Perceptrón

para Redes Neuronales

Verónica E. Arriola-Rios

Facultad de Ciencias, UNAM

5 de febrero de 2023

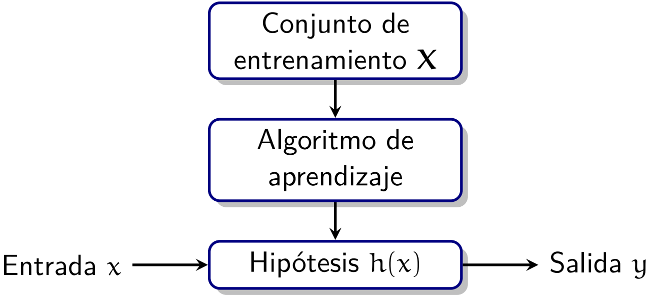
Perceptrón

Temas

1. Perceptrón
   - Clasificación lineal
   - Compuertas lógicas con neuronas

Problema

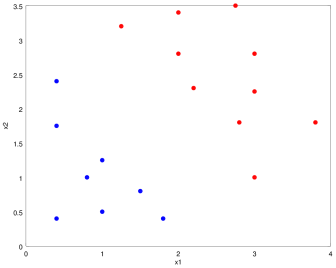
- En un problema de clasificación se pretende, dadas las características de un ejemplar, identificar la clase a la cual pertenece.



$$h(x) : \mathbb{R}^n \rightarrow \{e | e \text{ es una etiqueta de clase}\}$$

Clasificación con una sola frontera

- En el caso más sencillo, se trata de separar los datos de entrada en dos clases.



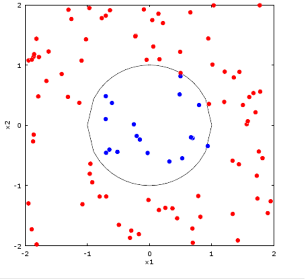


Figura: Problema de clasificación, ilustrado en el plano. Izq: frontera lineal, se dice que los datos son *linealmente separables*. Der: frontera circular.

Perceptrón (Rosenblatt 1958)

- Un *perceptrón* es una función matemática inspirada en el funcionamiento de una neurona.
- Sirve para obtener fronteras lineales entre datos multivariados mediante un algoritmo de aprendizaje.
- Fueron popularizados por el psicólogo Frank Rosenblatt a inicios de los 1960's, en un libro llamado *Principios de neurodinámica*, donde presentó varios modelos de perceptrones.
- Los perceptrones cayeron en desuso porque Minsky y Papert demostraron sus limitaciones en un libro llamado *Perceptrones*.

Perceptrón

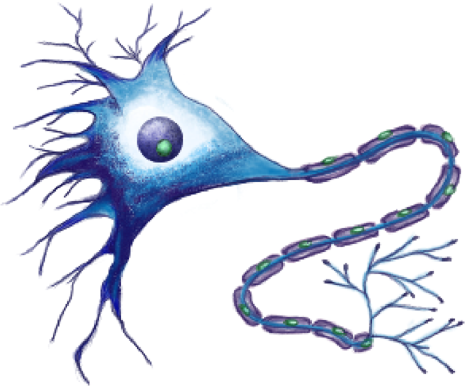


Figura: Neurona.

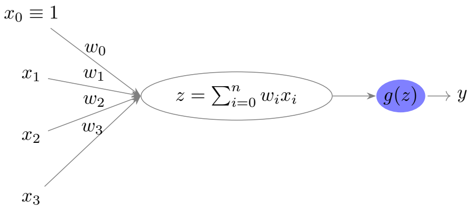
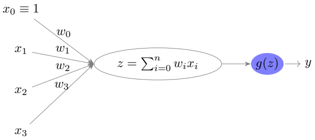


Figura: Perceptrón.

Donde:

- x_0 es el *sesgo* (*bias*),
- g es la *función de activación*,
- w_i son los *pesos* de las conexiones de entrada.

Función de activación



Donde g, la *función de activación*, puede ser alguna de:

- Función escalón

$$g = \begin{cases} 0 & \text{si } z < 0 \\ 1 & \text{si } z \geqslant 0 \end{cases}$$

- ReLU

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

- Logística

$$g = \frac{1}{1 + e^{-z}}$$

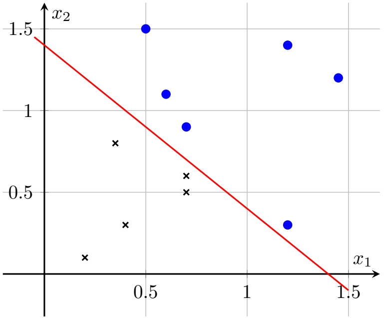
- Tangente hiperbólica

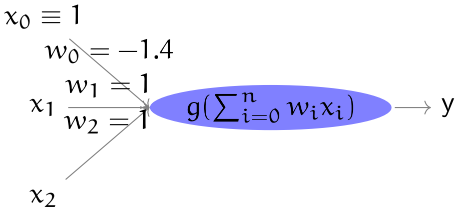
$$\tanh = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

Interpretación

El perceptrón modela una neurona que dispara o no dependiendo de los valores de sus señales de entrada.

Clasificación



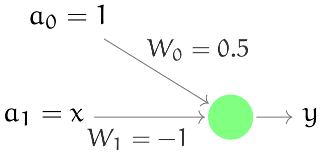


$$y = \frac{1}{1 + e^{-(-1.4 + x_1 + x_2)}}$$

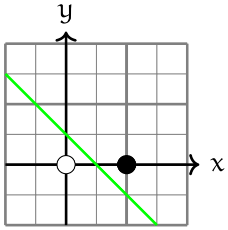
Temas

1. Perceptrón
   - Clasificación lineal
   - Compuertas lógicas con neuronas

NOT



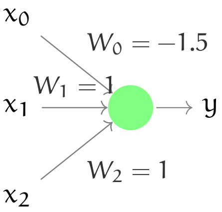
NOT

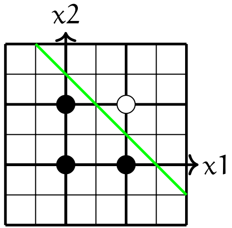


$$Salida = g(0.5 - x)$$

x	z	y
0	0.5	1
1	-0.5	0

AND





$$\text{Salida} = g(-1.5 + x_1 + x_2)$$

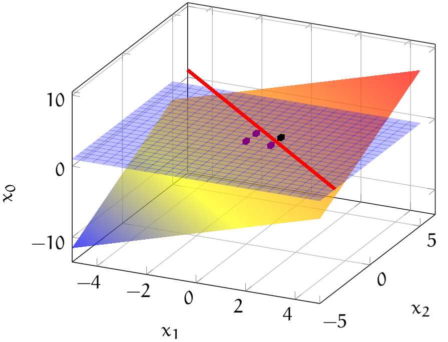
x_1	x_2	z	h_{umbral}
0	0	-1.5	0
0	1	-0.5	0
1	0	-0.5	0
1	1	0.5	1

$$\text{Salida} = g(-15 + 10x_1 + 10x_2)$$

x_1	x_2	z	$h_{sigmoide}$
0	0	-15	0.0000003
0	1	-5	0.0067
1	0	-5	0.0067
1	1	5	0.993

Linealidad

Para que un perceptrón en $\mathbb{R}^2$ represente una función lineal, debemos considerar el espacio tridimencional $\mathbb{R}^3$.



Entrenamiento

- En ocasiones, no es posible visualizar los datos de entrada, especialmente si son multidimensionales.
- El perceptrón incluye un algoritmo de entrenamiento que permite encontrar los pesos adecuados para realizar la predicción correcta.

Seleccionar casos de entrenamiento utilizando cualquier política que garantice que cada caso de entrenamiento continuará siendo elegido.

1. Si la salida es correcta, dejar los pesos como están.
2. Si la salida incorrectamente indica 0, sumar el vector de entrada al vector de pesos.
3. Si la salida incorrectamente indica 1, restar el vector de entrada al vector de salida.

Está garantizado que este procedimiento encontrará un conjunto de pesos que obtienen la respuesta correcta para todos los casos de entrenamiento, *si existen*. Esto dependerá de que se hayan elegido las características correctas.

Uso de los perceptrones

En reconocimiento de patrones estadísticos tradicionalmente se siguen tres pasos:

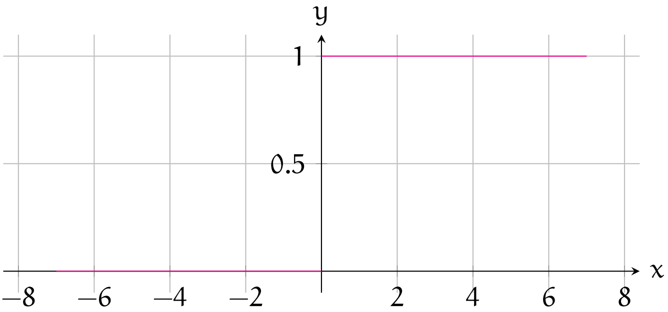
1. Se convierten los vectores de entrada en vectores de activación, que representan características. Estas características se eligen a mano.
2. Se aprende con qué peso contribuye cada una de las caracteríticas, y la suma pesada de las entradas determina un valor numérico.
3. Si esta cantidad se encuentra por encima de un valor numérico, se dice que el vector de entrada corresponde con la clase objetivo.

Funciones de activación

Función escalón

$$y = \begin{cases} 1 & \text{si } x \geqslant 0 \\ 0 & \text{si } x < 0 \end{cases} \tag{1}$$

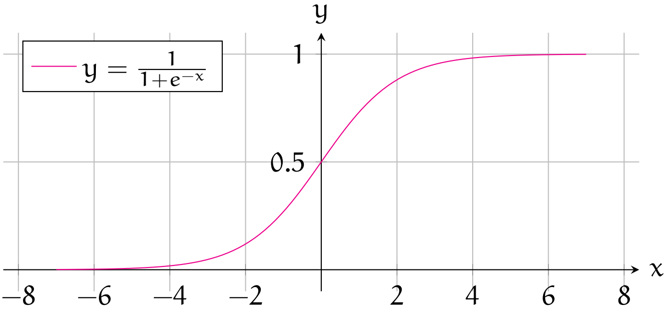
Rectificadora o función rampa



Función logística

$$y = \frac{1}{1 + e^{-x}} \quad (2)$$

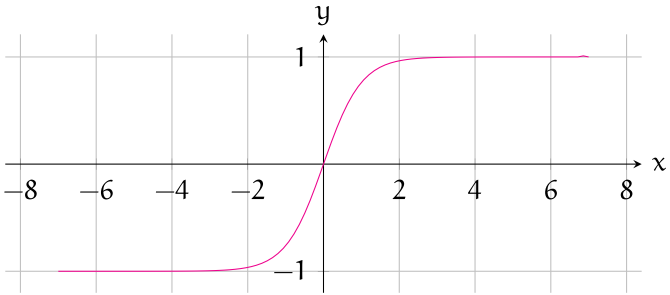
Función logística



Tangente hiperbólica

$$y = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{3}$$

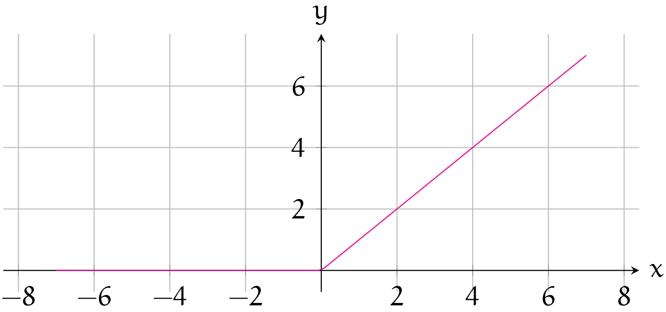
Tangente hiperbólica



Rectificadora lineal

$$y = max(0, x) \tag{4}$$

Rectificadora o función rampa



Funciones de error

Funciones de error

- Para determinar si el perceptrón está clasificando correctamente los datos o no, se definen funciones de error.
- Para entrenar un perceptrón basta con utilizar cualquier método de optimización de funciones para encontrar los parámetros θ que minimizan la función de error.
- Sea $y^{(m)}$ el valor de salida deseado para el m-ésimo ejemplar de entrenamiento para un perceptrón con pesos Θ, es decir, la clasificación correcta 0 ó 1.
- Sea $a^{(m)}$ el valor de salida del perceptrón.
- Existen varias funciones de error, pero se presentan a continuación las dos más utilizadas.

Diferencias al cuadrado

La función de error *diferencias al cuadrado* se define como:

$$L(\Theta) = \frac{1}{2m} \sum_{m=0}^{M-1} (y^{(m)} - a^{(m)})^2 \tag{5}$$

Históricamente fue la primera en popularizarse, pero no es la más adecuada para problemas de clasificación.

Entropía cruzada

La función de error *entropía cruzada* se define como:

$$L(\Theta) = -\frac{1}{m} \sum_{m=0}^{M-1} \left(y^{(m)} \log(a^{(m)}) + (1 - y^{(m)}) \log(1 - a^{(m)}) \right) \tag{6}$$

Esta función tiene un mejor comportamiento ya que define una superficie más suave en el espacio de parámetros para problemas de clasificación.

Referencias I

Haykin, Simon (2009). *Neural Networks and Learning Machines*. 3rd. Prentice Hall, Pearson.

Machine Learning, Andrew NG, https://www.coursera.org/learn/machine-learning

Licencia