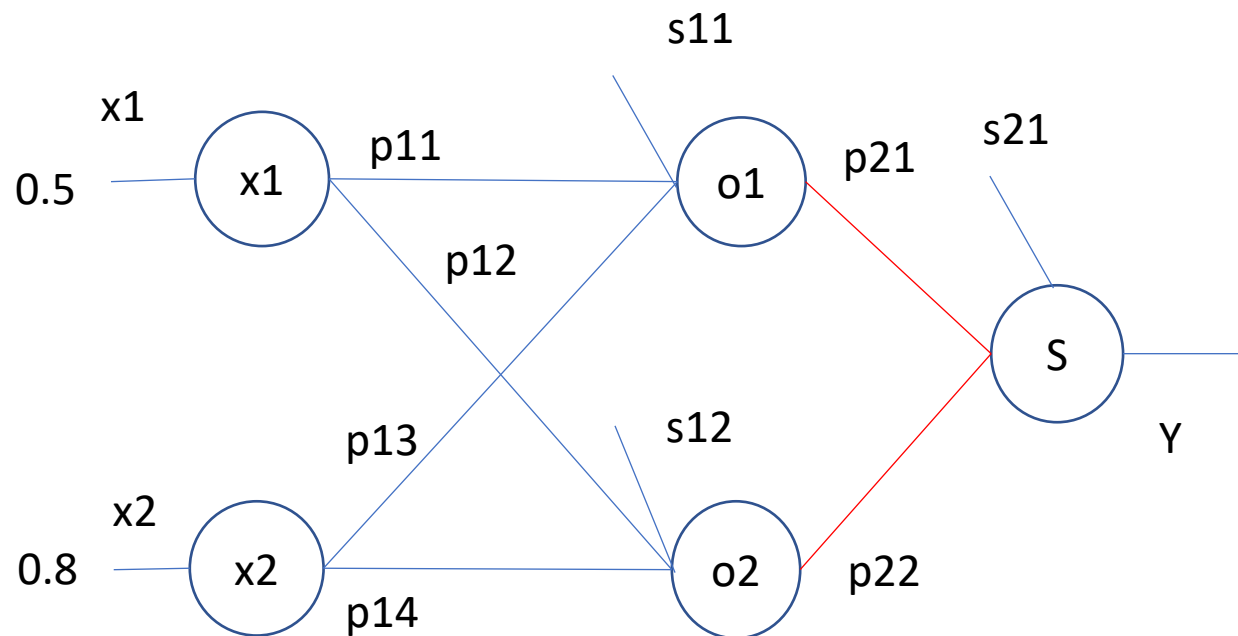
$$\frac{\partial E}{\partial \text{sesgo11}} = \delta_s * p_{21} * \text{salidao1}(1 - \text{salidao1}) * 1$$

$$\frac{\partial E}{\partial p_{12}} = \delta_s * p_{22} * \text{salidao2}(1 - \text{salidao2}) * x_1$$

$$\frac{\partial E}{\partial p_{14}} = \delta_s * p_{22} * \text{salidao2}(1 - \text{salidao2}) * x_2$$

$$\frac{\partial E}{\partial \text{ps sesgo12}} = \delta_s * p_{22} * \text{salidao2}(1 - \text{salidao2}) * 1$$

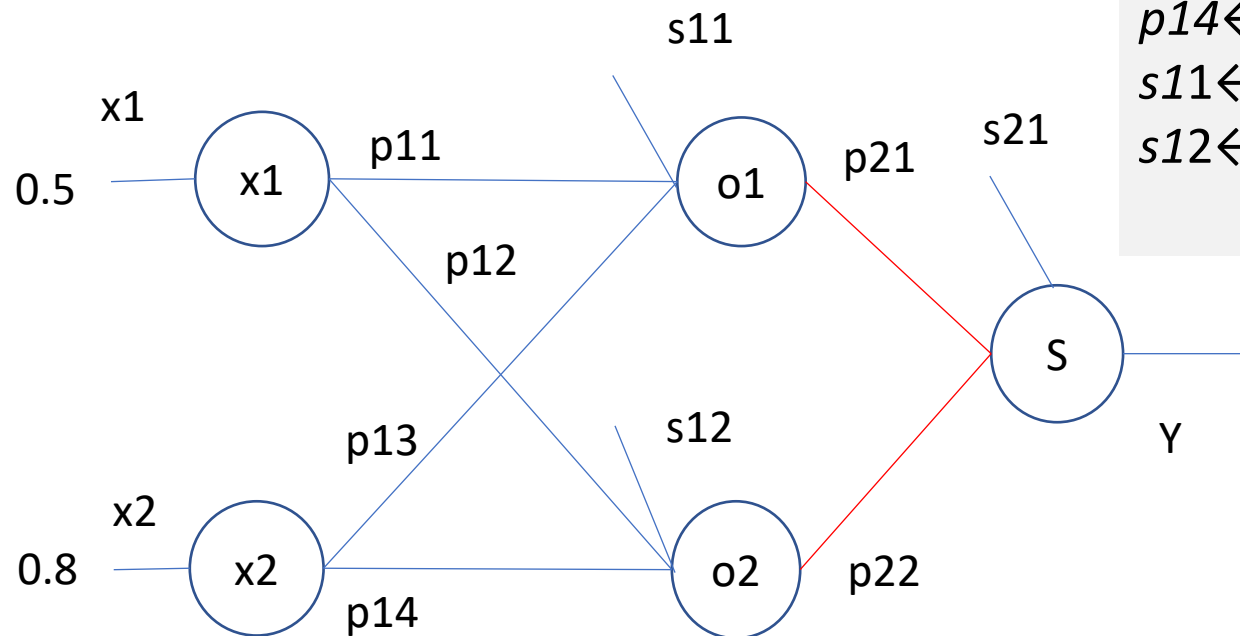
Aplicando Retro propagación a capa de salida



Actualizamos los pesos y sesgos de la capa de salida:

1. $p21 \leftarrow p21 - \eta \cdot \delta_s \cdot \text{salida}_{o1}$
2. $p22 \leftarrow p22 - \eta \cdot \delta_s \cdot \text{salida}_{o2}$
3. $s21 \leftarrow s21 - \eta \cdot \delta_s \cdot 1$

Aplicando Retro propagación a Capa oculta



Actualizamos los pesos y sesgos de la capa oculta:

$$\begin{aligned} p11 &\leftarrow p11 - \eta \cdot \delta_s \cdot p21 * salidao1(1 - salidao1) * x1 \\ p13 &\leftarrow p13 - \eta \cdot \delta_s \cdot p21 * salidao1(1 - salidao1) * x2 \\ p12 &\leftarrow p12 - \eta \cdot \delta_s \cdot p22 * salidao2(1 - salidao2) * x1 \\ p14 &\leftarrow p14 - \eta \cdot \delta_s \cdot p22 * salidao2(1 - salidao2) * x2 \\ s11 &\leftarrow s11 - \eta \cdot \delta_s \cdot p21 * salidao1(1 - salidao1) * 1 \\ s12 &\leftarrow s12 - \eta \cdot \delta_s \cdot p22 * salidao2(1 - salidao2) * 1 \end{aligned}$$

FIN

JAFH

Refs:

- <https://www.youtube.com/watch?v=iOsR-EC9z6I>