UNIVERSIDAD
NACIONAL
DE COLOMBIA

Analítica Predictiva

Carlos A. Madrigal

Profesor Ocasional

Departamento de Ciencias de la Computación y de la Decisión

Maestría en ingeniería - ingeniería de sistemas

Maestría en ingeniería - analítica

Especialización en sistemas

Nota: Este material se ha adaptado con base a diferentes fuentes de información académica

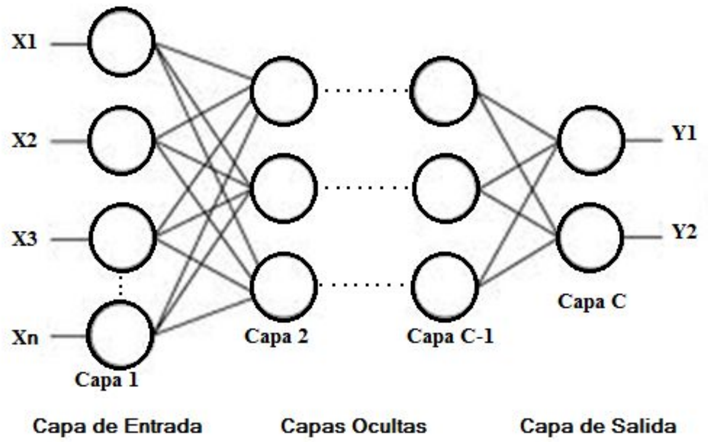
Contenido

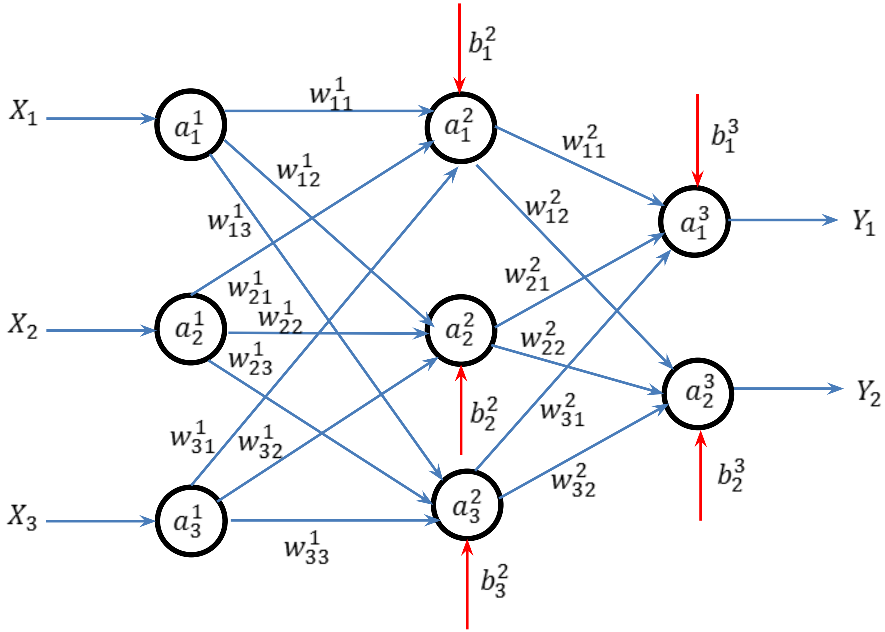
Introducción a las Redes Neuronales Profundas

- Perceptrón Multicapa
- Forward
- Funciones de activación, funciones de error, costo
- Descenso del gradiente
- Backpropagation
- Implementacíon

Introducción a las Redes Neuronales Profundas

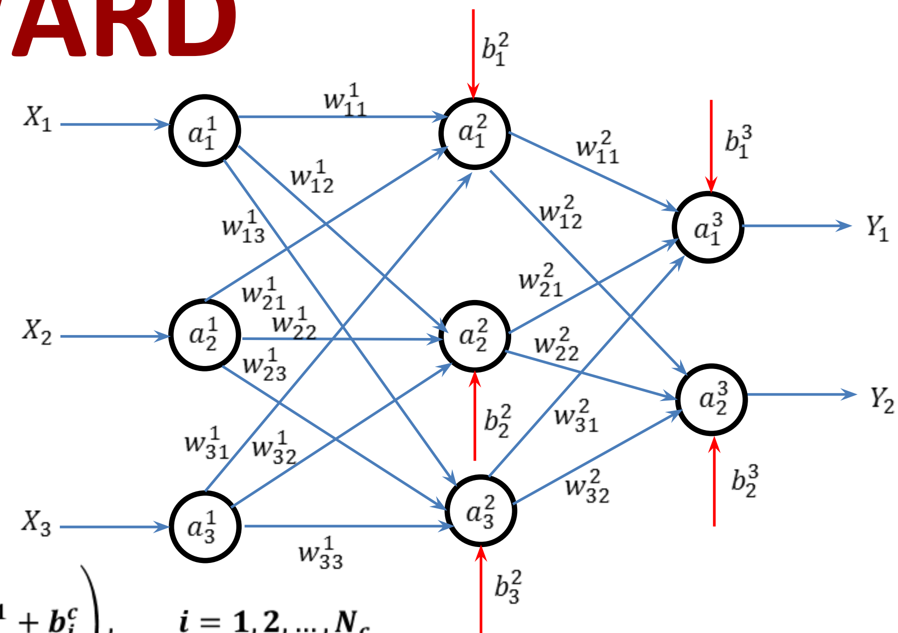
Perceptron Multicapa





- $C = \#\ de\ Capas$
- $Capas\ de\ Entrada = 1$
- $Capas\ Ocultas = C - 2$
- $Capas\ de\ Salida = 1$
- $W_{ij}^q = representa\ el\ peso\ de\ la\ conexion\ de\ la\ neurona\ i\ de\ la\ capa\ q\ con\ la\ neurona\ j\ de\ la\ capa\ q+1.\quad q = 1, 2, 3 \ldots, C\text{-}1$
- $b_i^q = Vector\ de\ umbrales\ de\ las\ neuronas\ dela\ capa\ q,\ \ q = 2, 3 \ldots, C$
- $a_i^q = Activación\ de\ la\ neurona\ i\ de\ la\ capa\ q.$
- $N_q = \#de\ Neuronas\ en\ la\ capa\ q, \qquad q = 1, 2, 3 \ldots, Q$
- $W^q = Matriz\ de\ Pesos\ de\ la\ Red\ de\ \ la\ capa\ q\ a\ q\text{+}1$

FORWARD



Activación de las Capas

Capa de Entrada:

$$a_i^q = a_i^1 = x_i, \qquad i = 1, 2, \ldots n$$

Capa de Salida:

$$a_i^c = y_i = f\left(\sum_{j=1}^{N_{(c-1)}} W_{ji}^{c-1}\, a_j^{c-1} + b_i^c\right), \qquad i = 1, 2, \ldots, N_c$$

C

C-1

j

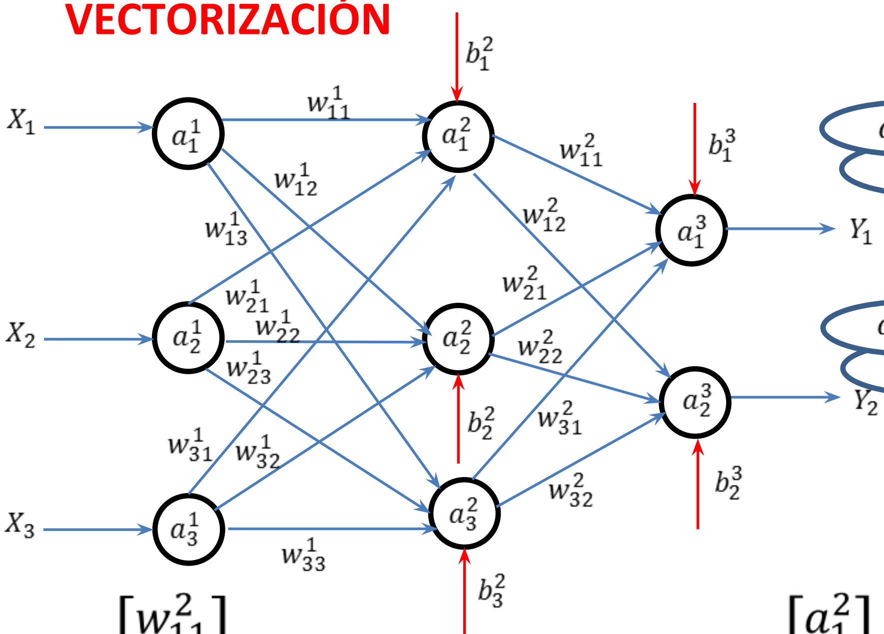
$$a_1^3 = f(w_{11}^2 * a_1^2 + w_{21}^2 * a_2^2 + w_{31}^2 * a_3^2 + b_1^3)$$

Capas Ocultas:

$$a_i^q = f\left(\sum_{j=1}^{N_{(q-1)}} W_{ji}^{q-1}\, a_j^{q-1} + b_i^q\right), \qquad i = 1, 2, \ldots, N_q, q = 2, 3, \ldots, C-1$$

FORWARD

VECTORIZACIÓN



$$a_1^3 = f(z_1^3) = f(w_{11}^2 * a_1^2 + w_{21}^2 * a_2^2 + w_{31}^2 * a_3^2 + u_1^3)$$

$$a_2^3 = f(z_2^3) = f(w_{12}^2 * a_1^2 + w_{22}^2 * a_2^2 + w_{32}^2 * a_3^2 + u_2^3)$$

$$z^3 = \underset{\substack{2x1 \\ N_3x1}}{\begin{bmatrix} z_1^3 \\ z_2^3 \end{bmatrix}} = \underset{2x3}{\begin{bmatrix} w_{11}^2 & w_{21}^2 & w_{31}^2 \\ w_{12}^2 & w_{22}^2 & w_{32}^2 \end{bmatrix}} \underset{3x1}{\begin{bmatrix} a_1^2 \\ a_2^2 \\ a_3^2 \end{bmatrix}} + \underset{2x1}{\begin{bmatrix} b_1^3 \\ b_2^3 \end{bmatrix}}$$

$$w_{:,1}^2 = \begin{bmatrix} w_{11}^2 \\ w_{21}^2 \\ w_{31}^2 \end{bmatrix} \qquad W^2 = \underset{\substack{3x2 \\ N_2xN_3}}{\begin{bmatrix} w_{11}^2 & w_{12}^2 \\ w_{21}^2 & w_{22}^2 \\ w_{31}^2 & w_{32}^2 \end{bmatrix}} \qquad a^2 = \underset{\substack{3x1 \\ N_2x1}}{\begin{bmatrix} a_1^2 \\ a_2^2 \\ a_3^2 \end{bmatrix}}$$

$$\boxed{\boldsymbol{z^3 = W^{2^T} a^2 + b^3}}$$

$$w_{:,2}^2 = \begin{bmatrix} w_{12}^2 \\ w_{22}^2 \\ w_{32}^2 \end{bmatrix} \qquad b^3 = \underset{\substack{2x1 \\ N_3x1}}{\begin{bmatrix} b_1^3 \\ b_2^3 \end{bmatrix}}$$

$$\boxed{\boldsymbol{a^3} = \begin{bmatrix} a_1^3 \\ a_2^3 \end{bmatrix} = f\left(\begin{bmatrix} z_1^3 \\ z_2^3 \end{bmatrix}\right) = \boldsymbol{f^3(z^3) = y}}$$

FORWARD

VECTORIZACIÓN

Capa 2

$$z^2 = W^{1^T} a^1 + b^2$$

$$a^2 = f^2(z^2)$$

Capa 3

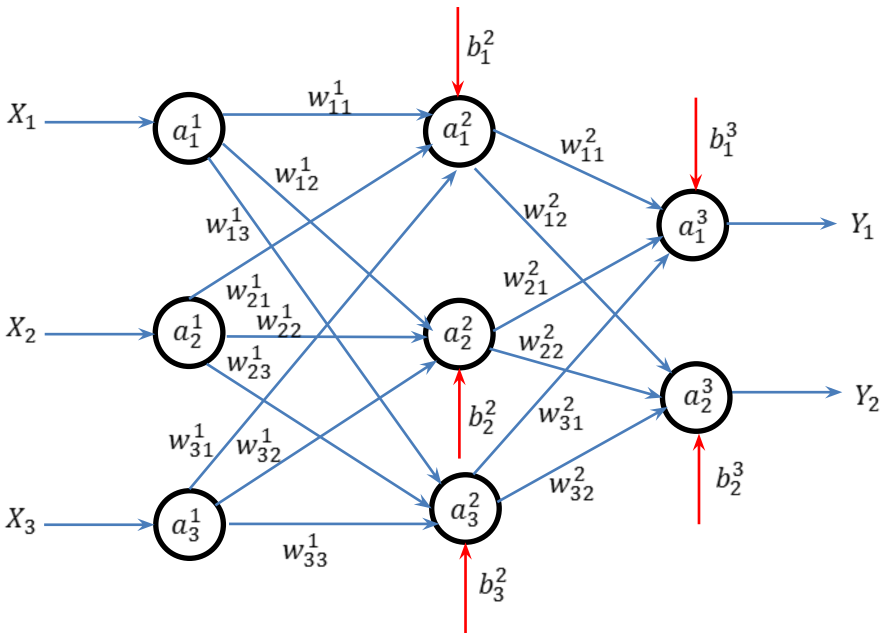
$$z^3 = W^{2^T} a^2 + b^3$$

$$a^3 = f^3(z^3)$$

Capa *l* (*cualquier Capa*)

$$z^l = W^{l-1^T} a^{l-1} + b^l$$

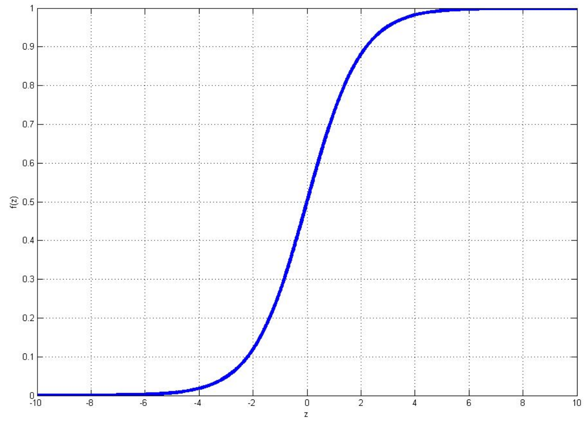
$$a^l = f^l(z^l)$$



Funciones de Activación

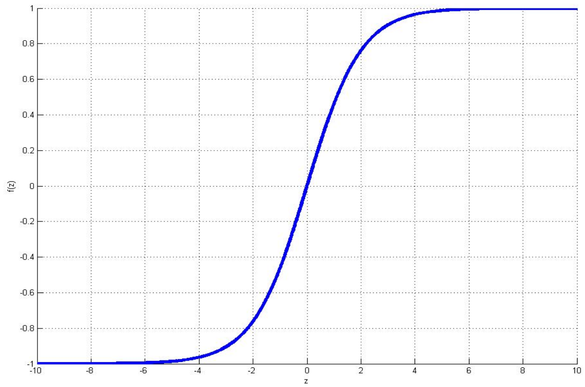
Función Sigmoidal:

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



Función Tangente Hiperbólica:

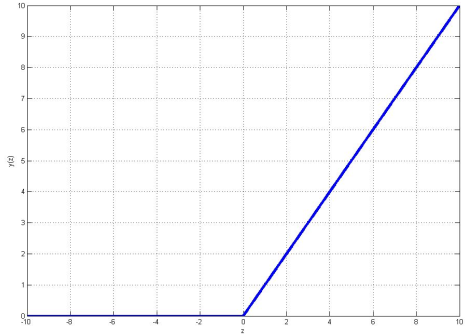
$$\mathrm{f}(z) = \frac{1- e^{-z}}{1+e^{-z}}$$



Funciones de Activación

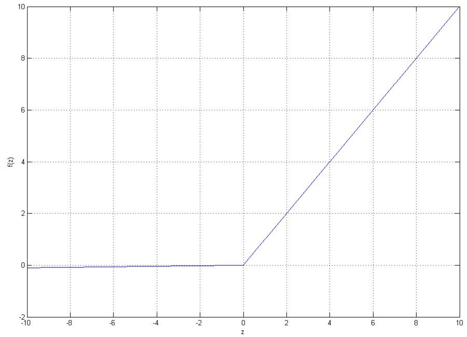
Función RELU:

$$f(z) = \max(0, z)$$



Función Leaky Relu:

$$f(z) = \max(0.01z, z)$$



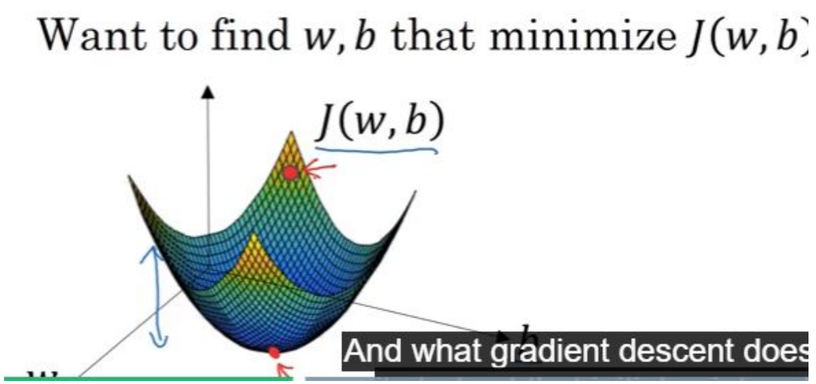
Funciones de Error

Median Squared Error

$$\mathcal{L}(\boldsymbol{S},\boldsymbol{Y}) = \boldsymbol{e}(\boldsymbol{n}) = \frac{\mathbf{1}}{\mathbf{2}}(\boldsymbol{S}-\boldsymbol{Y})^{\mathbf{2}}$$

$S(n)$: Salidas deseadas de la red para el patrón n

Y(n): Vector de salida de la red para el patrón n



Logistic Regression Loss Function

$$\mathcal{L}(\boldsymbol{S},\boldsymbol{Y}) = \boldsymbol{e}(\boldsymbol{n}) = -(\boldsymbol{S}\log \boldsymbol{Y} + (\mathbf{1}-\boldsymbol{S})\log(1-\boldsymbol{Y}))$$

Cost Function

$$\boldsymbol{J}(\boldsymbol{w},\boldsymbol{b}) = \boldsymbol{E} = \frac{\mathbf{1}}{\boldsymbol{N}}\sum_{\boldsymbol{n}=\mathbf{1}}^{\boldsymbol{N}} \mathcal{L}(\boldsymbol{S},\boldsymbol{Y}) = \frac{\mathbf{1}}{\boldsymbol{N}}\sum_{\boldsymbol{n}=\mathbf{1}}^{\boldsymbol{N}} \boldsymbol{e}(\boldsymbol{n})$$

Descenso del Gradiente

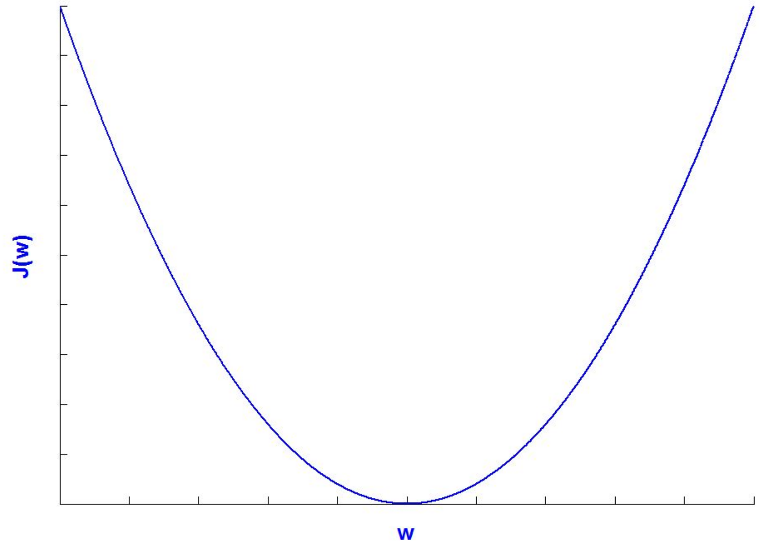
Método del Descenso del Gradiente Estocástico

$$\boldsymbol{w(n) = w(n-1) - \alpha \frac{\partial e(n)}{\partial w}},$$

α: factor de aprendizaje

Método del Descenso del Gradiente

$$\boldsymbol{w(n) = w(n-1) - \alpha \frac{\partial J(w,b)}{\partial w}},$$



BACKPROPAGATION

Como se modifican los Pesos a la Capa de Salida y Umbrales de la capa de salida

$$w_{ji}^{c-1}(n) = w_{ji}^{c-1}(n-1) - \alpha \frac{\partial e(n)}{\partial w_{ji}^{c-1}}$$

$$\frac{\partial e(n)}{\partial w_{ji}^{c-1}} = \frac{\partial e(n)}{\partial y_i(n)} \frac{\partial y_i(n)}{\partial w_{ji}^{c-1}} = \frac{\partial e(n)}{\partial a_i^c(n)} \frac{\partial a_i^c(n)}{\partial w_{ji}^{c-1}} =$$

¡Recordemos!

$$a_i^c = y_i = f(z_i^c) = f\left(\sum_{j=1}^{N_{(c-1)}} W_{ji}^{c-1}\, a_j^{c-1} + b_i^c\right)$$

$$\frac{d}{dx} g(f(x)) = g'(f(x)) \cdot f'(x)$$

$$\frac{\partial a_i^c(n)}{\partial w_{ji}^{c-1}} = f'\left(\sum_{j=1}^{N_{c-1}} W_{ji}^{c-1}\, a_j^{c-1} + b_i^c\right) a_j^{c-1}(n)$$

$$\frac{\partial a_i^c(n)}{\partial w_{ji}^{c-1}} = f'(z_i^c)\, a_j^{c-1}(n)$$

$$\frac{\partial e(n)}{\partial w_{ji}^{c-1}} = \frac{\partial e(n)}{\partial y_i(n)} f'(z_i^c)\, a_j^{c-1}(n)$$

BACKPROPAGATION

Cómo se modifican los Pesos a la Capa de Salida y Umbrales de la capa de salida

Con Error Cuadrático Medio como función de error

$$\frac{\partial e(n)}{\partial w_{ji}^{c-1}} = -\big(S_i(n) - Y_i(n)\big) f'(z_i^c)\, a_j^{c-1}(n)$$

Error Cuadrático Medio

$$e(n) = \frac{1}{2}\sum_{i=1}^{N_c}\big(S_i(n) - Y_i(n)\big)^2$$

$$\delta_i^c(n) = -\big(S_i(n) - Y_i(n)\big) f'(z_i^c)$$

Término que contiene el error cometido por la red para la neurona i de la capa c

$$\frac{\partial e(n)}{\partial w_{ji}^{c-1}} = \delta_i^c(n) a_j^{c-1}(n) \qquad \frac{\partial e(n)}{\partial z_i^c} = \delta_i^c(n)$$

$$w_{ji}^{c-1}(n) = w_{ji}^{c-1}(n-1) - \alpha\delta_i^c(n) a_j^{c-1}(n), \qquad j = 1, 2, \ldots, N_{c-1}, i = 1, .., N_c$$

$$b_i^c(n) = b_i^c(n-1) - \alpha\delta_i^c(n), \qquad , i = 1, .., N_c$$

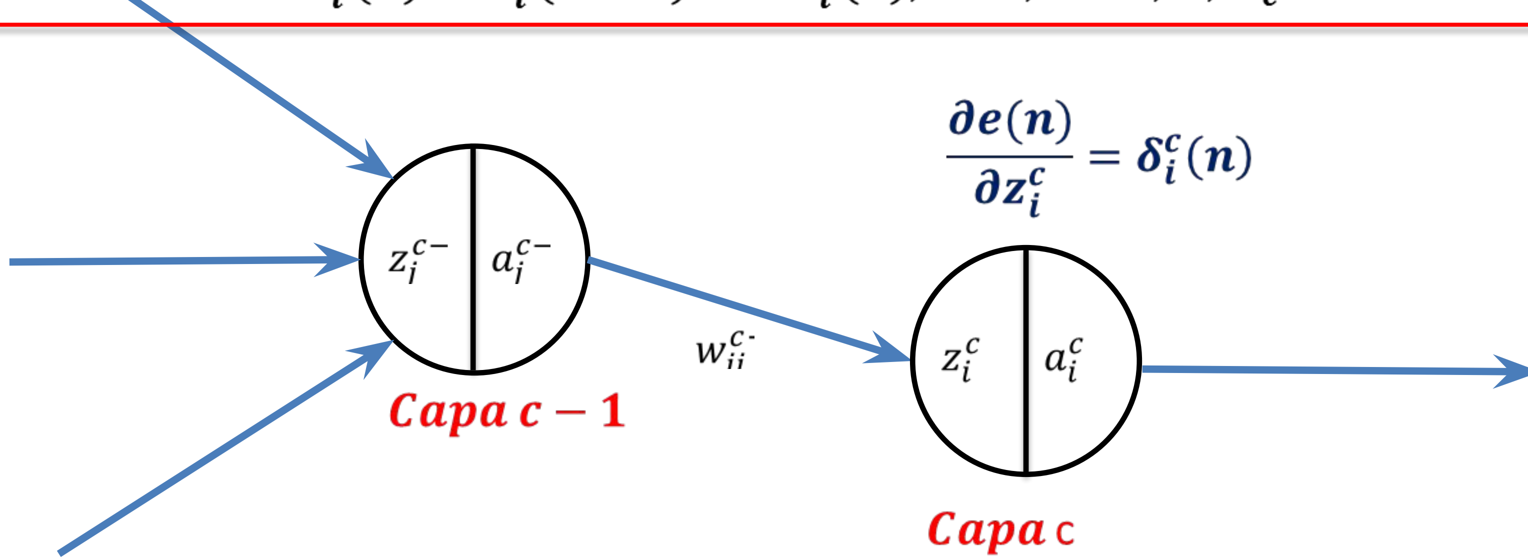
BACKPROPAGATION

Cómo se modifican los Pesos a la Capa de Salida y Umbrales de la capa de salida

Con Error Cuadrático Medio como función de error

$$w_{ji}^{c-1}(n) = w_{ji}^{c-1}(n-1) - \alpha \delta_i^c(n) a_j^{c-1}(n), \qquad j = 1, 2, \ldots, N_{c-1}, i = 1, .., N_c$$

$$b_i^c(n) = b_i^c(n-1) - \alpha \delta_i^c(n), \qquad , i = 1, .., N_c$$



BACKPROPAGATION

VECTORIZACIÓN

Como se modifican los Pesos a la Capa de Salida y Umbrales de la capa de salida

$$w^{c-1}(n) = w^{c-1}(n-1) - \alpha \frac{\partial e(n)}{\partial w^{c-1}} \longrightarrow \frac{\partial e(n)}{\partial w^{c-1}} = \frac{\partial e(n)}{\partial y(n)} \frac{\partial y(n)}{\partial w^{c-1}} = \frac{\partial e(n)}{\partial a^c(n)} \frac{\partial a^c(n)}{\partial w^{c-1}}$$

$$z^c = {W^{c-1}}^T a^{c-1} + b^c$$

$$\longrightarrow \frac{\partial a^c(n)}{\partial w^{c-1}} = f'(z^c) a^{c-1}$$

$$a^c(n) = y(n) = f(z^c)$$

$$\boxed{\frac{\partial e(n)}{\partial w^{c-1}} = \frac{\partial e(n)}{\partial a^c(n)} f'(z^c) a^{c-1}}$$

BACKPROPAGATION

Modificación de los pesos de Otras Capas (C-2)->(C-1).

$$w_{kj}^{c-2}(n) = w_{kj}^{c-2}(n-1) - \alpha \frac{\partial e(n)}{\partial w_{kj}^{c-2}} \longrightarrow \frac{\partial e(n)}{\partial w_{kj}^{c-2}} = \frac{\partial e(n)}{\partial a_j^{c-1}} \frac{\partial a_j^{c-1}(n)}{\partial w_{kj}^{c-2}}$$

$$a_j^{c-1}(n) = f\left(z_j^{c-1}\right) = f\left(\sum_{k=1}^{N_{c-2}} W_{kj}^{c-2}\, a_k^{c-2} + b_j^{c-1}\right)$$

$$\frac{\partial a_j^{c-1}}{\partial w_{kj}^{c-2}} = f'\left(z_j^{c-1}\right) a_k^{c-2}(n) \longrightarrow \frac{\partial e(n)}{\partial w_{kj}^{c-2}} = \frac{\partial e(n)}{\partial a_j^{c-1}} f'\left(z_j^{c-1}\right) a_k^{c-2}(n)$$

$$\frac{\partial e(n)}{\partial w_{kj}^{c-2}} = \frac{\partial e(n)}{\partial z_j^{c-1}} a_k^{c-2}(n)$$

Con Error Cuadrático Medio como función de error

$$\frac{\partial e(n)}{\partial a_j^{c-1}} = -\sum_{i=1}^{Nc} \left(S_i(n) - Y_i(n)\right) \frac{\partial y_i(n)}{\partial a_j^{c-1}} = -\sum_{i=1}^{Nc} \left(S_i(n) - Y_i(n)\right) f'(z_i^c) w_{ji}^{c-1}(n)$$

$$\frac{\partial e(n)}{\partial a_j^{c-1}} = \sum_{i=1}^{Nc} \delta_i^c(n) w_{ji}^{c-1}$$

BACKPROPAGATION

Modificación de los pesos de Otras Capas (C-2)->(C-1).

Con Error Cuadrático Medio como función de error

$$\frac{\partial e(n)}{\partial w_{kj}^{c-2}} = f'(z_j^{c-1}) a_k^{c-2}(n) \sum_{i=1}^{Nc} (\delta_i^c(n) w_{ji}^{c-1})$$

$$\delta_j^{c-1}(n) = f'\left(\sum_{k=1}^{N_{c-2}} W_{kj}^{c-2} a_k^{c-2} + b_j^{c-1}\right)\left(\sum_{i=1}^{N_c} \delta_i^c(n) W_{ji}^{c-1}\right)$$

Término que contiene el error cometido por la red para la neurona j de la capa c-1

$$\boxed{\begin{aligned} w_{kj}^{c-2}(n) &= w_{kj}^{c-2}(n-1) - \alpha \delta_j^{c-1}(n) a_k^{c-2}(n), \qquad k = 1, 2, \ldots, N_{c-2}, j = 1, .., N_{c-1} \\ b_j^{c-1}(n) &= b_j^{c-1}(n-1) - \alpha \delta_j^{c-1}(n), \qquad , j = 1, .., N_{c-1} \end{aligned}}$$

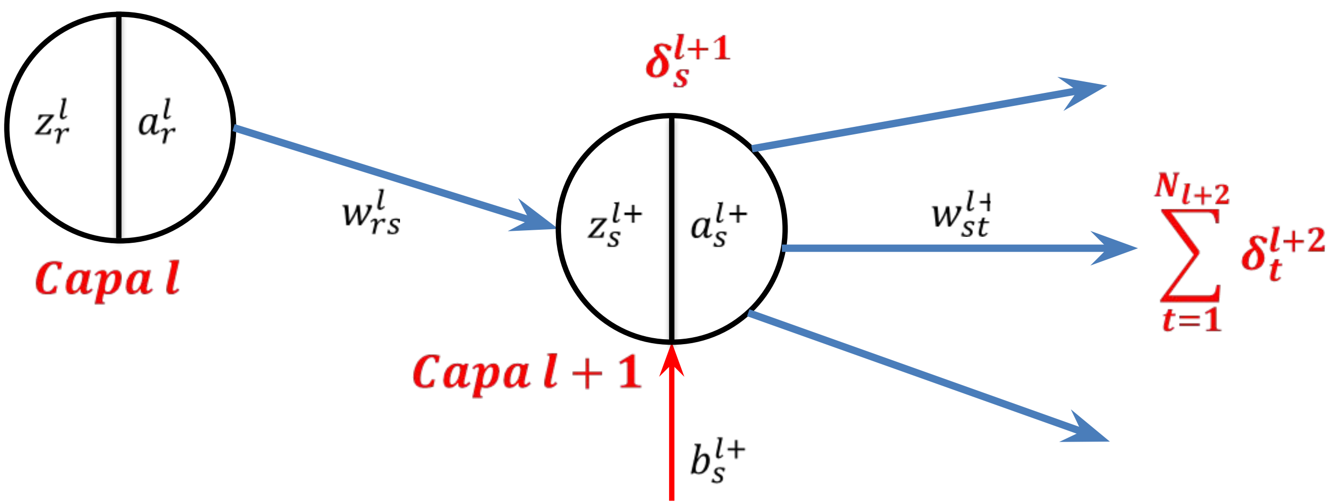
BACKPROPAGATION

Generalizando para cualquier capa

$$w_{rs}^{l}(n) = w_{rs}^{l}(n-1) - \alpha\delta_s^{l+1}(n)a_r^{l}(n)$$

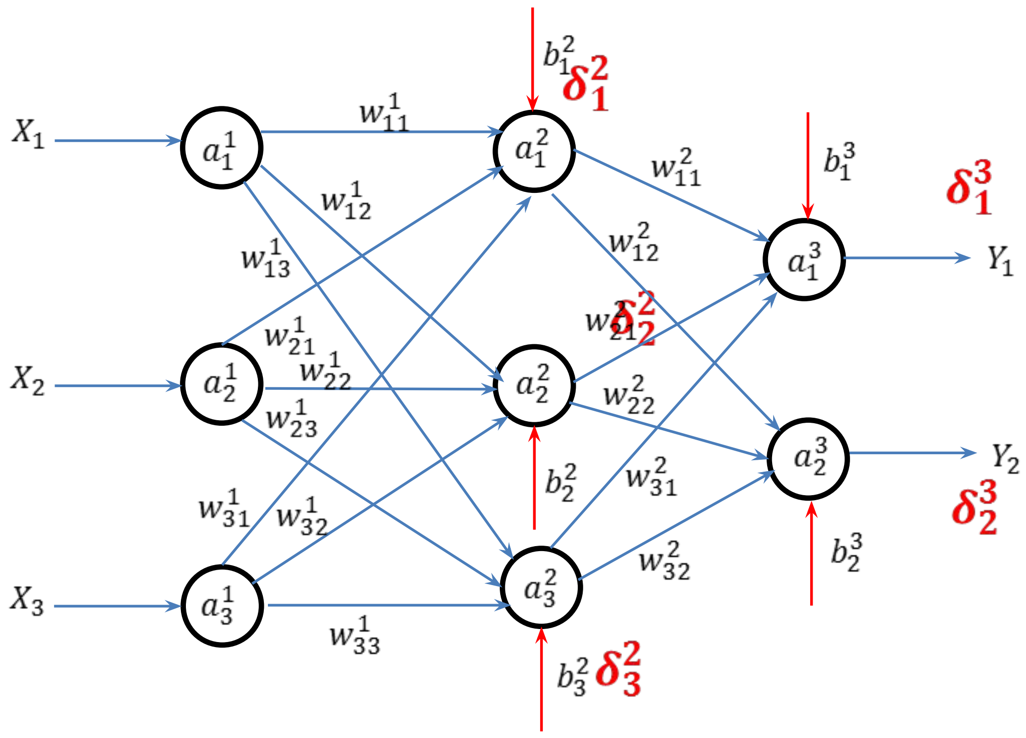
$$\delta_s^{l+1}(n) = f'\left(z_s^{l+1}\right)\left(\sum^{N_{l+2}} \delta_t^{l+2}(n)W_{st}^{l+1}\right) = f'\left(\sum^{N_l} W_{rs}^{l}\, a_r^{l} + b_s^{l+1}\right)\left(\sum^{N_{l+2}} \delta_t^{l+2}(n)W_{st}^{l+1}\right)$$

$$b_s^{l+1}(n) = b_s^{l+1}(n-1) - \alpha\delta_s^{l+1}(n) \qquad s = 1, 2, \ldots, N_{l+1},$$



BACKPROPAGATION

Generalizando para cualquier capa



BACKPROPAGATION

VECTORIZACIÓN

Modificación de los pesos de Otras Capas (C-2)->(C-1).

$$w^{c-2}(n) = w^{c-2}(n-1) - \alpha \frac{\partial e(n)}{\partial w^{c-2}} \longrightarrow \frac{\partial e(n)}{\partial w^{c-2}} = \frac{\partial e(n)}{\partial a^{c-1}(n)} \frac{\partial a^{c-1}(n)}{\partial w^{c-2}}$$

$$a^{c-1}(n) = f\left(z^{c-1}\right) \longrightarrow \frac{\partial a^{c-1}(n)}{\partial w^{c-2}} = f'\left(z^{c-1}\right) a^{c-2}(n)$$

$$z^{c-1}(n) = W^{c-2^{T}} a^{c-2} + b^{c-1}$$

$$\boxed{\frac{\partial e(n)}{\partial w^{c-2}} = \frac{\partial e(n)}{\partial a^{c-1}(n)} f'\left(z^{c-1}\right) a^{c-2}(n)} \longrightarrow \frac{\partial e(n)}{\partial a^{c-1}(n)} = \frac{\partial e(n)}{\partial a^{c}(n)} \frac{\partial a^{c}(n)}{\partial z^{c}(n)} \frac{\partial z^{c}(n)}{\partial a^{c-1}(n)}$$

$$\boxed{\frac{\partial e(n)}{\partial w^{c-2}} = \frac{\partial e(n)}{\partial z^{c-1}(n)} a^{c-2}(n)}$$

BACKPROPAGATION

VECTORIZACIÓN

Generalizando para cualquier capa

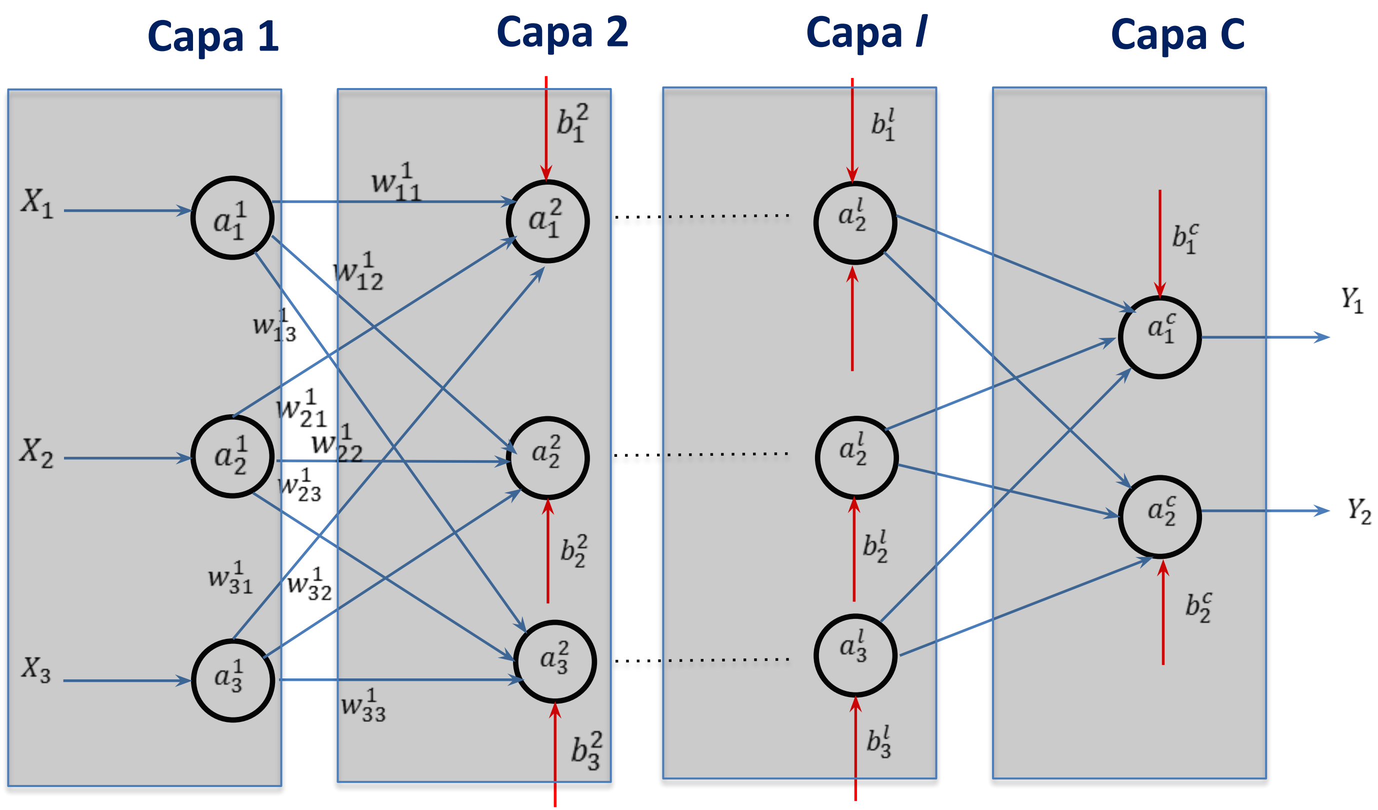
$$\boldsymbol{w}^{l}(\boldsymbol{n}) = \boldsymbol{w}^{l}(\boldsymbol{n}-\mathbf{1}) - \boldsymbol{\alpha}\frac{\partial \boldsymbol{e}(\boldsymbol{n})}{\partial \boldsymbol{w}^{l}} \longrightarrow \frac{\partial \boldsymbol{e}(\boldsymbol{n})}{\partial \boldsymbol{w}^{l}} = \frac{\partial \boldsymbol{e}(\boldsymbol{n})}{\partial \boldsymbol{a}^{l+1}(\boldsymbol{n})}\frac{\partial \boldsymbol{a}^{l+1}(\boldsymbol{n})}{\partial \boldsymbol{w}^{l}}$$

$$\boldsymbol{z}^{l+1}(\boldsymbol{n}) = \boldsymbol{W}^{l^{T}}\boldsymbol{a}^{l} + \boldsymbol{b}^{l+1}$$

$$\boldsymbol{a}^{l+1}(\boldsymbol{n}) = \boldsymbol{f}\left(\boldsymbol{z}^{l+1}\right) \longrightarrow \frac{\partial \boldsymbol{a}^{l+1}(\boldsymbol{n})}{\partial \boldsymbol{w}^{l}} = \boldsymbol{f}'\left(\boldsymbol{z}^{l+1}\right)\boldsymbol{a}^{l}(\boldsymbol{n})$$

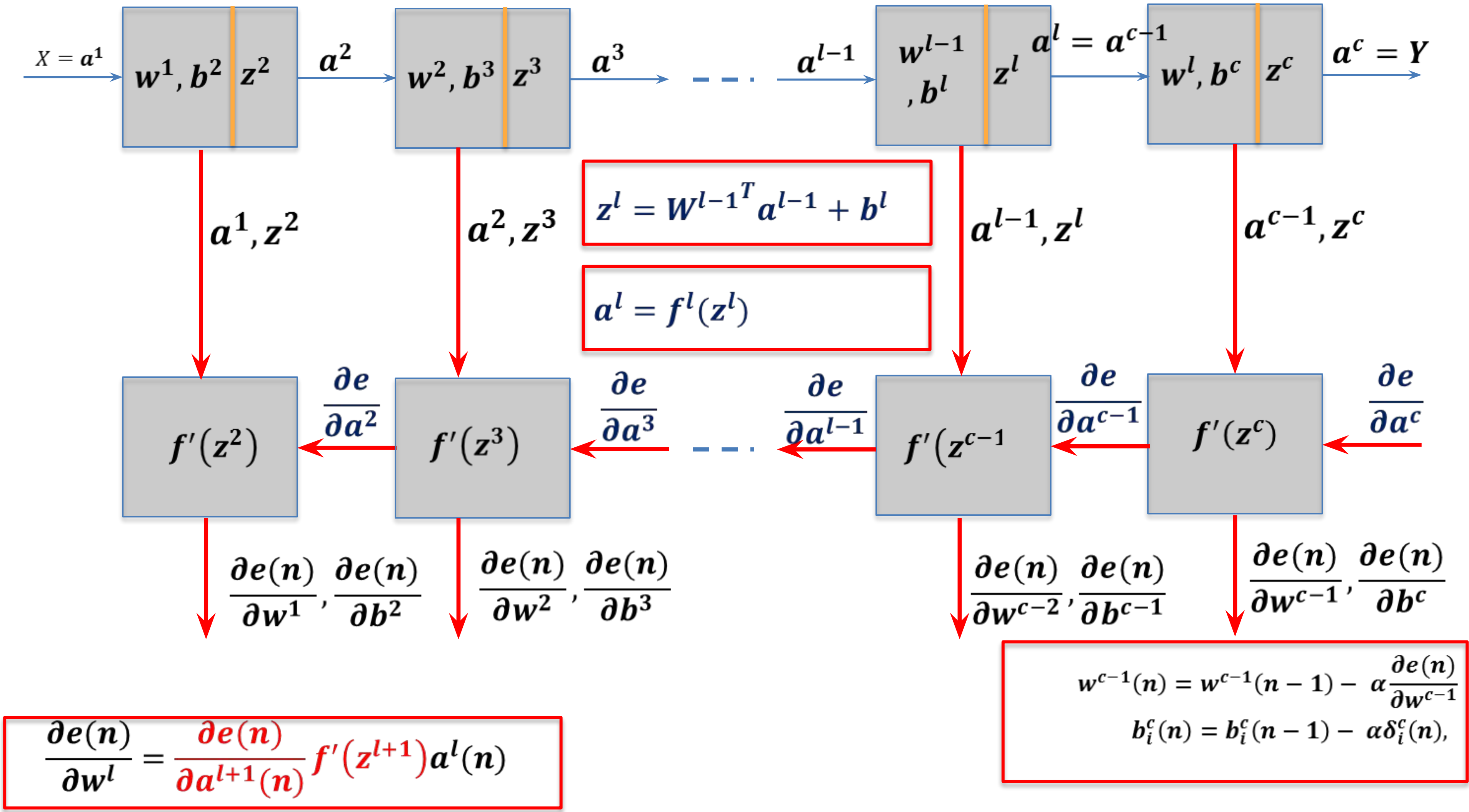
$$\boxed{\frac{\partial \boldsymbol{e}(\boldsymbol{n})}{\partial \boldsymbol{w}^{l}} = \frac{\partial \boldsymbol{e}(\boldsymbol{n})}{\partial \boldsymbol{a}^{l+1}(\boldsymbol{n})}\boldsymbol{f}'\left(\boldsymbol{z}^{l+1}\right)\boldsymbol{a}^{l}(\boldsymbol{n})} \longrightarrow \boxed{\frac{\partial \boldsymbol{e}(\boldsymbol{n})}{\partial \boldsymbol{a}^{l+1}(\boldsymbol{n})} = \frac{\partial \boldsymbol{e}(\boldsymbol{n})}{\partial \boldsymbol{a}^{l+2}(\boldsymbol{n})}\frac{\partial \boldsymbol{a}^{l+2}(\boldsymbol{n})}{\partial \boldsymbol{z}^{l+2}(\boldsymbol{n})}\frac{\partial \boldsymbol{z}^{l+2}(\boldsymbol{n})}{\partial \boldsymbol{a}^{l+1}(\boldsymbol{n})}}$$

FORWARD



Capa 1
Capa 2
Capa *l*
Capa C
X_1
X_2
X_3
a_1^1
a_2^1
a_3^1
a_1^2
a_2^2
a_3^2
a_2^l
a_2^l
a_3^l
a_1^c
a_2^c
w_{11}^1
w_{12}^1
w_{13}^1
w_{21}^1
w_{22}^1
w_{23}^1
w_{31}^1
w_{32}^1
w_{33}^1
b_1^2
b_2^2
b_3^2
b_1^l
b_2^l
b_3^l
b_1^c
b_2^c
Y_1
Y_2

FORWARD-BACKWARD



PREGUNTAS

UNIVERSIDAD
NACIONAL
DE COLOMBIA