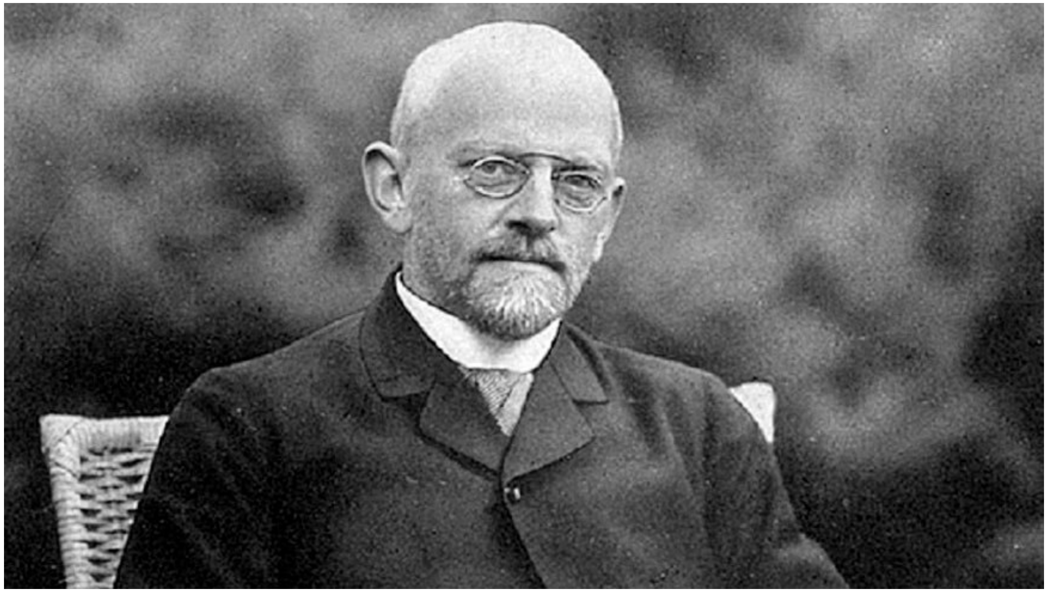
Hilbert, Kolmogorov (y Arnold) y las redes neuronales



París, 1900

Es posible que las funciones de 7º grado no puedan representarse por una composición finita de funciones de dos argumentos.

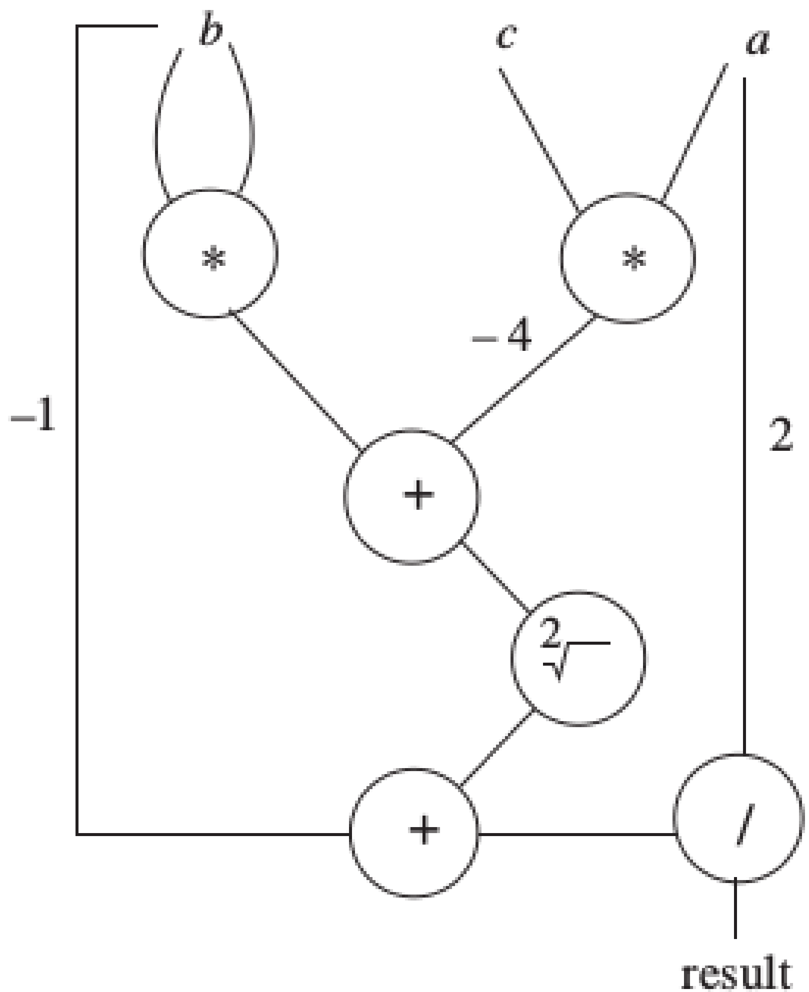
Ósea,

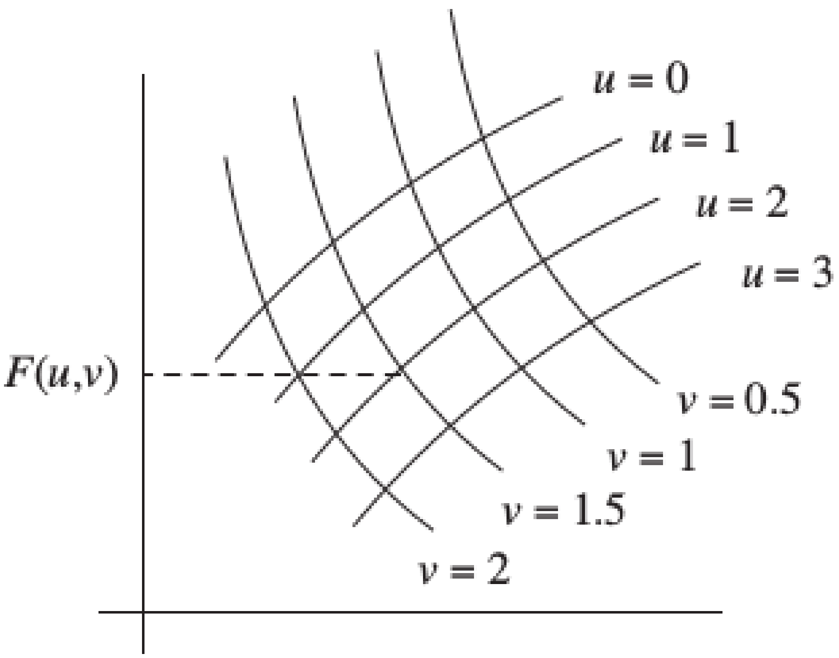
$$f^7 + xf^3 + yf^2 + zf + 1 = 0$$

No puede resolverse usando funciones de dos argumentos.

(Problema 13)

Rojas R, 1994. - Cap 10



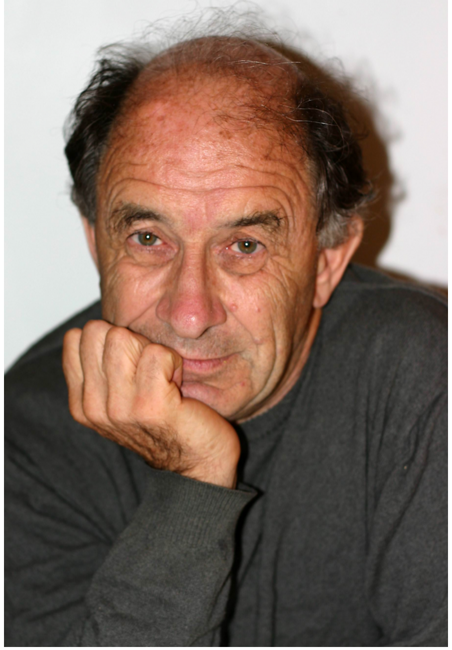


Para $x^2 + bx + c = 0$ se soluciona algebraicamente

Para la quíntica se ocupa el método nomográfico en F(u,v)

http://myreckonings.com/wordpress/2008/01/09/the-art-of-nomography-i-geometric-design/

Rojas R, 1994. - Cap 10



Rojas R, 1994. - Cap 10

Kolmogorov (y Arnold) en 1957

Una función continua de n argmentos puede *representarse* usando únicamente usando una composición finita de funciones y la suma.

De otra forma, la suma es la única función de más de un argumento requerida, junto con funciones de un solo argumento, para representar otras operaciones:

$$xy = \exp(\ln(x) + \ln(y))$$

Proposición

Sea $f:[0,1]^n \rightarrow [0,1]$ *función continua. Existen funciones* g *y* $\varphi_q\ ; q=1,\ldots,2n+1$ *de un argumento y constantes* $\lambda_p\ ;\ p=1,\ldots,n$ *tales que*

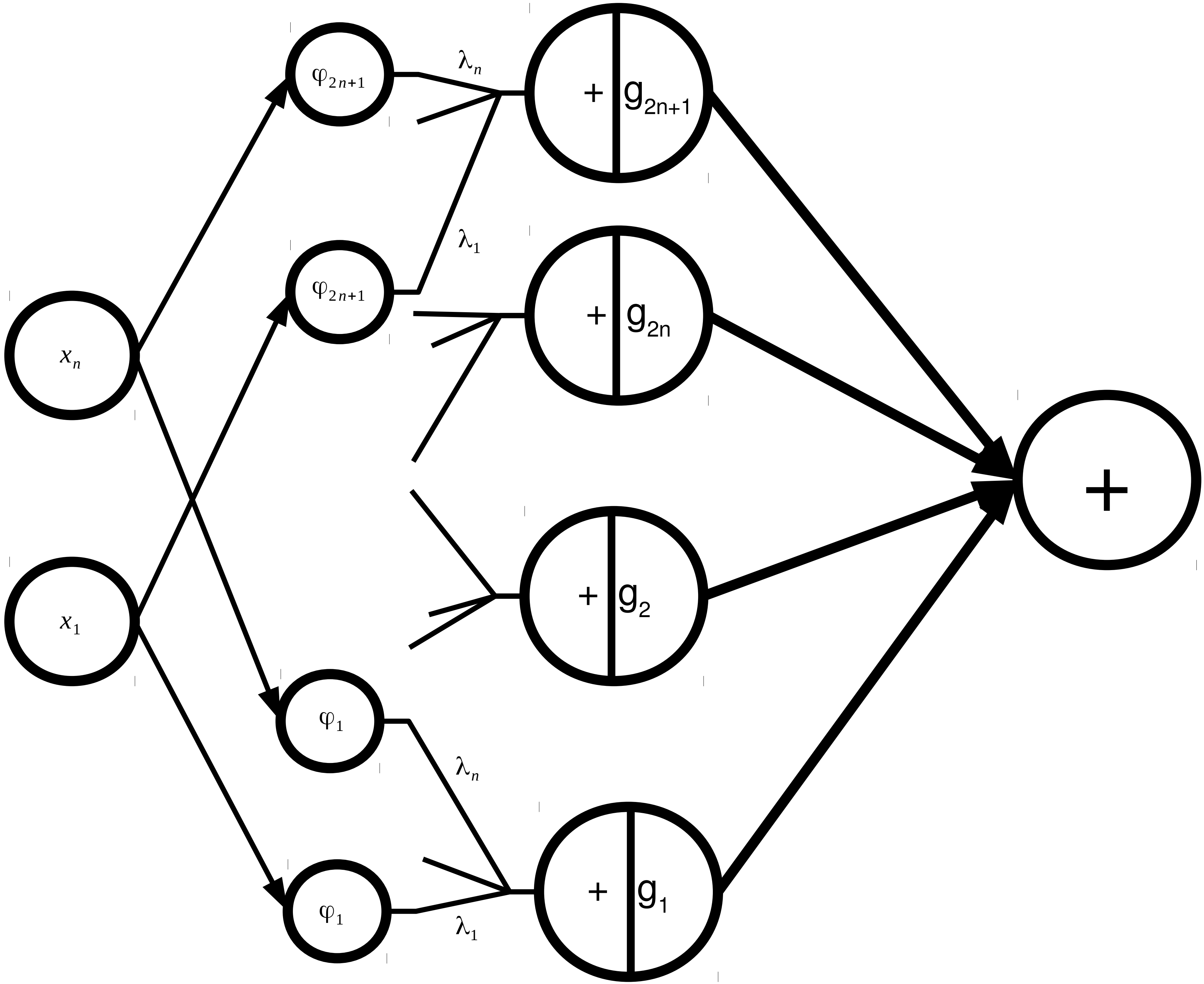
$$f(x_1,\ldots,x_n)=\sum_{q=1}^{2n+1} g\left(\sum_{p=1}^{n}\left(\lambda_p \varphi_q(x_p)\right)\right)$$

En el contexto de las redes neuronales podemos distinguir dos capas, la última:

$$y=\sum_{q}^{2n+1} g(z_q)$$

Y la capa anterior está computando

$$z_q=\sum_{p}^{n} \lambda_p \varphi_q(x_p)$$



x_n
x_1
φ_{2n+1}
φ_{2n+1}
φ_1
φ_1
λ_n
λ_1
λ_n
λ_1
+ g_{2n+1}
+ g_{2n}
+ g_2
+ g_1
+

Aproximar una función

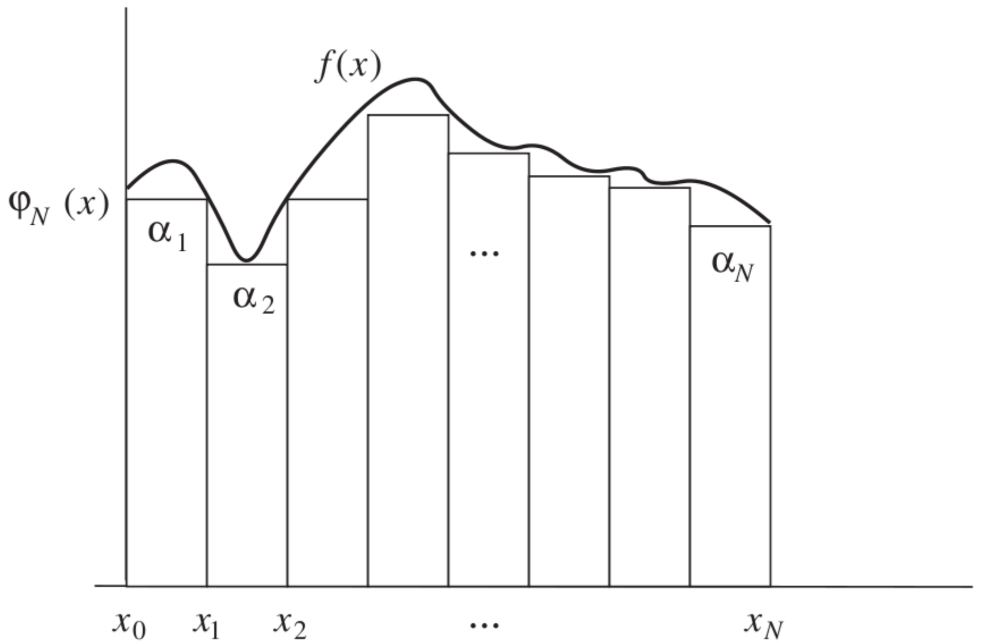
$$E = \int \left| f(x) - \hat{f}(x) \right| < \epsilon$$

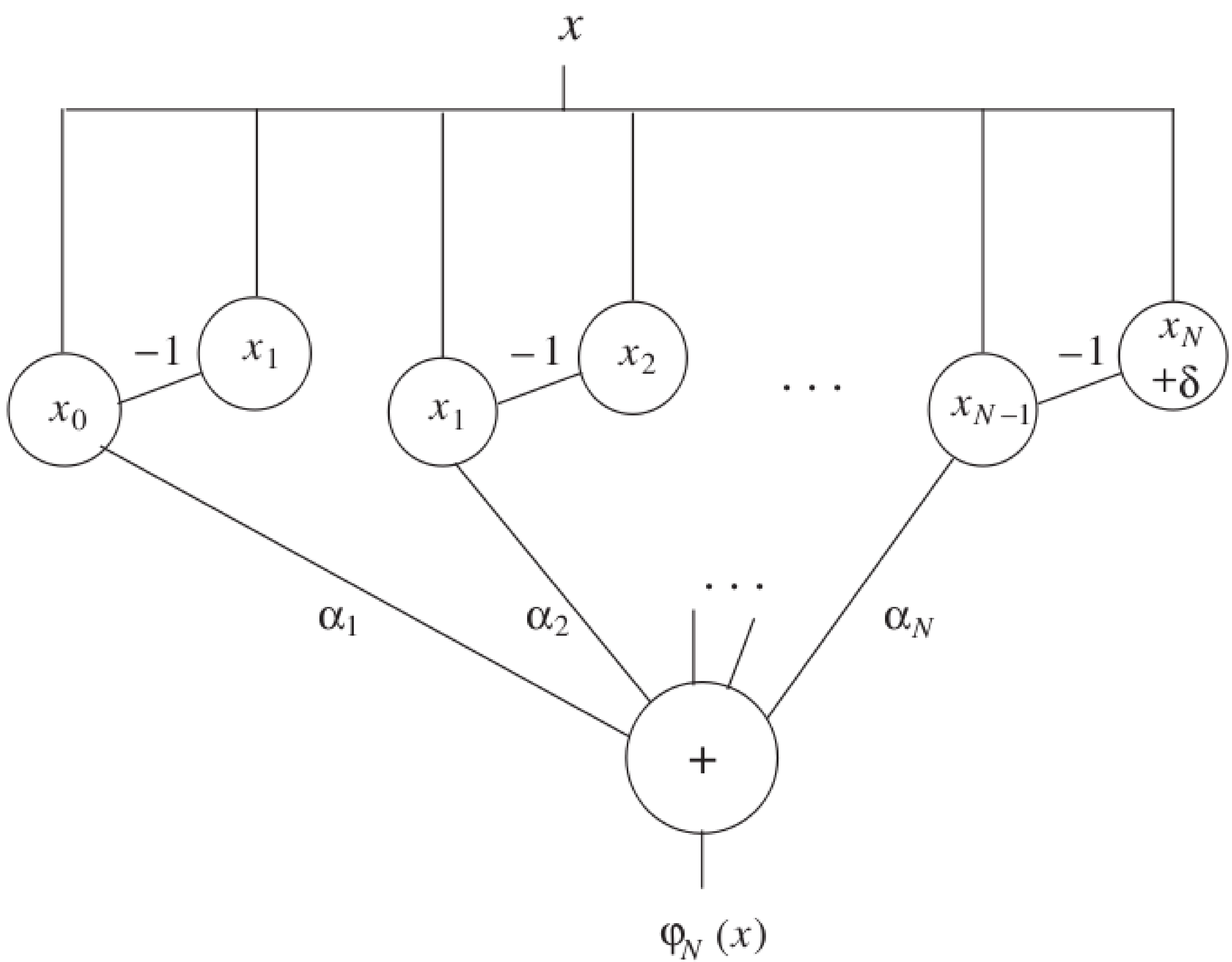
$$\varphi_N(x) = min\{f(x'); x \in [x_i, x_{i+1}]\}$$

$$E_N = \int \left| f(x) - \varphi_N(x) \right| < \epsilon$$

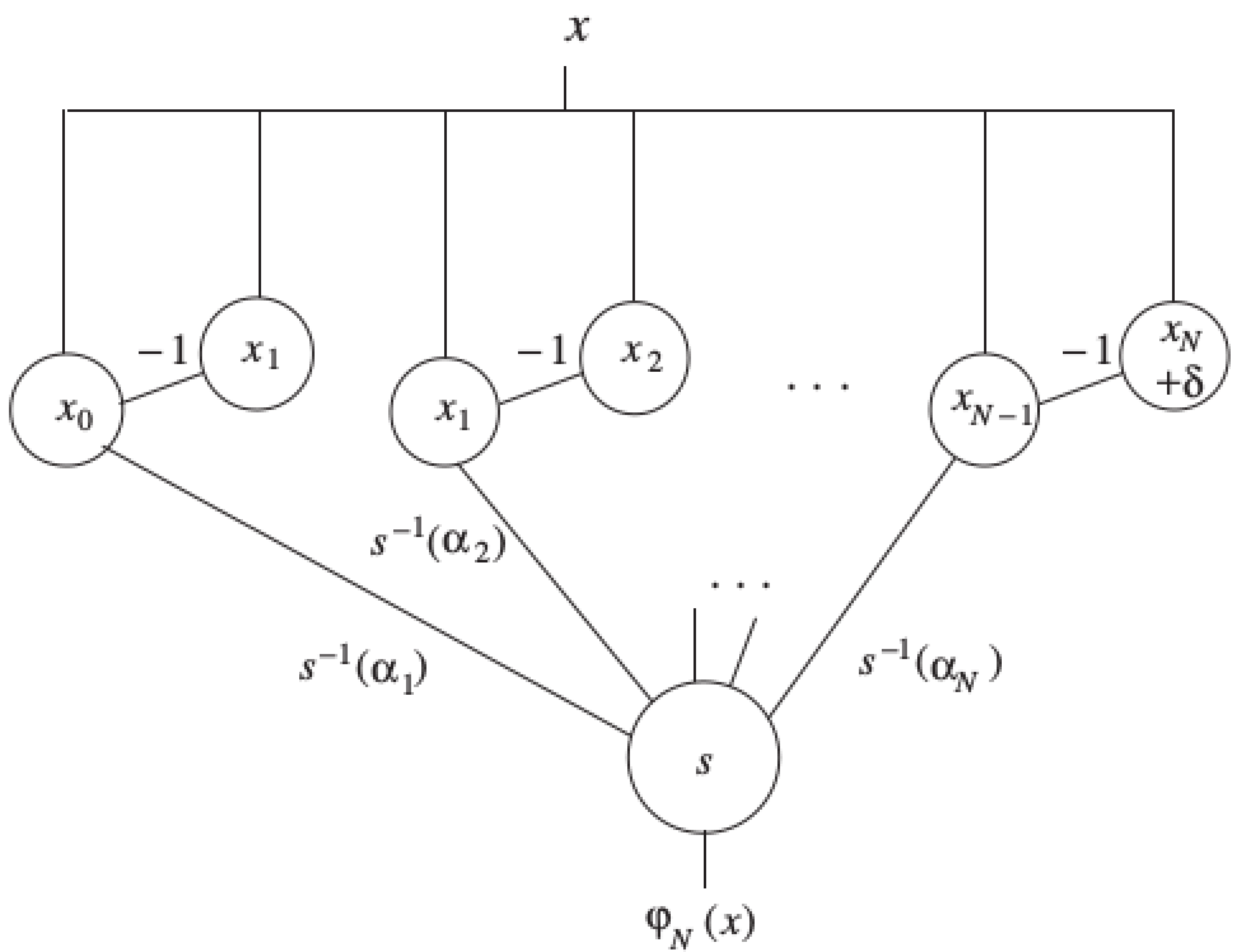
$$como\ f(x) > \varphi_N(x)\ \forall\ x$$

$$E_N = \int f(x) - \int \varphi_N(x) < \epsilon$$

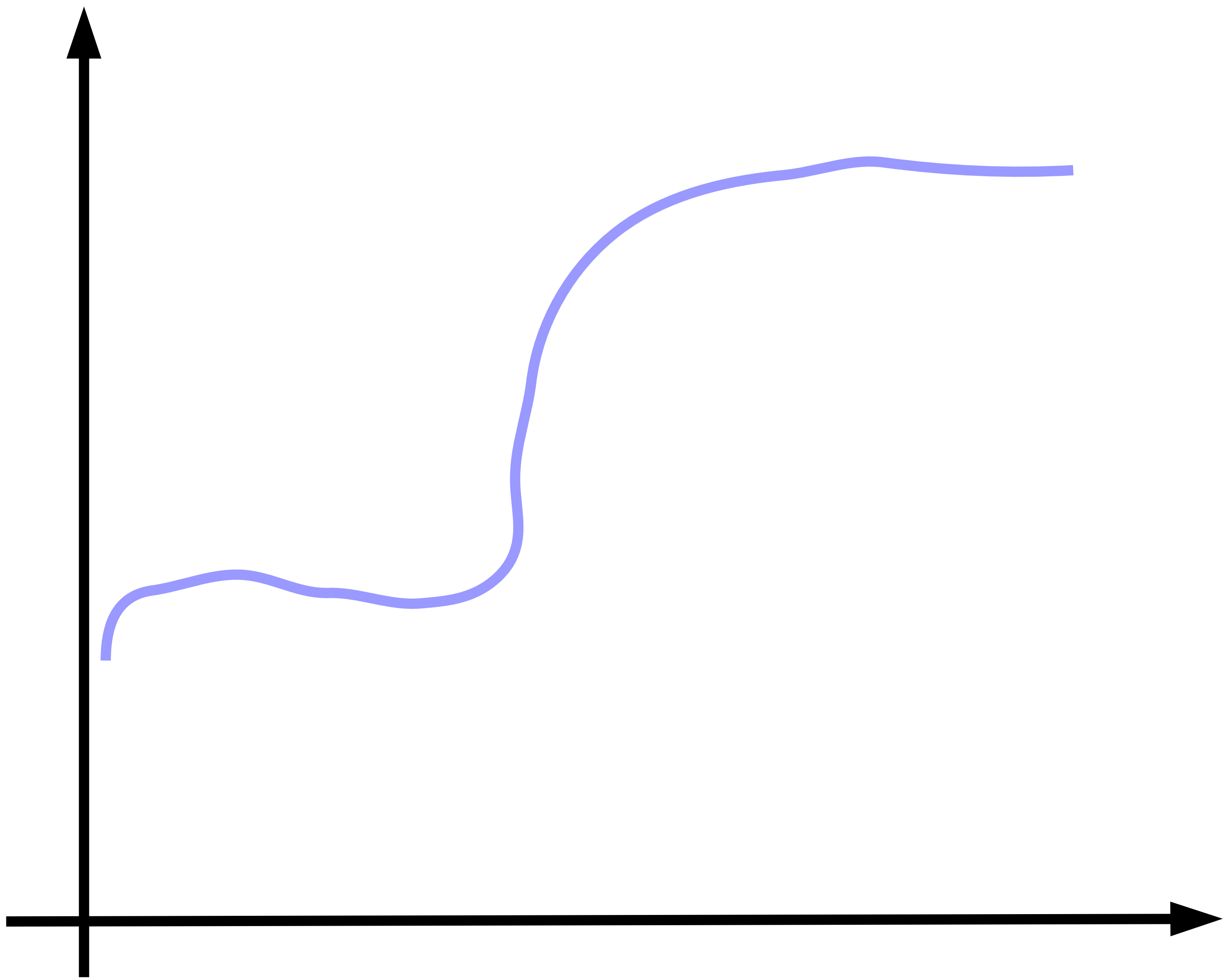


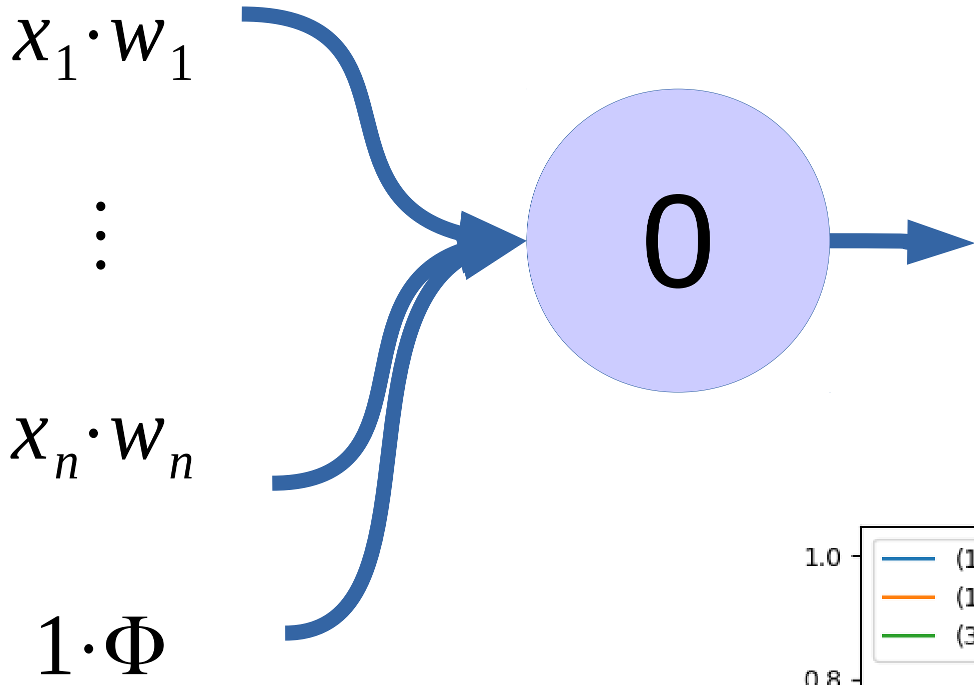


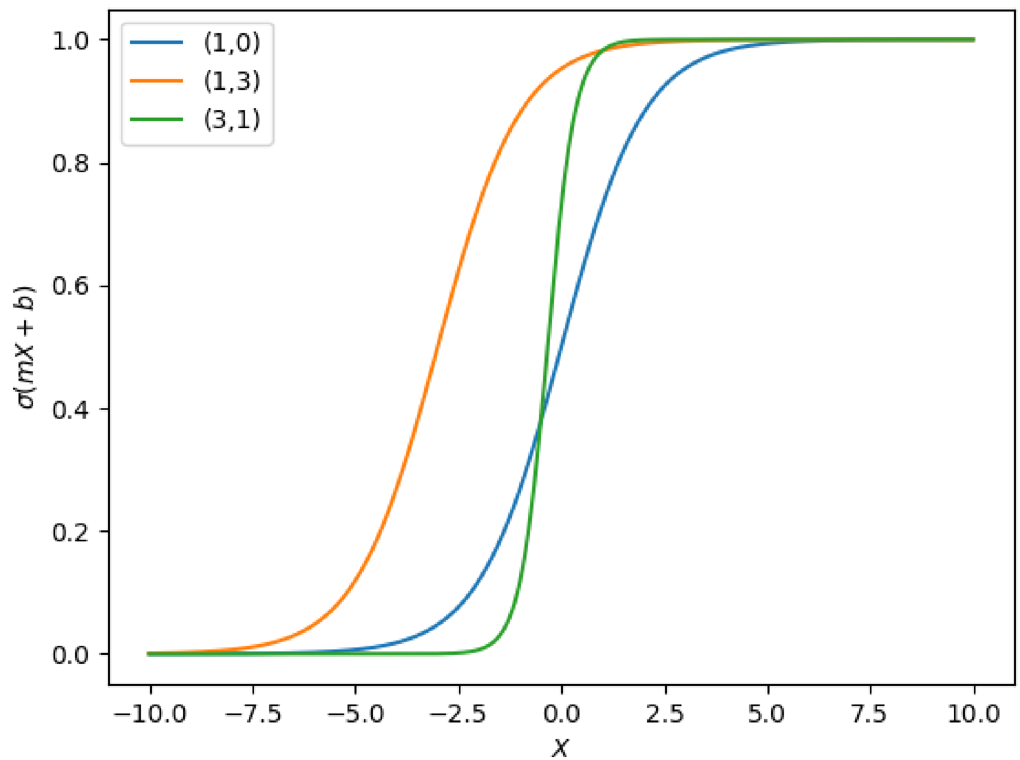
x
−1
−1
−1
x_0
x_1
x_1
x_2
x_{N-1}
x_N
$+\delta$
. . .
. . .
α_1
α_2
α_N
+
$\varphi_N(x)$



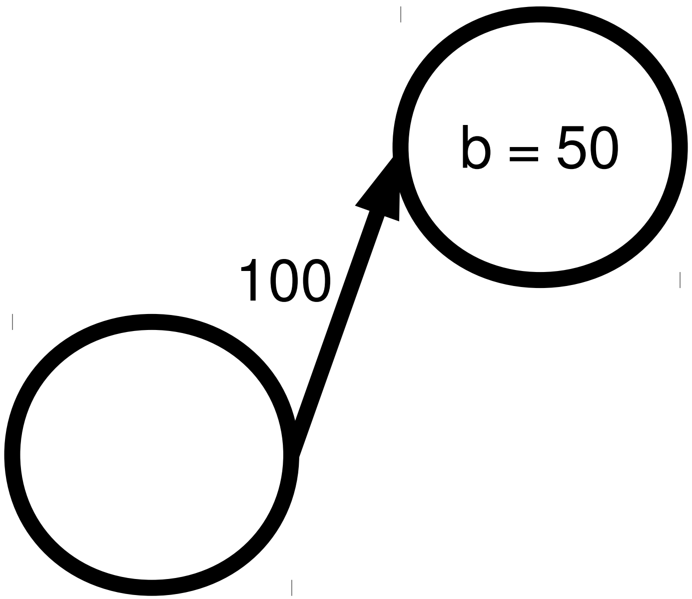
x
-1
x_1
x_0
-1
x_2
x_1
$\cdots$
x_{N-1}
-1
x_N $+\delta$
$s^{-1}(\alpha_2)$
$s^{-1}(\alpha_1)$
$s^{-1}(\alpha_N)$
$\cdots$
s
$\varphi_N(x)$

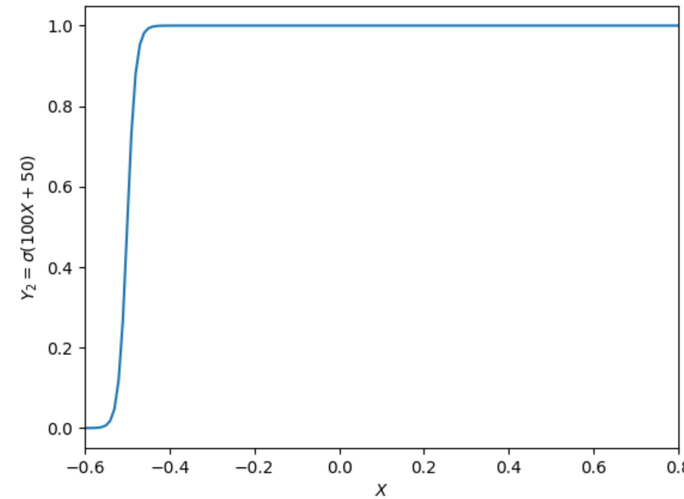


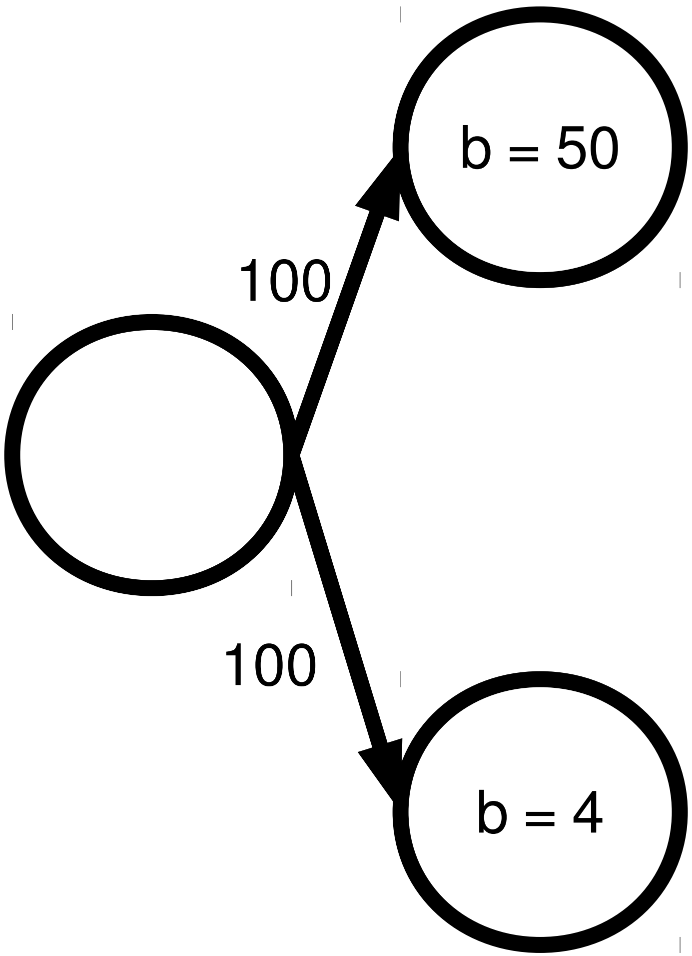


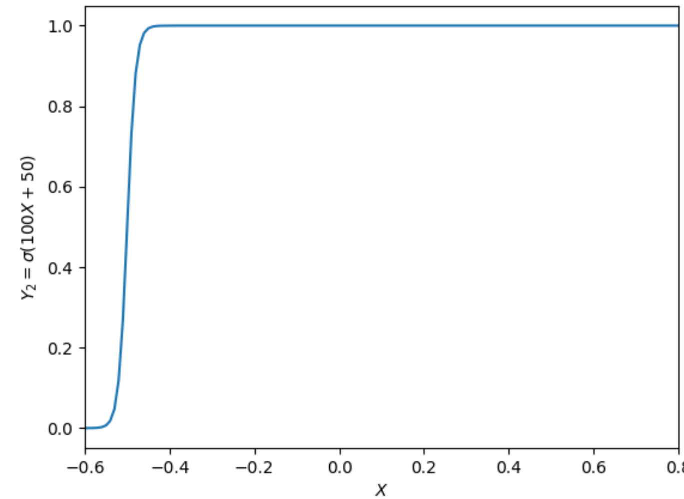


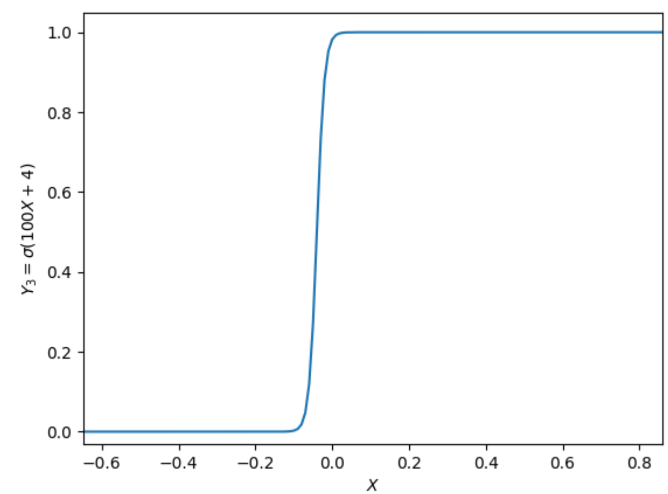
$$\sigma(W x) = \frac{1}{1+\exp(-W \cdot x)}$$

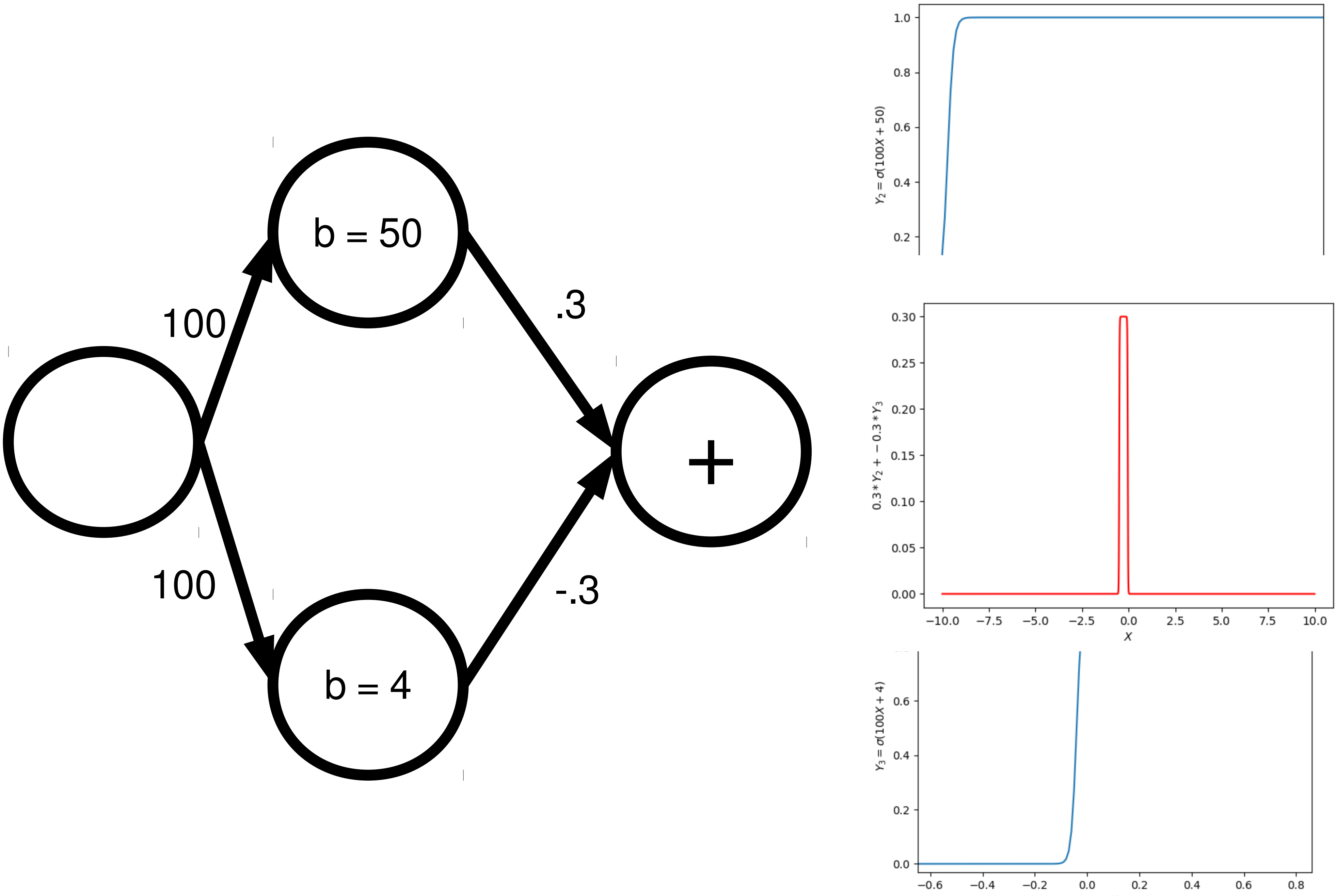

b = 50
100

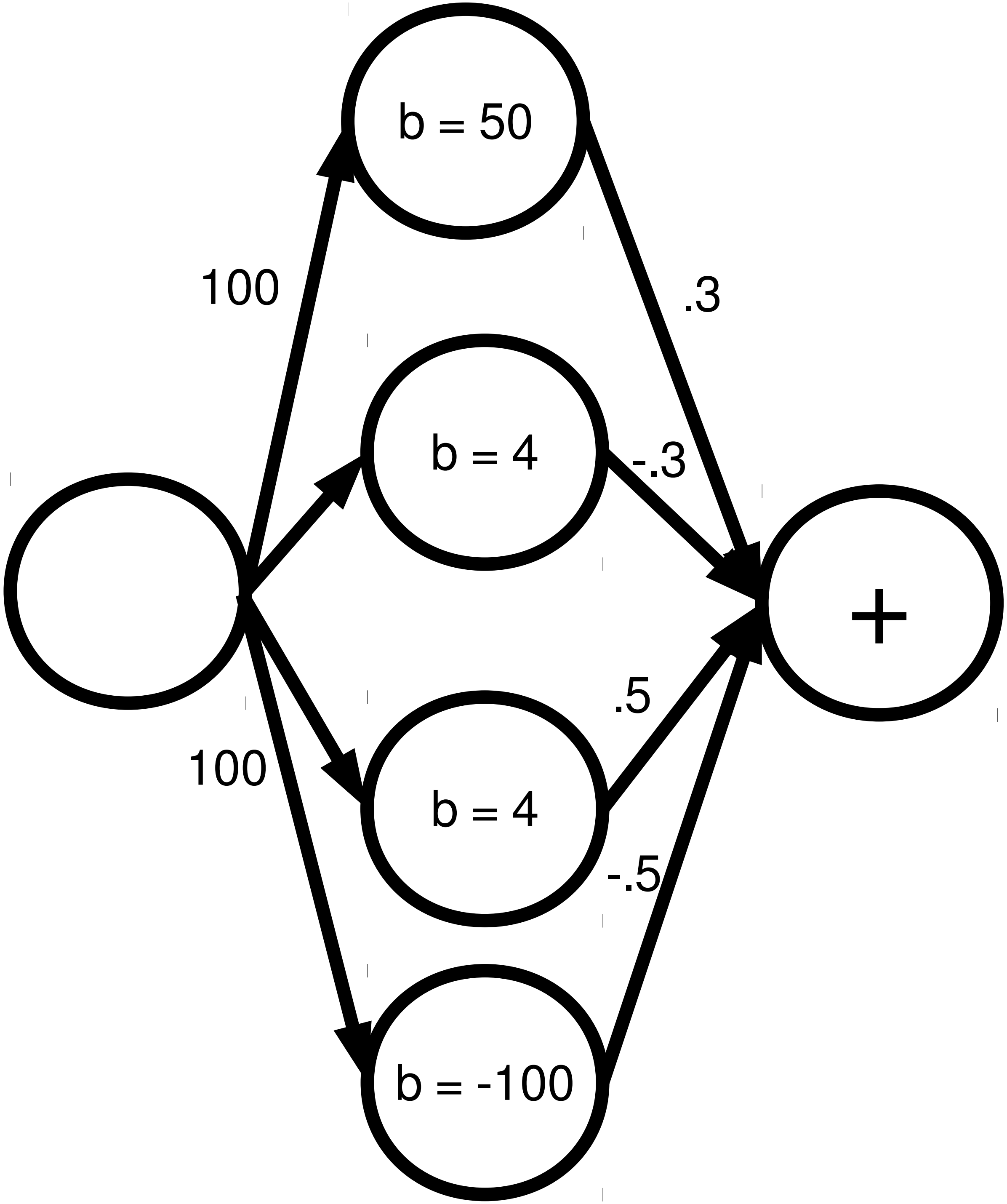

$Y_2 = \sigma(100X + 50)$
X
1.0
0.8
0.6
0.4
0.2
0.0
−0.6
−0.4
−0.2
0.0
0.2
0.4
0.6
0.8

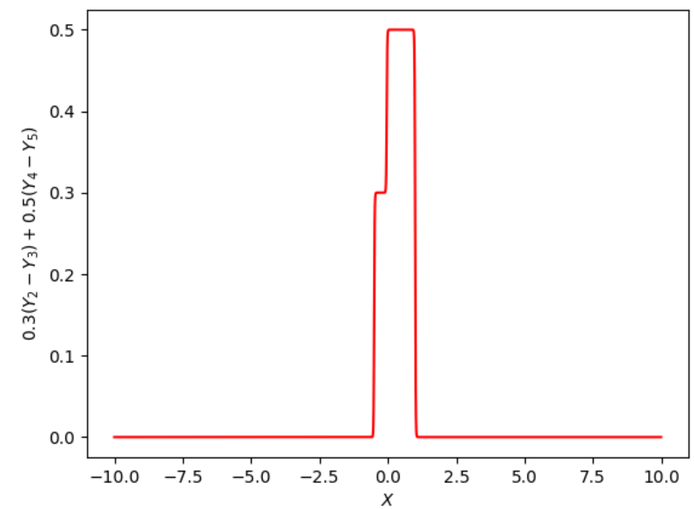

b = 50
100
100
b = 4

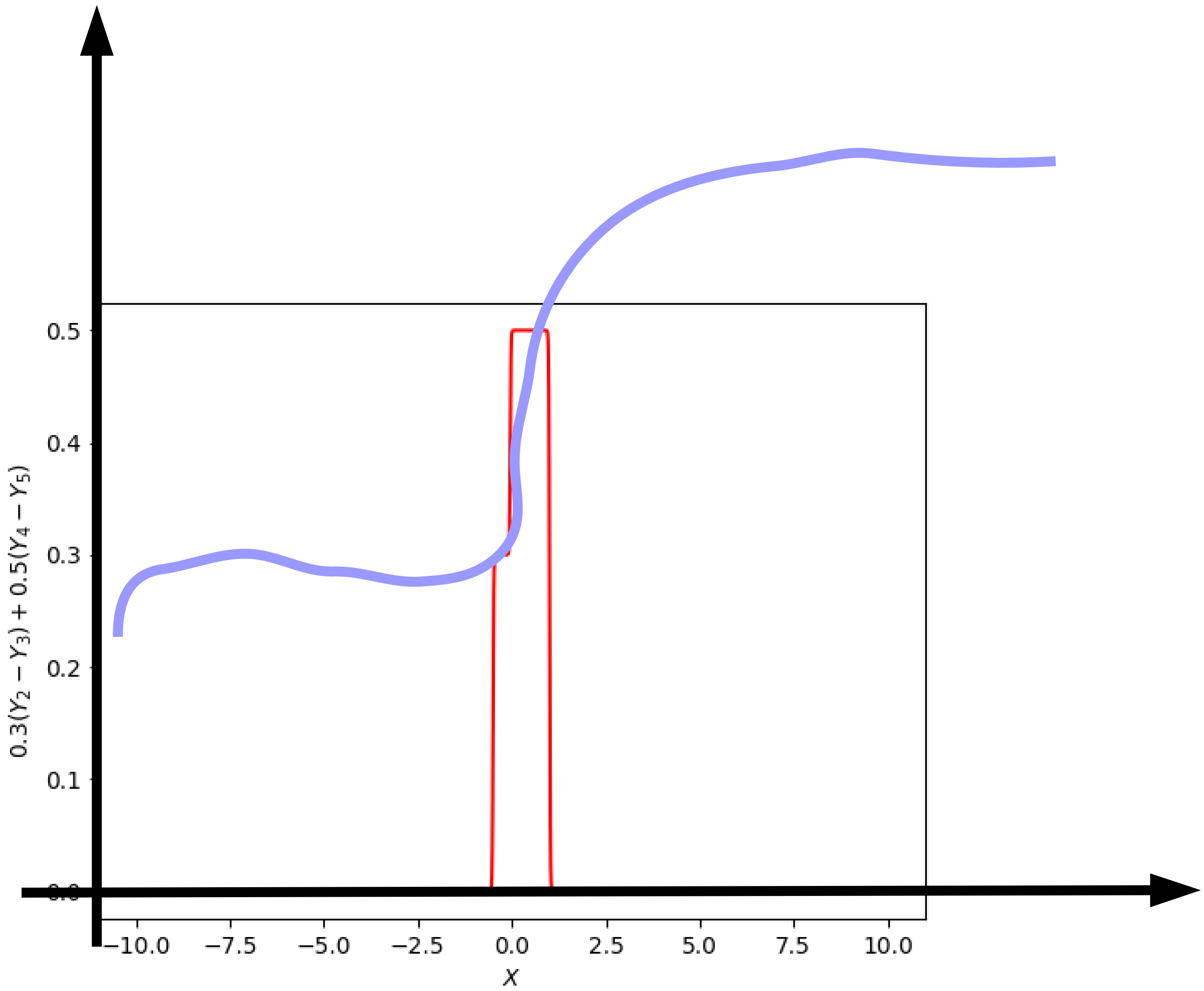

$Y_2 = \sigma(100X + 50)$
X


$Y_3 = \sigma(100X + 4)$
X


b = 50
b = 4
100
100
.3
-.3
+
$Y_2 = \sigma(100X + 50)$
$0.3 * Y_2 + -0.3 * Y_3$
X
$Y_3 = \sigma(100X + 4)$
X


b = 50
b = 4
b = 4
b = -100
100
100
.3
-.3
.5
-.5
+


0.3(Y2 − Y3) + 0.5(Y4 − Y5)
x
0.0
0.1
0.2
0.3
0.4
0.5
−10.0
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5
10.0



0.5
0.4
0.3
0.2
0.1
0.0
−10.0
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5
10.0
X
$0.3(Y_2 - Y_3) + 0.5(Y_4 - Y_5)$

Caso Multidimensional