Red Neuronal Multicapas

Una capa de neuronas representa un modelo de múltiples salidas de la forma:

$$y_j = f_j\left(w_{j0} + \sum_{l=1}^{n} w_{jl}\, x_l\right) \qquad (1)$$

donde el subíndice j representa las j=1, 2, .., S salidas que se pretenden modelar y l representa las l=1, 2,.., n entradas.

Cuando se desean incorporar varias capas, es necesario incorporar un índice que indique a que capa pertenecen los parámetros y las variables de cada neurona. Supóngase que se desea crear una red neuronal para modelar un sistema con 4 variables de entrada y 3 variables de salida, con tres capas. Una posible configuración puede ser:

Capa 1: Tres neuronas.
Capa 2: Dos neuronas.
Capa 3: Tres neuronas (Obligatorias, por el número de salidas del sistema).

La anterior configuración puede representarse por: 4 – 3 – 2 – 3, que se crea a partir de la forma:

$$n - S^1 - S^2 - \cdots - S^C \qquad (2)$$

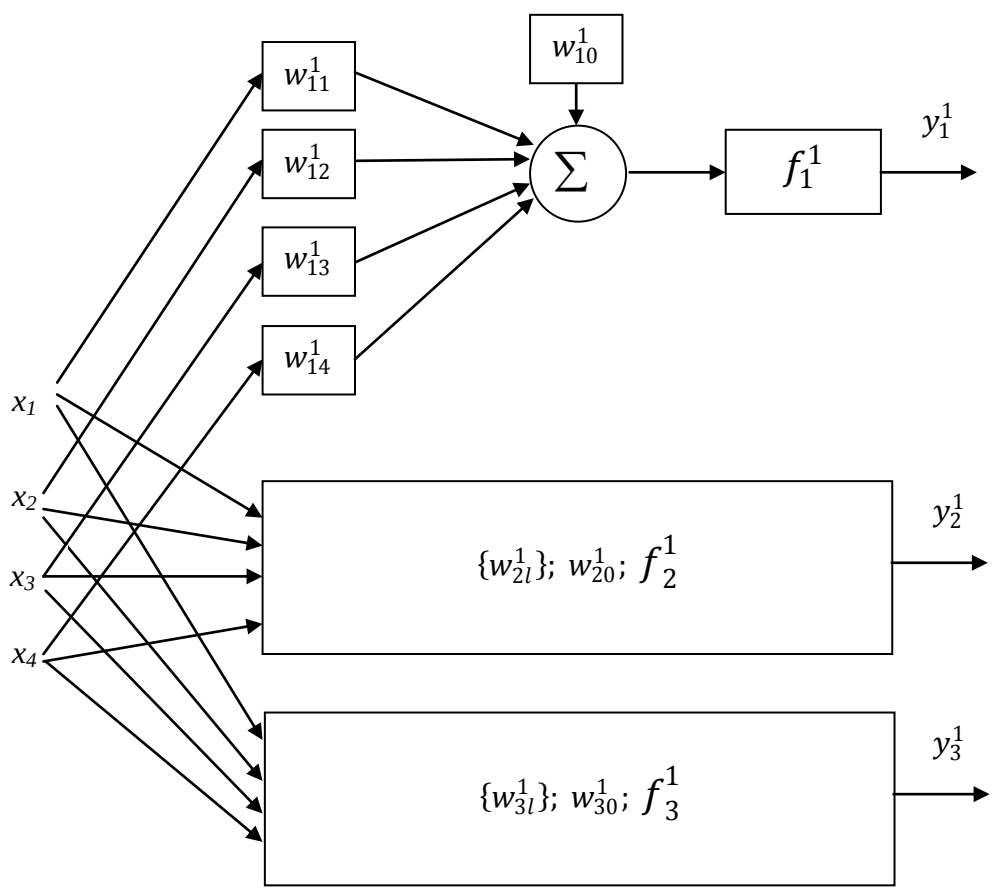
donde:
n: Número de variables de entrada.
S^c : Número de neuronas de la capa c.
C: Número de capas.

Para la configuración que se pretende crear, de acuerdo a (2), se cumple:

$$n = 4;\ S^1 = 3;\ S^2 = 2;\ S^3 = 3 \text{ ; siendo } C = 3 \qquad (3)$$

Las tres capas pueden representarse a través de las siguientes figuras,

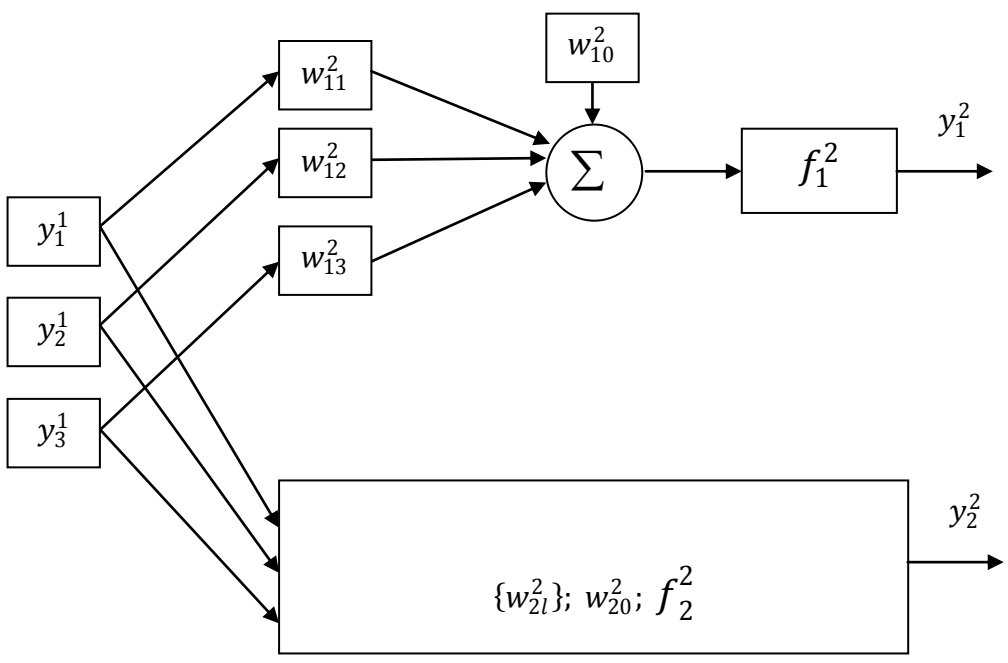
Capa 1:



Con la incorporación del supraíndice indicador de la capa, (1) se expresa por:

$$y_j^1 = f_j^1\left(w_{j0}^1 + \sum_{l=1}^{n=4} w_{jl}^1\, x_l\right) \quad \text{para} \quad j = 1, 2, S^1 = 3 \qquad (4)$$

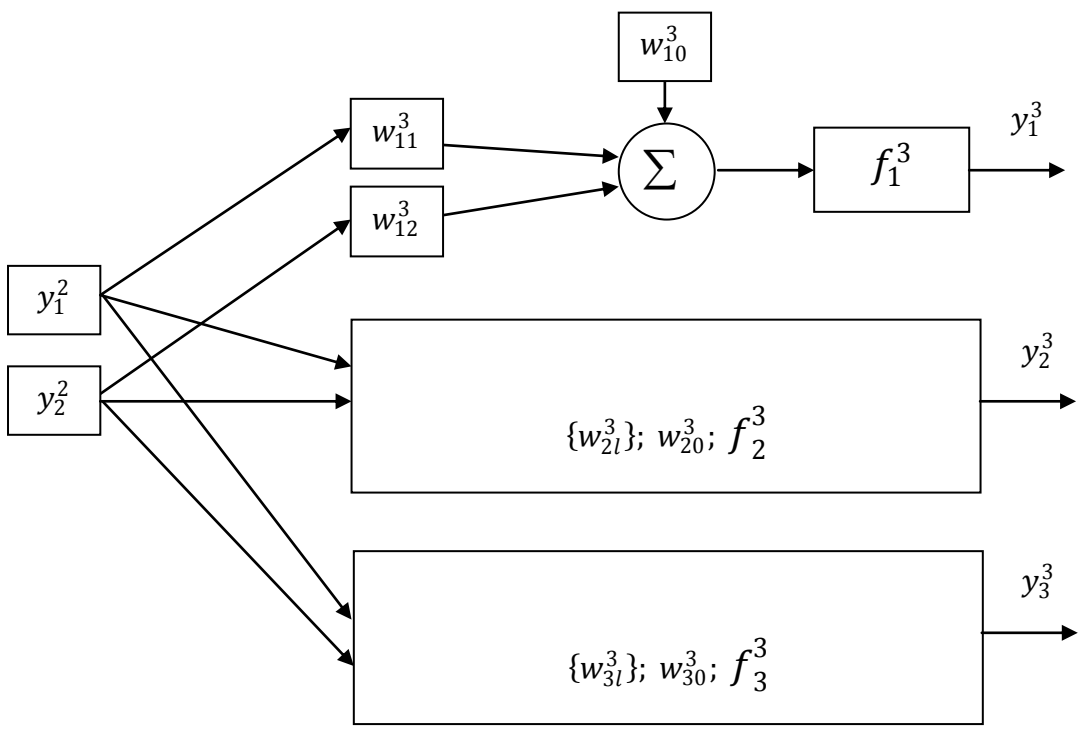
Capa 2:



La expresión (4) puede redefinirse para la capa 2 como:

$$y_j^2 = f_j^2\left(w_{j0}^2 + \sum_{l=1}^{S^1=3} w_{jl}^2\, y_l^1\right) \quad \text{para} \quad j = 1, S^2 = 2 \qquad (5)$$

Capa 3:



La expresión (5) puede redefinirse para la capa 3 como:

$$y_j^3 = f_j^3\left(w_{j0}^3 + \sum_{l=1}^{S^2=2} w_{jl}^3\, y_l^2\right) \qquad \text{para} \qquad j = 1,2, S^3 = 3 \qquad (6)$$

Si se analizan las expresiones (5) y (6) se puede crear una expresión general para las capas intermedias y la capa de salida (excepto la capa 1):

$$y_j^c = f_j^c\left(w_{j0}^c + \sum_{l=1}^{S^{c-1}} w_{jl}^c\, y_l^{c-1}\right) \quad \text{para} \quad j = 1,2,..,S^c \quad \text{y} \quad c = 2,..,C \qquad (7)$$

La ecuación (4) puede representarse de forma genérica por:

$$y_j^1 = f_j^1\left(w_{j0}^1 + \sum_{l=1}^{n} w_{jl}^1\, x_l\right) \qquad \text{para} \qquad j = 1,2,\dots,S^1 \qquad (8)$$

Las expresiones (7) y (8) representan una red neuronal multicapa.

Ejemplo: Salida de la función 1.

Para calcular la función de salida de una red es necesario evaluar las funciones de las capas previas, como ejemplo se determinará la salida 1 de la capa de salida (3), para ello:

Capa 1:
Basado en la expresión (8), como existen tres neuronas, $j = 1,2,3$ y cuatro variables de entrada (n=4):

$$y_1^1 = f_1^1(w_{10}^1 + w_{11}^1 x_1 + w_{12}^1 x_2 + w_{13}^1 x_3 + w_{14}^1 x_4) \qquad (9)$$

$$y_2^1 = f_2^1(w_{20}^1 + w_{21}^1 x_1 + w_{22}^1 x_2 + w_{23}^1 x_3 + w_{24}^1 x_4) \qquad (10)$$

$$y_3^1 = f_3^1(w_{30}^1 + w_{31}^1 x_1 + w_{32}^1 x_2 + w_{33}^1 x_3 + w_{34}^1 x_4) \qquad (11)$$

Capa 2 (c=2):
Basado en la expresión (7), como existen dos neuronas, $j = 1,2$. Existirán tres variables de entrada ($S^{c-1} = S^1 = 3$):

$$y_1^2 = f_1^2\begin{pmatrix} w_{10}^2 + w_{11}^2 f_1^1(w_{10}^1 + w_{11}^1 x_1 + w_{12}^1 x_2 + w_{13}^1 x_3 + w_{14}^1 x_4) \\ +w_{12}^2 f_2^1(w_{20}^1 + w_{21}^1 x_1 + w_{22}^1 x_2 + w_{23}^1 x_3 + w_{24}^1 x_4) \\ +w_{13}^2 f_3^1(w_{30}^1 + w_{31}^1 x_1 + w_{32}^1 x_2 + w_{33}^1 x_3 + w_{34}^1 x_4) \end{pmatrix} \qquad (12)$$

$$y_2^2 = f_2^2\begin{pmatrix} w_{20}^2 + w_{21}^2 f_1^1(w_{10}^1 + w_{11}^1 x_1 + w_{12}^1 x_2 + w_{13}^1 x_3 + w_{14}^1 x_4) \\ +w_{22}^2 f_2^1(w_{20}^1 + w_{21}^1 x_1 + w_{22}^1 x_2 + w_{23}^1 x_3 + w_{24}^1 x_4) \\ +w_{23}^2 f_3^1(w_{30}^1 + w_{31}^1 x_1 + w_{32}^1 x_2 + w_{33}^1 x_3 + w_{34}^1 x_4) \end{pmatrix} \qquad (13)$$

Capa 3 (c=3):
Basado en la expresión (7), como existen tres neuronas, $j = 1,2,3$ (en este caso se calcula sólo la salida de la primera función, por lo que $j = 1$). Existirán dos variables de entrada ($S^{c-1} = S^2 = 2$):

$$y_1^3 = f_1^3\begin{pmatrix} w_{10}^3 + w_{11}^3 f_1^2\begin{pmatrix} w_{10}^2 + w_{11}^2 f_1^1(w_{10}^1 + w_{11}^1 x_1 + w_{12}^1 x_2 + w_{13}^1 x_3 + w_{14}^1 x_4) \\ +w_{12}^2 f_2^1(w_{20}^1 + w_{21}^1 x_1 + w_{22}^1 x_2 + w_{23}^1 x_3 + w_{24}^1 x_4) \\ +w_{13}^2 f_3^1(w_{30}^1 + w_{31}^1 x_1 + w_{32}^1 x_2 + w_{33}^1 x_3 + w_{34}^1 x_4) \end{pmatrix} \\ +w_{12}^3 f_2^2\begin{pmatrix} w_{20}^2 + w_{21}^2 f_1^1(w_{10}^1 + w_{11}^1 x_1 + w_{12}^1 x_2 + w_{13}^1 x_3 + w_{14}^1 x_4) \\ +w_{22}^2 f_2^1(w_{20}^1 + w_{21}^1 x_1 + w_{22}^1 x_2 + w_{23}^1 x_3 + w_{24}^1 x_4) \\ +w_{23}^2 f_3^1(w_{30}^1 + w_{31}^1 x_1 + w_{32}^1 x_2 + w_{33}^1 x_3 + w_{34}^1 x_4) \end{pmatrix} \end{pmatrix} \qquad (14)$$

Las componentes derivativas:

Para calcular el gradiente, es necesario derivar con respecto a los parámetros, se deriva aquí con respecto a un parámetro de cada capa.

Parámetro 1: Corresponde a la capa 1, segunda variable de entrada, tercera neurona ($w_{32}^1 = w_{jl}^c$):

$$\frac{\partial y_1^3}{\partial w_{32}^1} = \frac{\partial f_1^3}{\partial w_{32}^1} = \frac{\partial f_1^3}{\partial f_1^3(.)}\left(w_{11}^3\left(\frac{\partial f_1^2}{\partial f_1^2(.)}\right) w_{13}^2\left(\frac{\partial f_3^1}{\partial f_3^1(.)}\right) x_2 + w_{12}^3\left(\frac{\partial f_2^2}{\partial f_2^2(.)}\right) w_{23}^2\left(\frac{\partial f_3^1}{\partial f_3^1(.)}\right) x_2\right)$$

Para generalizar se convierten a variables:

$$\frac{\partial y_{j3}^{c+2}}{\partial w_{jl}^c} = \frac{\partial f_{j3}^{c+2}}{\partial w_{jl}^c} = \frac{\partial f_{j3}^{c+2}}{\partial f_{j3}^{c+2}(.)}\left(w_{(j3)1}^{c+2}\left(\frac{\partial f_1^{c+1}}{\partial f_1^{c+1}(.)}\right) w_{1j}^{c+1} + w_{(j3)2}^{c+2}\left(\frac{\partial f_2^{c+1}}{\partial f_2^{c+1}(.)}\right) w_{2j}^{c+1}\right)\left(\left(\frac{\partial f_j^c}{\partial f_j^c(.)}\right) x_l\right)$$

de lo que resulta:

$$\frac{\partial y_{j3}^{c+2}}{\partial w_{jl}^c} = \frac{\partial f_{j3}^{c+2}}{\partial w_{jl}^c} = \frac{\partial f_{j3}^{c+2}}{\partial f_{j3}^{c+2}(.)}\left(\sum_{j2=1}^{S^2} w_{(j3)(j2)}^{c+2}\left(\frac{\partial f_{j2}^{c+1}}{\partial f_{j2}^{c+1}(.)}\right) w_{(j2)j}^{c+1}\right)\left(\left(\frac{\partial f_j^c}{\partial f_j^c(.)}\right) x_l\right) \qquad (15)$$

donde:

$j = 1,2,\dots,S^1 \equiv$ Número de neuronas de la capa 1.
$j_2 = 1,\ S^2 \equiv$ Número de neuronas de la capa 2.
$j_3 = 1,2,\dots,S^3 \equiv$ Número de neuronas de la capa 3 (coincide con el número de salidas).

Parámetro 2: Corresponde a la capa 2, primera variable de entrada (salida de la primera neurona de la capa 1), segunda neurona ($w_{21}^2 = w_{(j2)j}^{c+1}$):

$$\frac{\partial y_1^3}{\partial w_{21}^2} = \frac{\partial f_1^3}{\partial w_{21}^2} = \frac{\partial f_1^3}{\partial f_1^3(.)}\left(w_{12}^3\left(\frac{\partial f_2^2}{\partial f_2^2(.)}\right) f_1^1(w_{10}^1 + w_{11}^1 x_1 + w_{12}^1 x_2 + w_{13}^1 x_3 + w_{14}^1 x_4)\right)$$

Si se generaliza:

$$\frac{\partial y_{j3}^{c+2}}{\partial w_{(j2)j}^{c+1}} = \frac{\partial f_{j3}^{c+2}}{\partial w_{(j2)j}^{c+1}} = \frac{\partial f_{j3}^{c+2}}{\partial f_{j3}^{c+2}(.)}\left(w_{j(j2)}^{c+2}\left(\frac{\partial f_{j2}^{c+1}}{\partial f_{j2}^{c+1}(.)}\right) f_j^c(.)\right) \qquad (16)$$

Parámetro 3: Corresponde a la capa 3, segunda variable de entrada (segunda neurona de la capa 2), primera neurona ($w_{12}^3 = w_{(j3)(j2)}^{c+2}$):

$$\frac{\partial y_1^3}{\partial w_{12}^3} = \frac{\partial f_1^3}{\partial w_{12}^3} = \frac{\partial f_1^3}{\partial f_1^3(.)}\left(f_2^2\begin{pmatrix} w_{20}^2 + w_{21}^2 f_1^1(w_{10}^1 + w_{11}^1 x_1 + w_{12}^1 x_2 + w_{13}^1 x_3 + w_{14}^1 x_4) \\ +w_{22}^2 f_2^1(w_{20}^1 + w_{21}^1 x_1 + w_{22}^1 x_2 + w_{23}^1 x_3 + w_{24}^1 x_4) \\ +w_{23}^2 f_3^1(w_{30}^1 + w_{31}^1 x_1 + w_{32}^1 x_2 + w_{33}^1 x_3 + w_{34}^1 x_4) \end{pmatrix}\right)$$

Si se generaliza:

$$\frac{\partial y_{j3}^{c+2}}{\partial w_{(j3)(j2)}^{c+2}} = \frac{\partial f_{j3}^{c+2}}{\partial w_{(j3)(j2)}^{c+2}} = \frac{\partial f_{j3}^{c+2}}{\partial f_{j3}^{c+2}(.)}\left(f_{j2}^{c+1}(.)\right) \qquad (17)$$

Pasos para crear una red neuronal multicapa:

1.- Se definen el número de variables de entrada (n) y variables de salida (S^C) del sistema.
2.- Se definen el número de capas (C).
3.- Se definen el número de neuronas de cada capa S^c, para $c = 1,2,..,C-1$ (excepto la de salida, que ha sido definida en el paso 1).
4.- Se definen las funciones de activación para cada capa $f_1^c, f_2^c, .., f_{S^c}^c$ (la única condición, en teoría, para definir estas funciones, es que sean derivables).