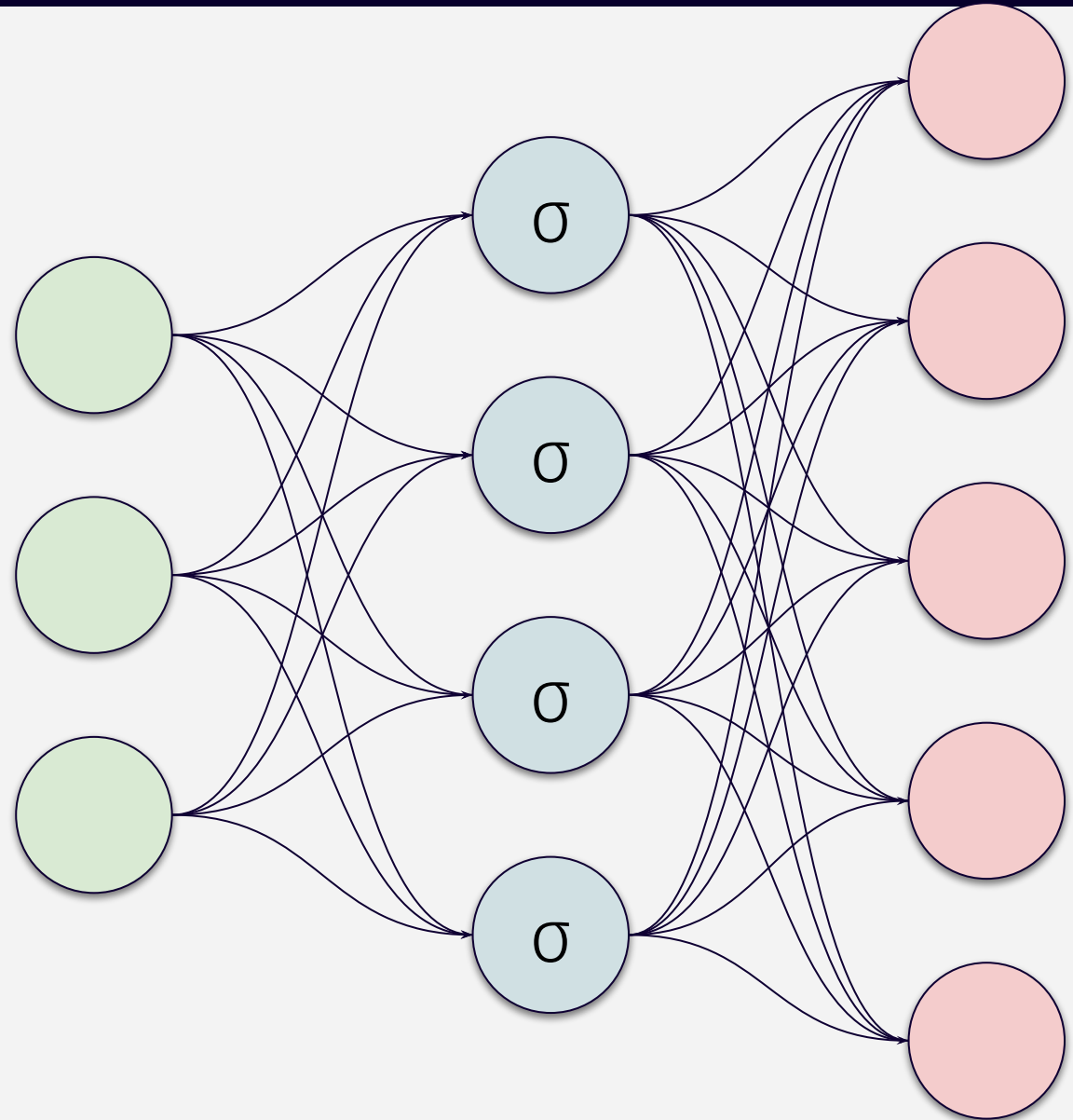


No linealidad en Redes
Neuronales.
Aproximación Universal

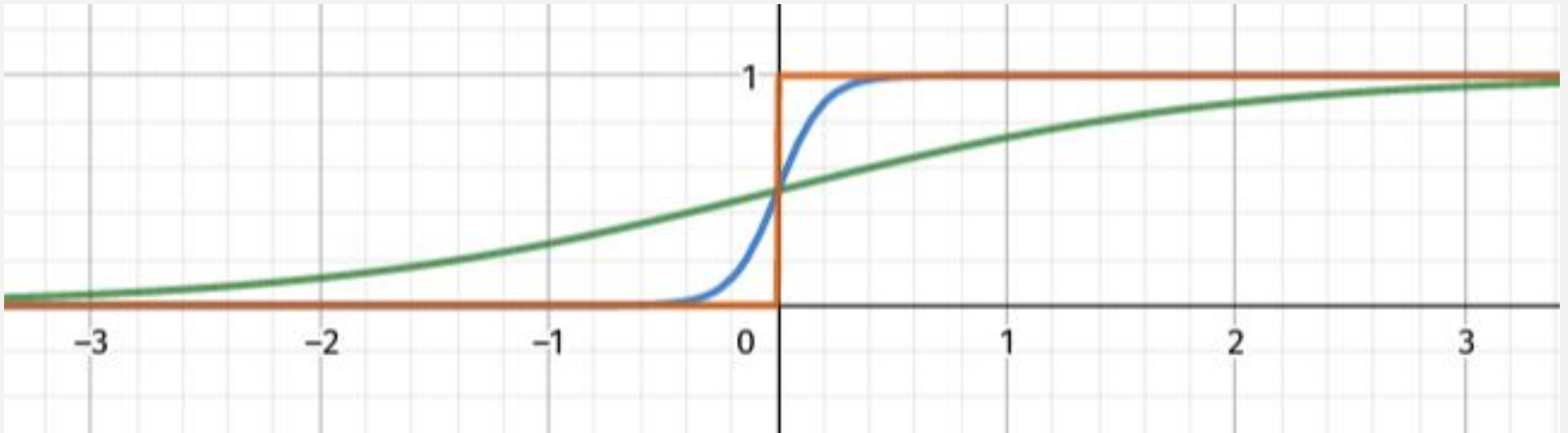
Teorema(s) de Cybenko

- Datos
 - Red
 - 2 capas
 - Fully connected
 - Act. sigmoidea (σ)
 - Conj. de datos
 - Función de error
- Aumentando #parámetros
 - Decrementa error
 - de forma arbitraria
- Teoremas
 - Para otras f. activación



Función sigmoidea

- Sigmoidea
 - $\sigma(x) = 1 / (1+e^{-x})$
 - $\sigma(x) = 1 / (1+e^{-(wx+b)})$
- Variando w y b
 - Cambia salida
- w varia pendiente
 - $\sigma(x)$
 - $\sigma(10x)$
 - $\sigma(1000x)$ (función *step*)



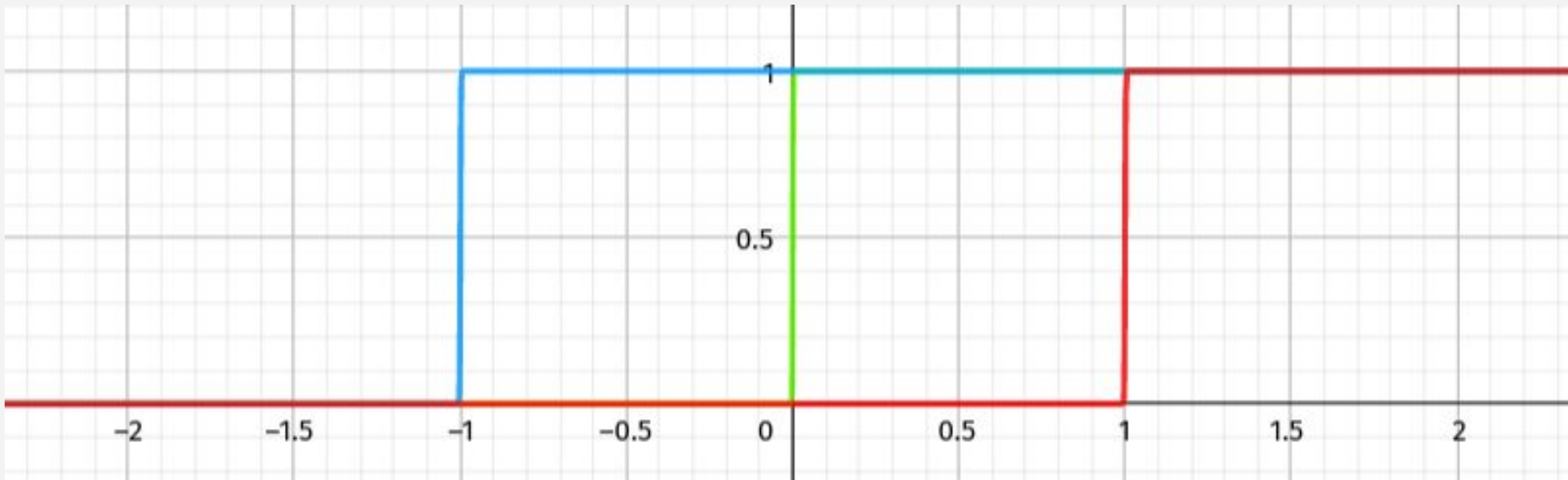
Función sigmoidea

- Sigmoidea
 - $\sigma(x) = 1 / (1+e^{-x})$
 - $\sigma(x) = 1 / (1+e^{-(wx+b)})$
- Variando w y b
 - Cambia salida
- b varia posición
 - $\sigma(x)$
 - $\sigma(x+10)$
 - $\sigma(x-10)$



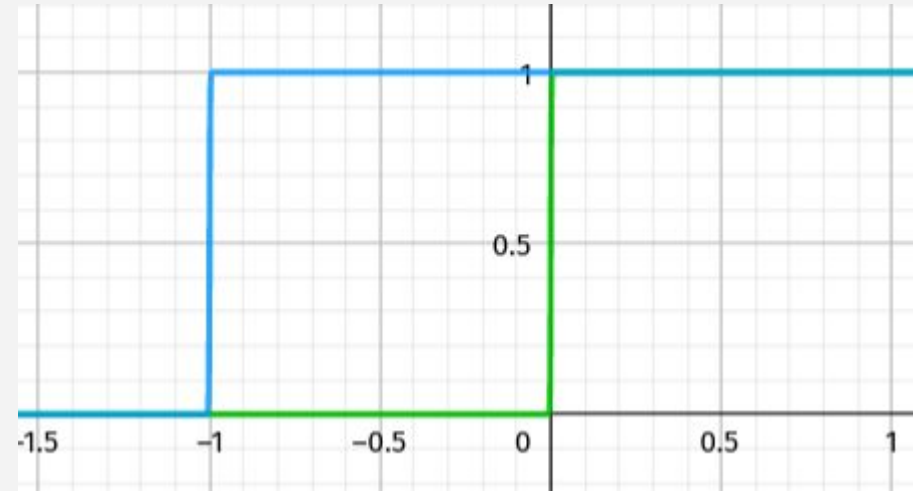
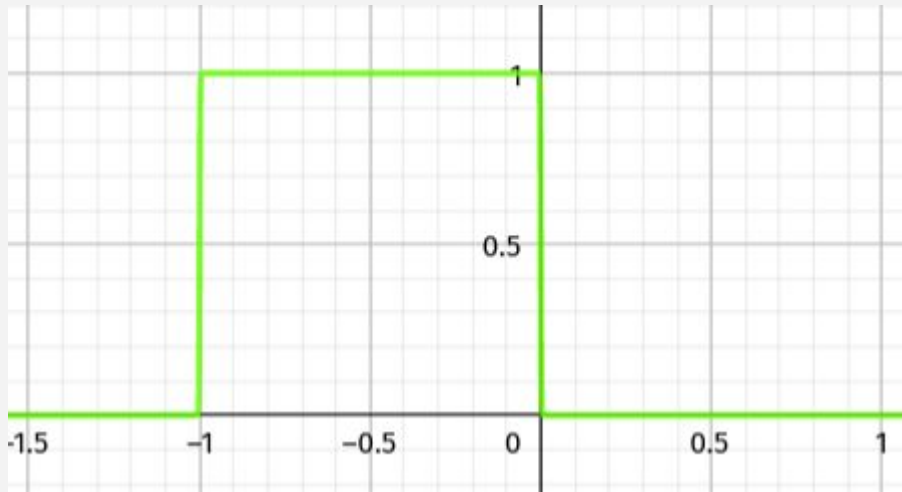
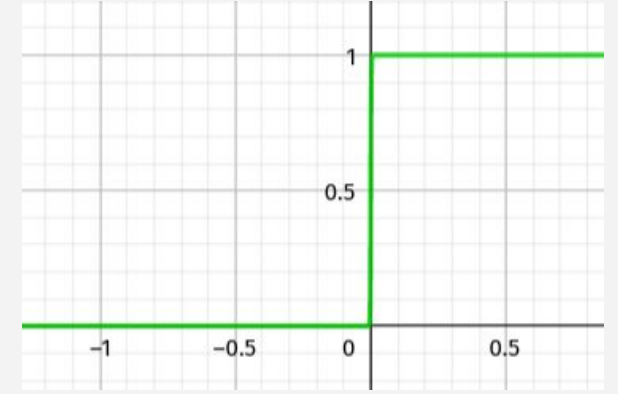
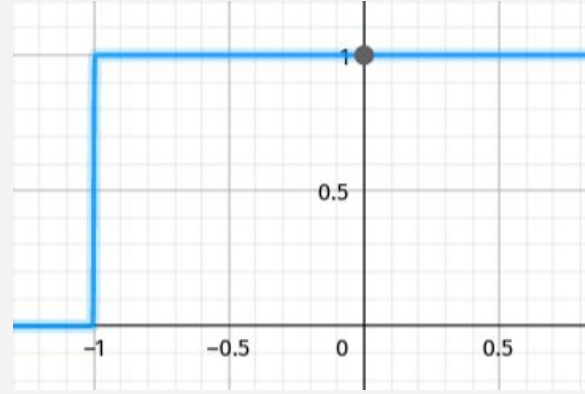
Función sigmoidea

- Sigmoidea
 - $\sigma(x) = 1 / (1+e^{-x})$
 - $\sigma(x) = 1 / (1+e^{-(wx+b)})$
- Variando w y b
 - Cambia salida
- w y b combinadas
 - $\sigma(1000x)$
 - $\sigma(1000x+1000)$
 - $\sigma(1000x-1000)$



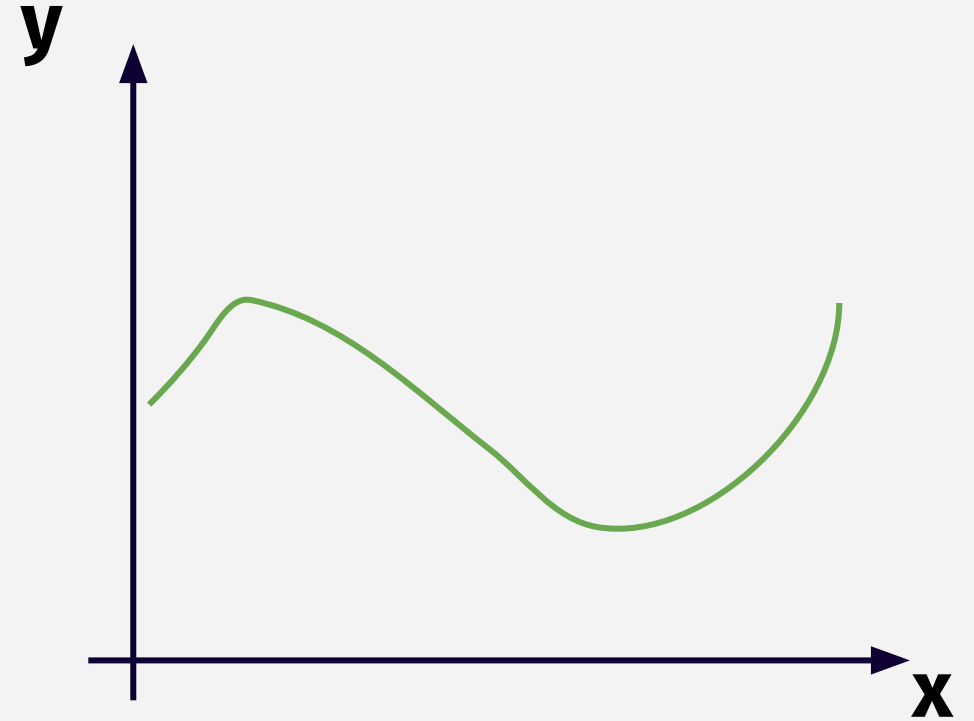
Truco: combinación de sigmoideas

- Sigmoidea
 - $\sigma(x) = 1 / (1 + e^{-x})$
- Selector de región
 - $\sigma(1000x)$
 - $\sigma(1000x+1000)$
 - $\sigma(1000x+1000) - \sigma(1000x)$



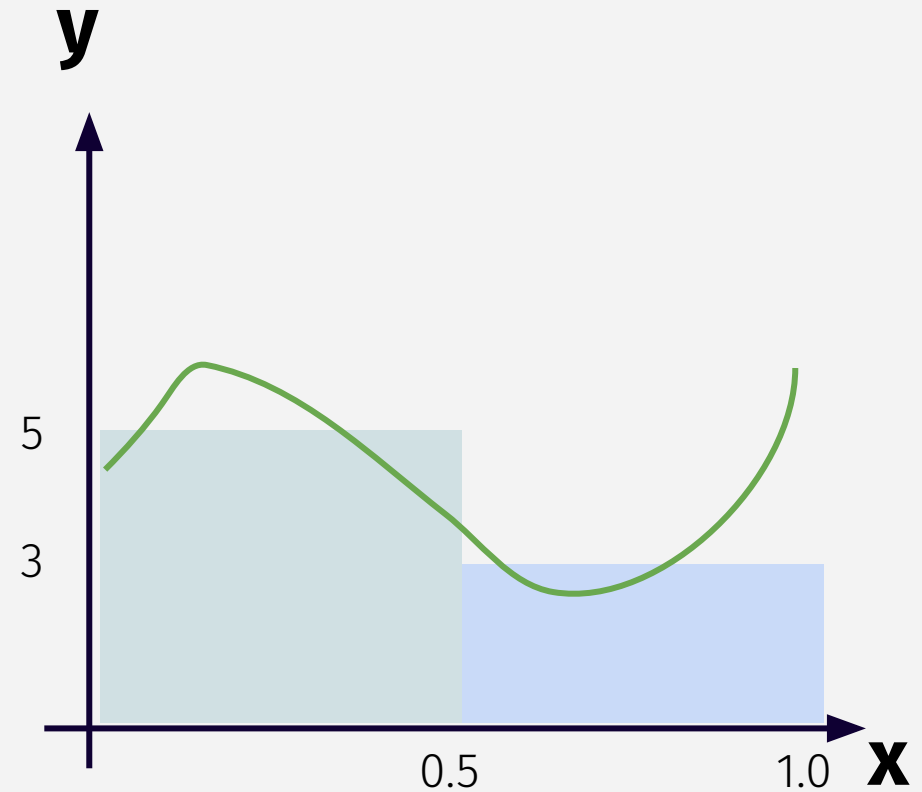
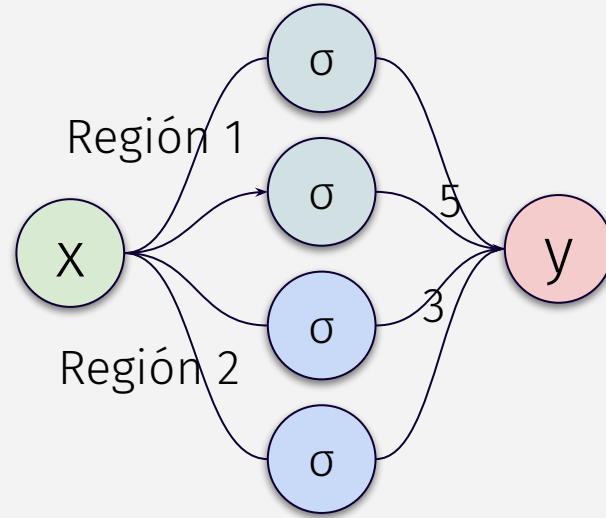
Teorema(s) de Cybenko - Intuición

- Problema de regresión 1D
 - Predecir **y** en base a **x**
- Red:
 - $y = w_2 \sigma(w_1 x + b_1) + b$
 - Equivalente
 - $o = \sigma(w_1 x + b_1)$
 - $y = w_2 o + b_2$
- Estrategia
 - $\sigma(w_1 x + b)$: selector de *región* de x
 - $w_2 o + b_2$: valor y para *región*



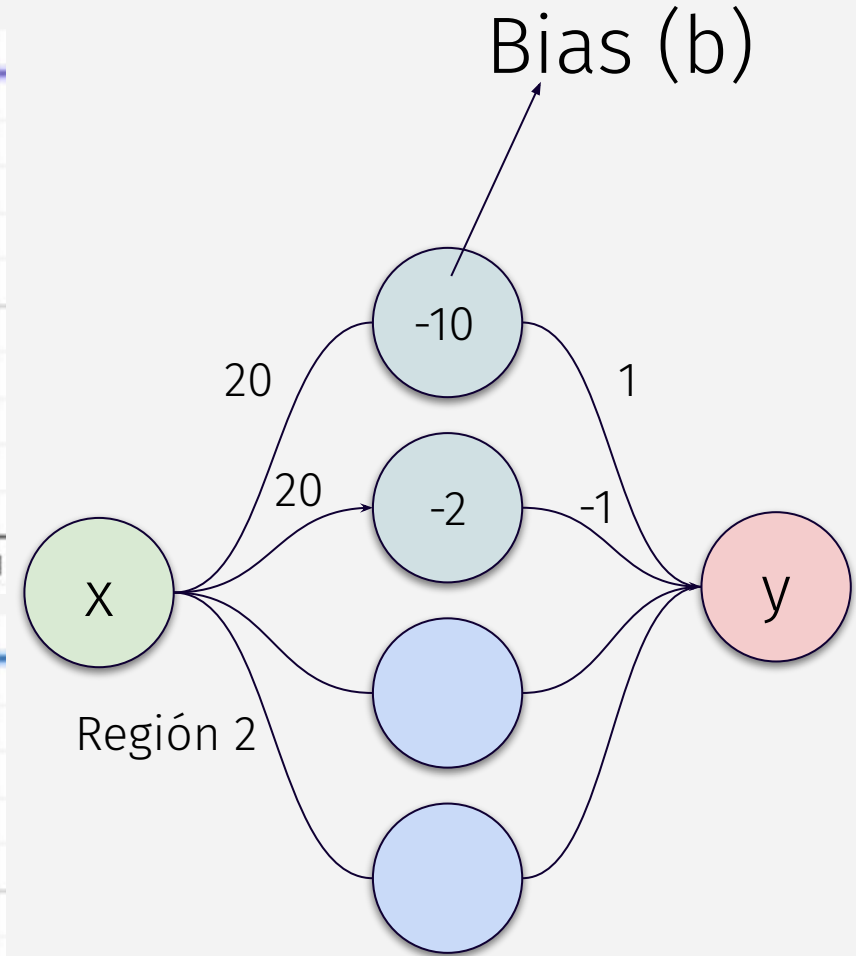
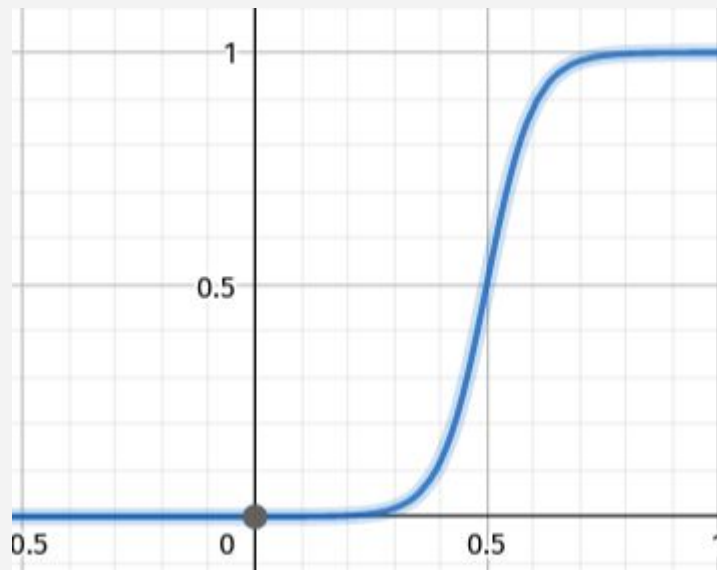
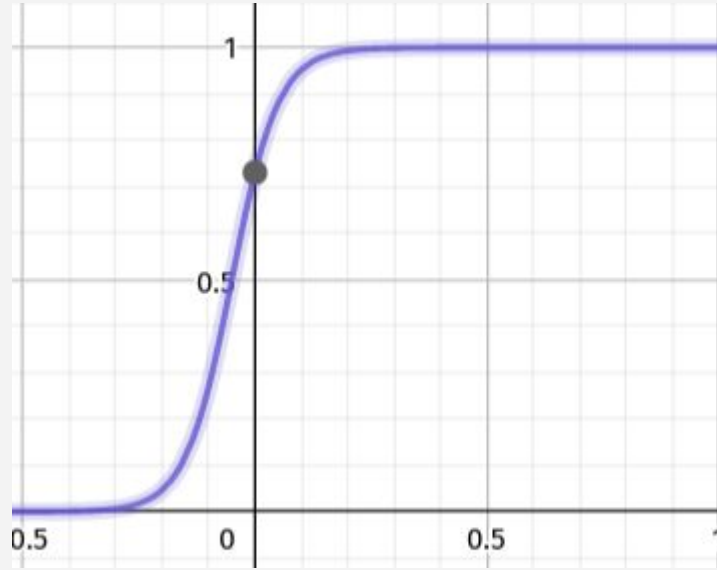
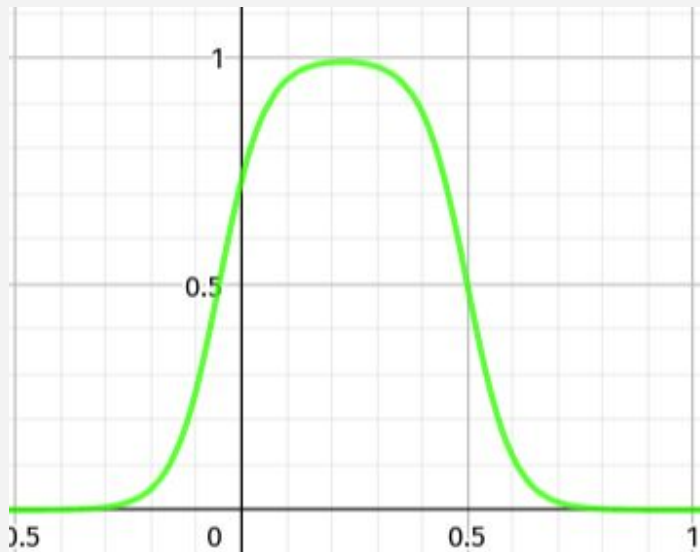
Teorema(s) de Cybenko - Intuición

- Ejemplo con 4 neuronas ocultas
 - w_1 es de 1×4
 - b_1 es de 4
 - w_2 es de 4×1
 - b_2 es de 1
- $4/2=2$ regiones
 - 2 neuronas por región
 - **Selectores** de región
- Región 1: \mathbf{x} de 0 a 0.5
 - valor = 5
- Región 2: \mathbf{x} de 0.5 a 1
 - valor = 3



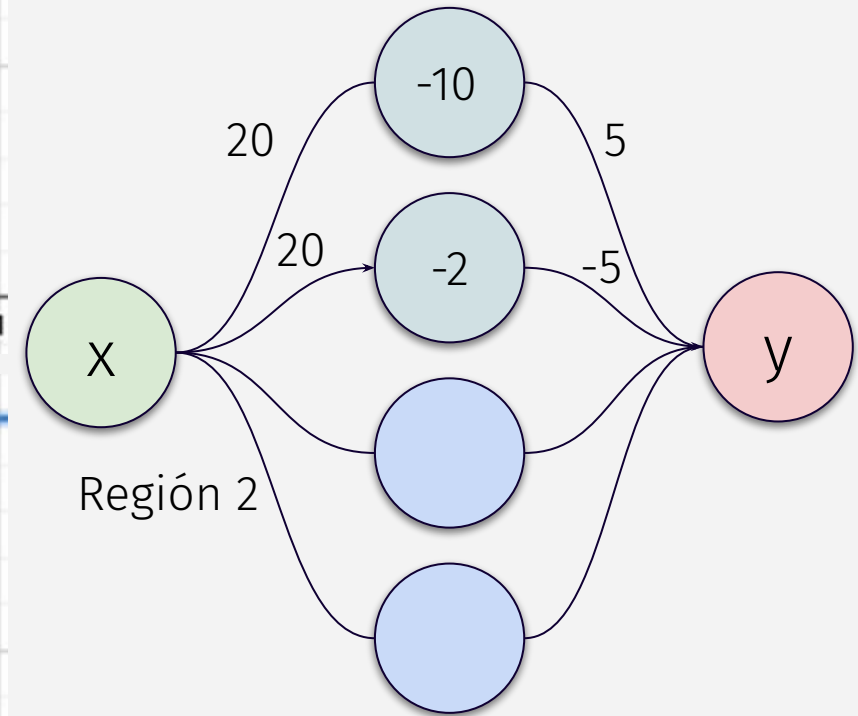
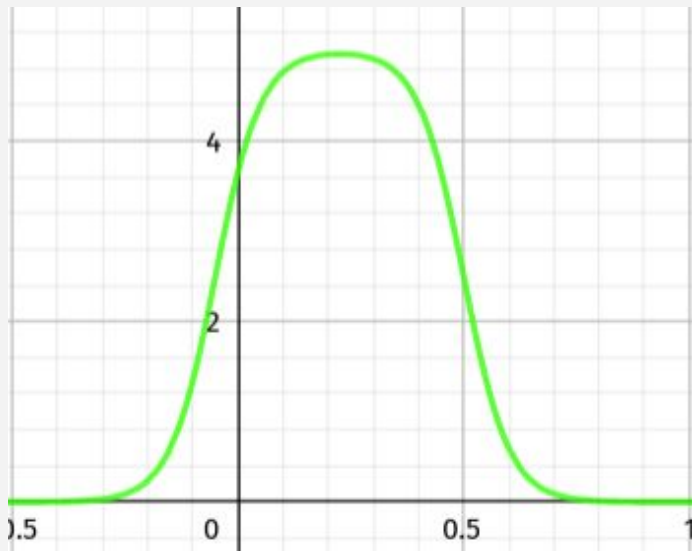
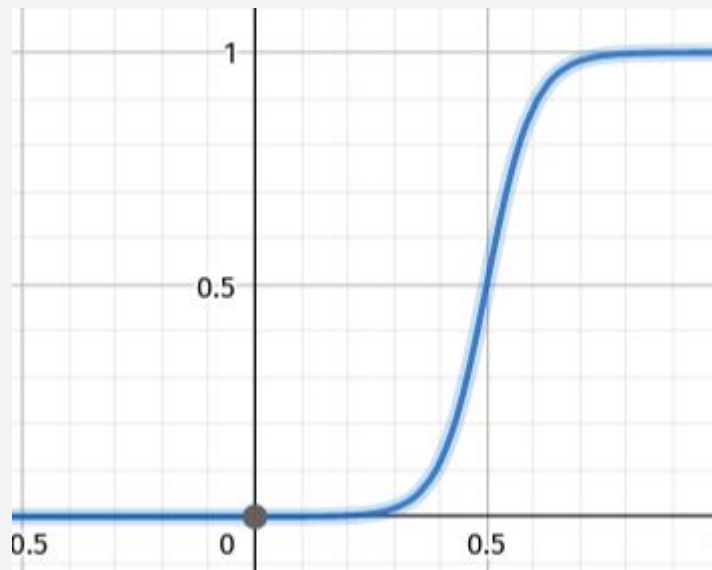
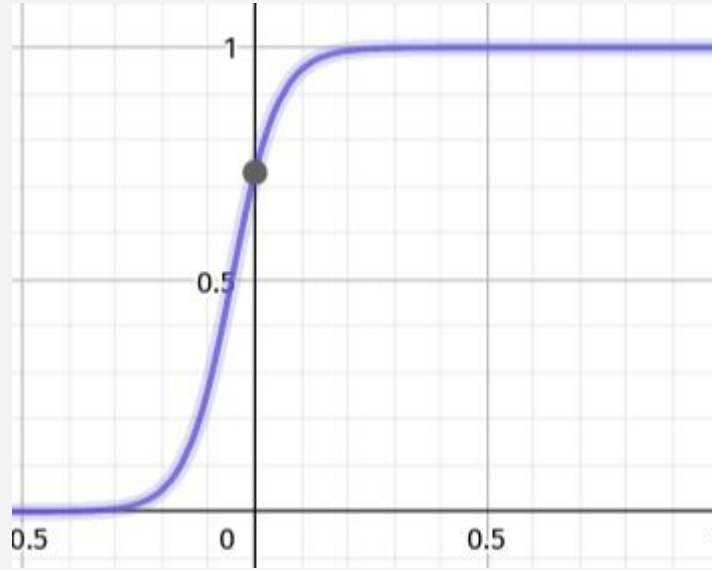
Selector de región 1

- Sigmoidea
 - $\sigma(x) = 1 / (1+e^{-x})$
 - $\sigma(x) = 1 / (1+e^{-(wx+b)})$
- $g(x) = \sigma(20x-10)$
- $h(x) = \sigma(20x-2)$
- $g(x)-h(x)$



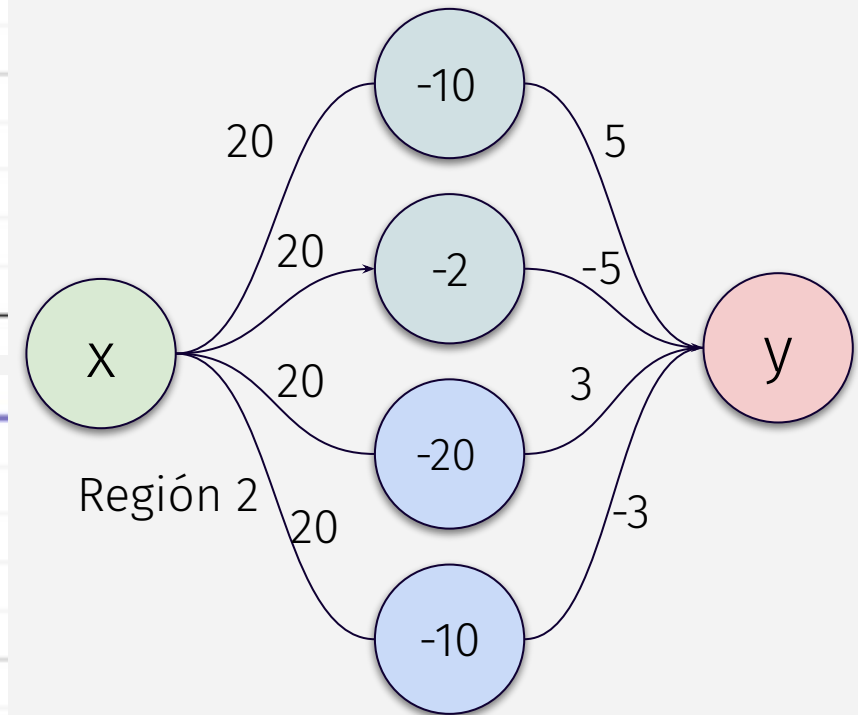
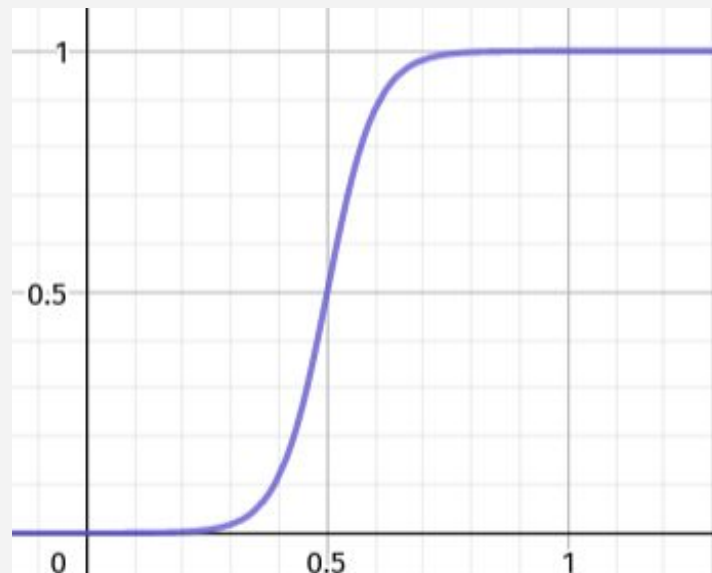
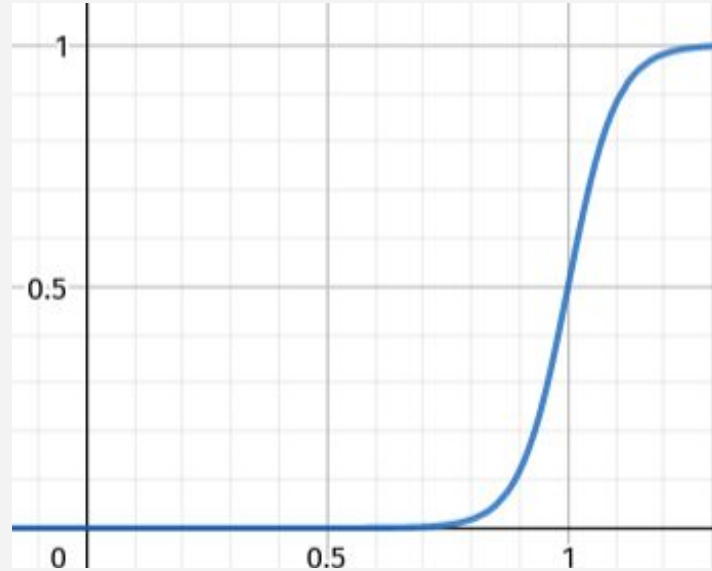
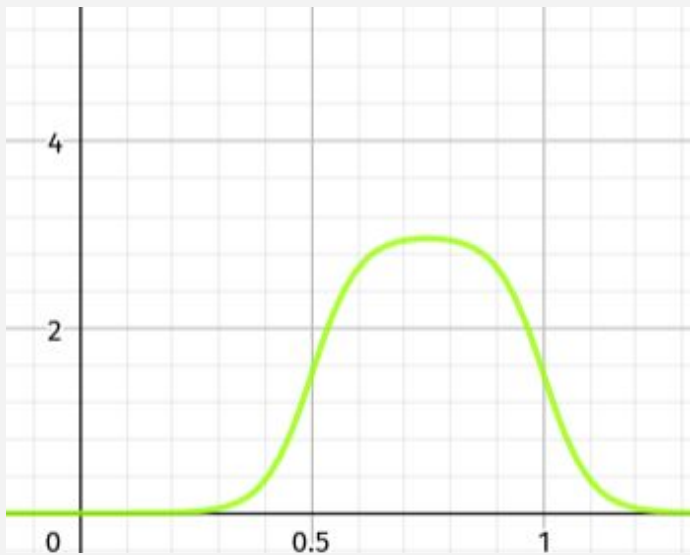
Valor para región 1: 5

- Sigmoidea
 - $\sigma(x) = 1 / (1+e^{-x})$
 - $\sigma(x) = 1 / (1+e^{-(wx+b)})$
- $g(x) = \sigma(20x-10)$
- $h(x) = \sigma(20x-2)$
- $5g(x) - 5h(x)$



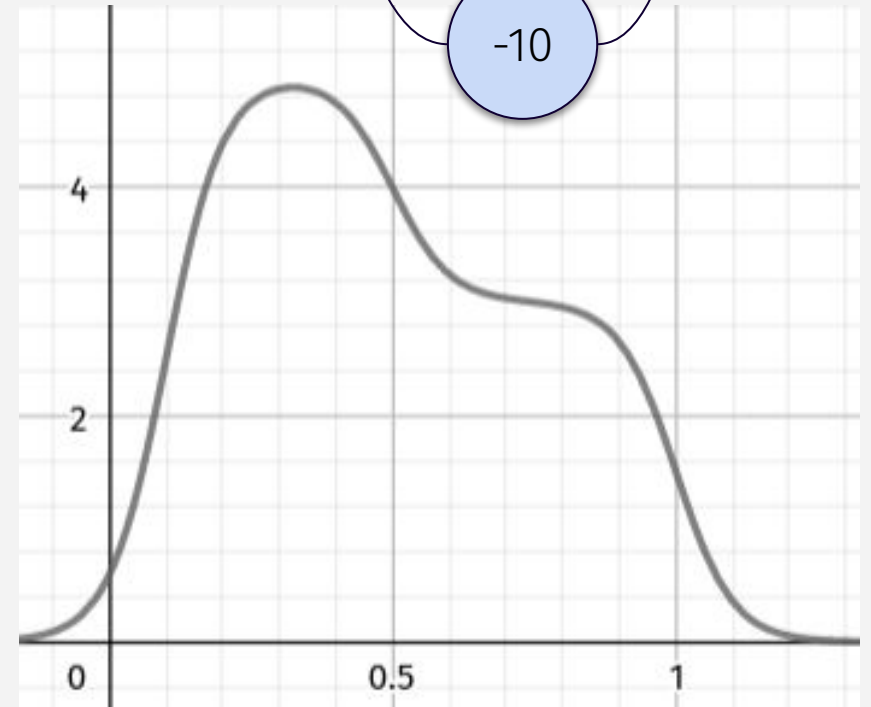
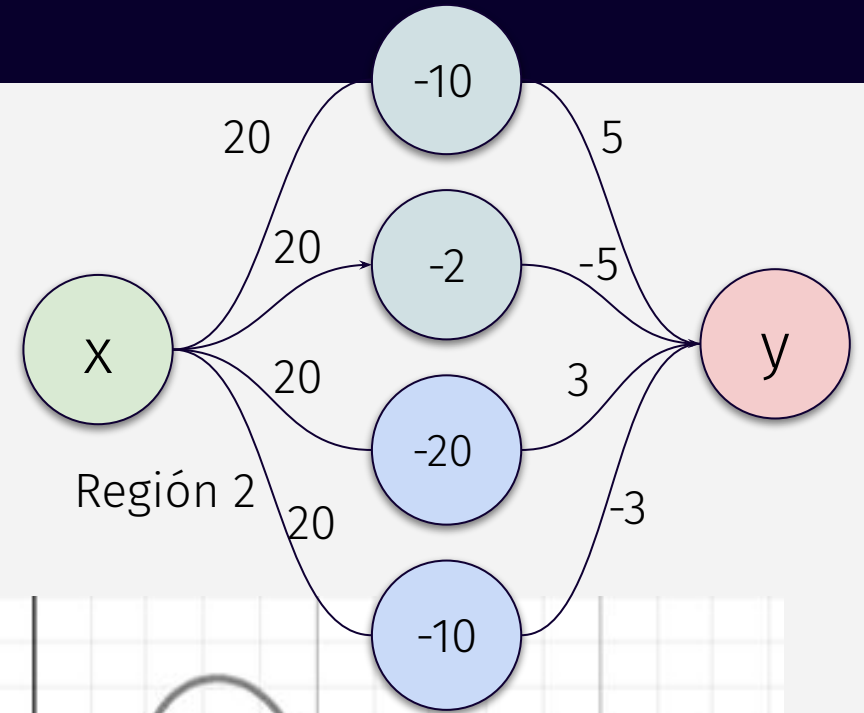
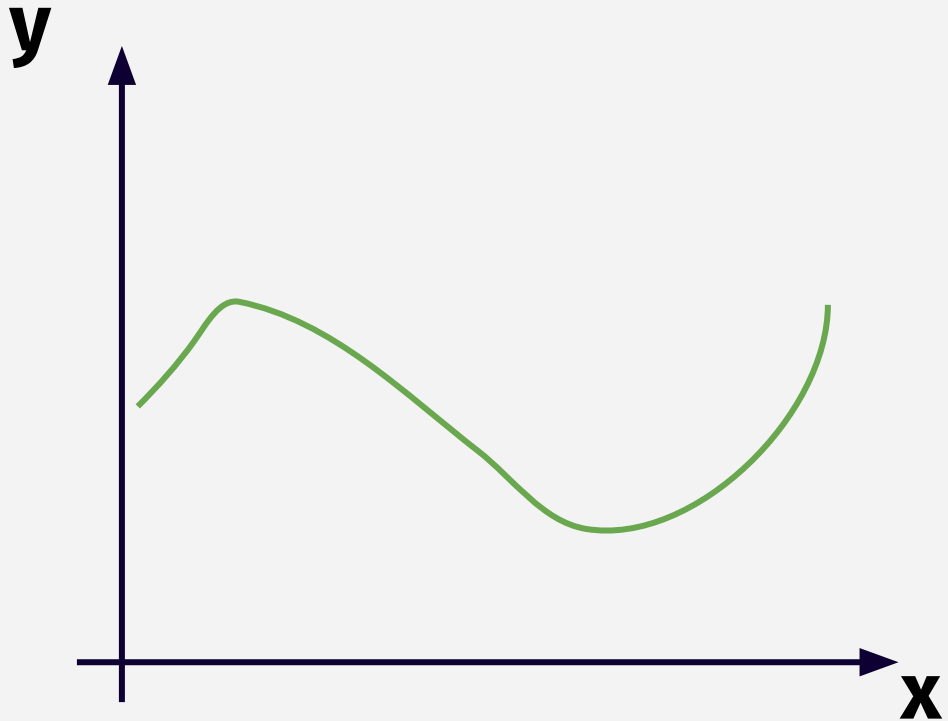
Selector región 2 (0.5 a 1) y valor (3)

- Sigmoidea
 - $\sigma(x) = 1 / (1+e^{-x})$
 - $\sigma(x) = 1 / (1+e^{-(wx+b)})$
- $g(x) = \sigma(20x-20)$
- $h(x) = \sigma(20x-10)$
- $3g(x) - 3h(x)$



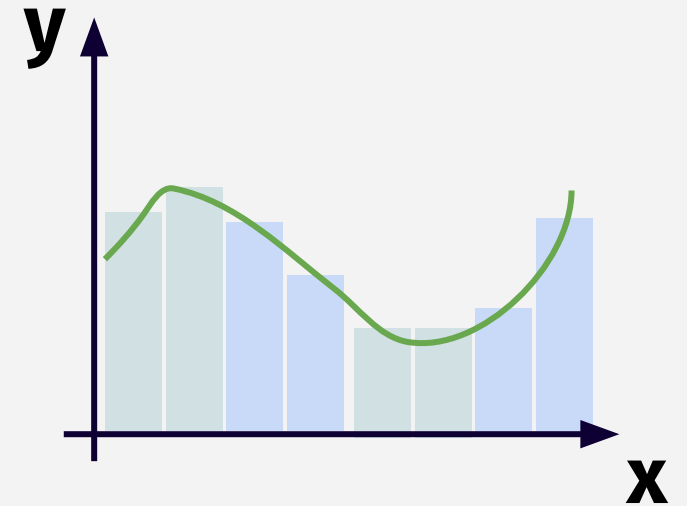
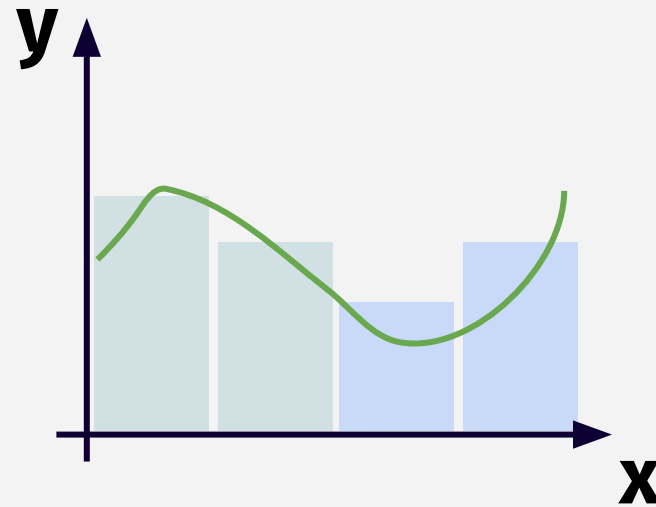
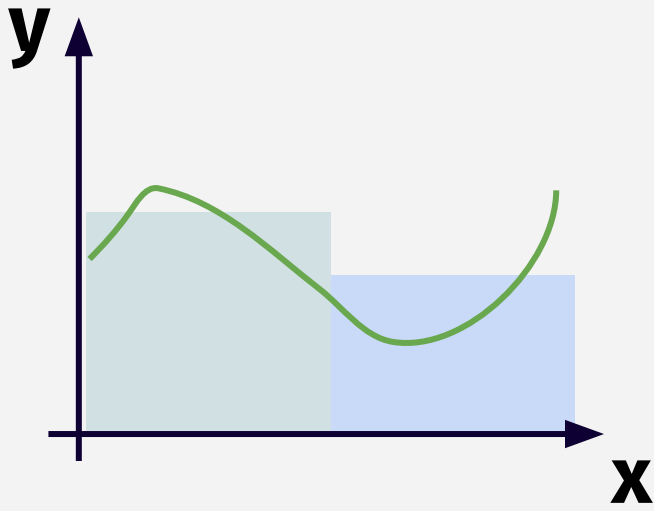
Función final

- $$y = 5 \sigma(20x-10) - 5 \sigma(20x-2) + 3 \sigma(20x-20) - 3 \sigma(20x-10)$$



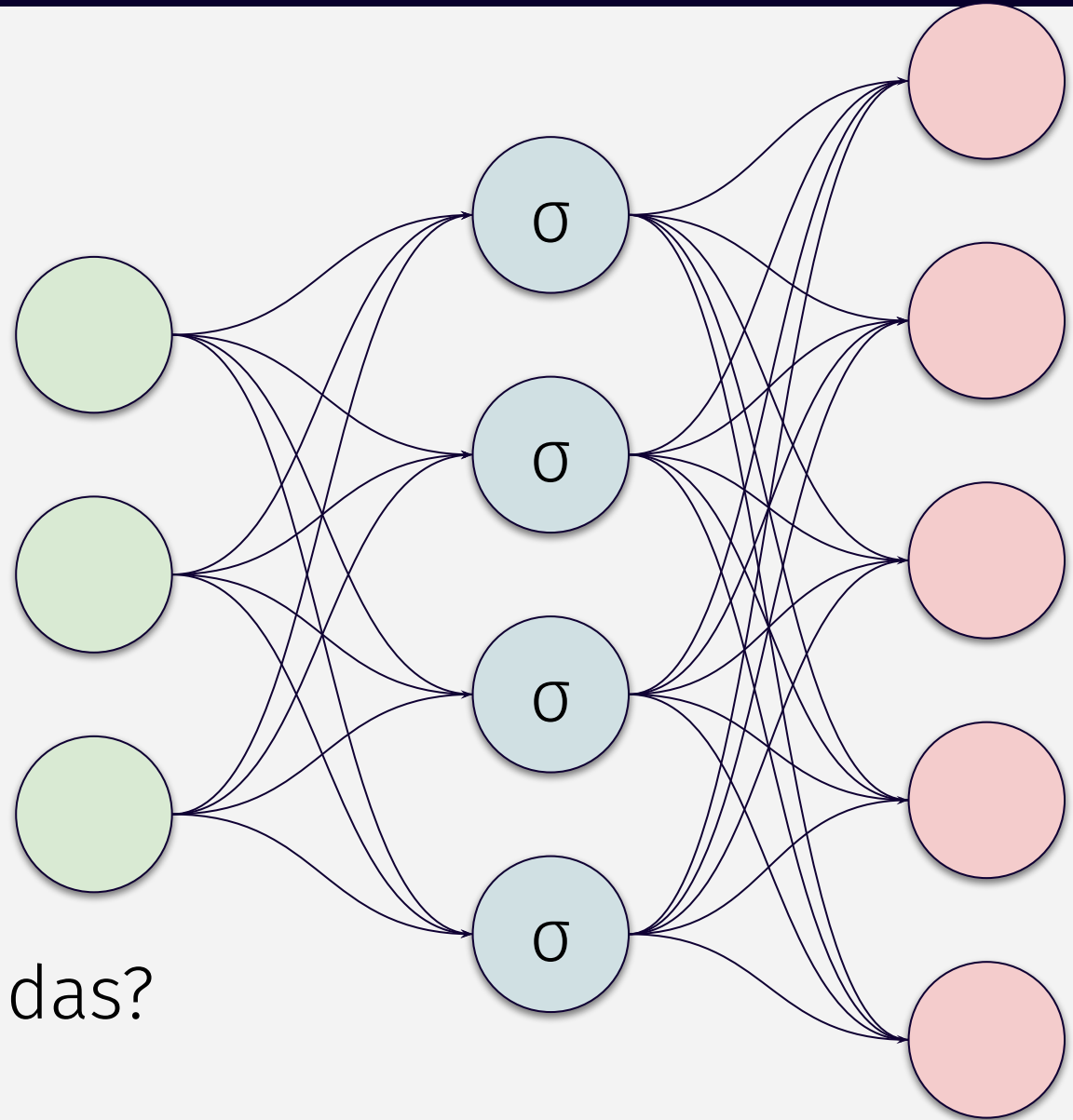
¿Usando más regiones?

- Más regiones
 - Más parámetros
- Aproximación arbitraria



Teorema(s) de Cybenko - Resumen

- Red 2 capas
 - Aproximador universal
- Capa 1
 - **Selectores** de región
- Capa 2
 - Asignan valor por región
- + regiones
 - + neuronas
 - + parámetros
 - + precisión
- ¿Otros problemas? ¿más salidas?
- ¿Reutilización de cómputo?



Recursos

- Visor interactivo de sigmoidea
- Visor interactivo de selector de región y asignador de valor con 2 sigmoideas
- Explicación más detallada (Lorke et al)
- Explicación con visualizaciones interactivas (Nielsen)