

Redes Neuronales

Jorge Gallego

Facultad de Economía, Universidad del Rosario

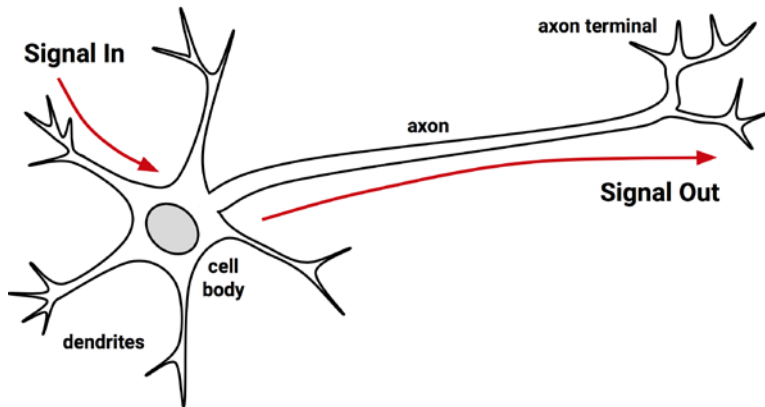
Mayo 5 de 2020

Introducción

- El cerebro humano se compone de una red de neuronas interconectadas que crean un procesador paralelo masivo
- Cada neurona cuenta con una estructura que le permite captar las señales provenientes del exterior
- Dichos *inputs* sensoriales son procesados, y bajo ciertas circunstancias, un estímulo eléctrico es emitido al sistema
- Todo lo cual redunda en una serie de reacciones en el cerebro y cuerpo de un ser vivo
- Las redes neuronales artificiales buscan imitar esa relación entre señales de entrada y de salida en los cerebros

Introducción

Figure: Neuronas Biológicas

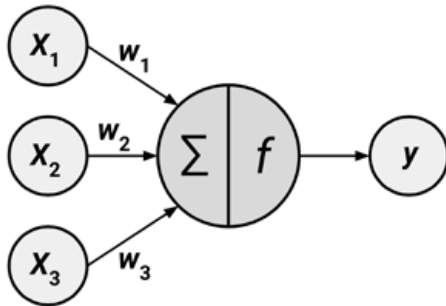


Neuronas Biológicas

- Las señales de entrada son recibidas por dendritas de las células a través de procesos bioquímicos
- El impulso es ponderado de acuerdo a su importancia o frecuencia
- El soma acumula señales de entrada y si se alcanza cierto umbral, se transmite una señal de salida
- Dicha señal es transmitida vía un proceso electroquímico a través del axón de la neurona
- En las terminales del axón, la señal eléctrica es nuevamente procesada como señal química y pasada a otras neuronas

Neuronas Artificiales

Figure: Neuronas Artificiales



Neuronas Artificiales

- Las neuronas artificiales buscan imitar el proceso de las biológicas
- La figura ilustra un diagrama de redes dirigidas
- Se modela la relación entre señales de entrada (variables x) y de salida (y)
- Cada señal de la dendrita es ponderada (w) de acuerdo con su importancia relativa
- Las entradas son sumadas por el soma y son transmitidas de acuerdo con una función de activación $f(\cdot)$

Neuronas Artificiales

La típica neurona artificial con n dendritas de entrada puede representarse por medio de la función:

$$y(x) = f \left(\sum_{i=1}^n w_i x_i \right)$$

- Los pesos w determinan cuánto importa cada variable de entrada
- La función de activación $f(\cdot)$ determina la señal (y) emitida en función de la suma neta de entradas

Neuronas Artificiales

Las neuronas de este tipo determinan las redes neuronales artificiales, que suelen tener tres componentes:

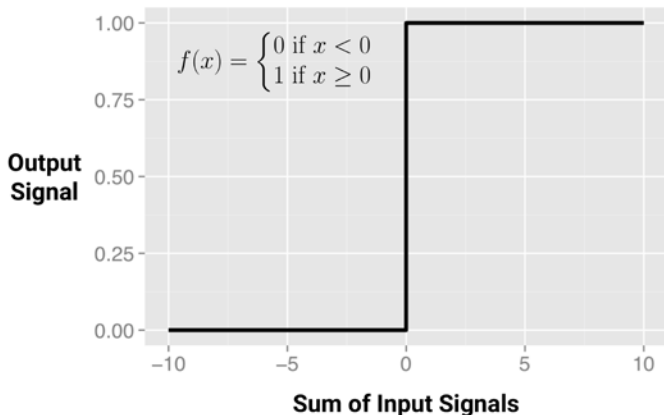
1. Función de activación: transforma la combinación de señales de entrada en una de salida transmitida a resto de la red
2. Topología de la red: número de neuronas, capas y la manera en la que están conectadas
3. Algoritmo de entrenamiento: especifica cómo se fijan los pesos de conexión para inhibir o excitar a las neuronas

Funciones de activación

- Son el mecanismo a través del cual se procesa la información de entrada y se pasa al resto de la red
- En biología, se modela sumando las entradas y determinando si se cruza el umbral de activación
- Cuando se cruza, se produce una señal de salida al resto de la red
- El proceso es dicotómico: se cruza o no el umbral, luego se activa o no la señal de salida

Funciones de Activación

Figure: Función de Activación de un Solo Paso

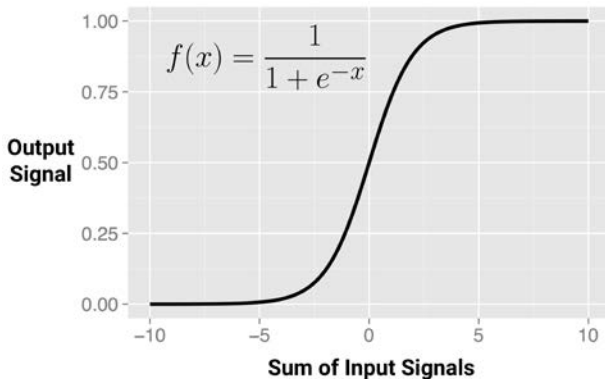


Funciones de activación

- Las redes neuronales tienen un funcionamiento ligeramente distinto
- El resultado de la función de activación no tiene por qué asumirse dicotómico
- Se permite más flexibilidad, que depende de la capacidad del modelo para ajustarse a los datos
- Por tanto, existen diferentes tipos de funciones de activación que relacionan señales de entrada con señales de salida
- Una de las más populares es la sigmoideal logística

Función de Activación Sigmoidal

Figure: Función de Activación de un Solo Paso

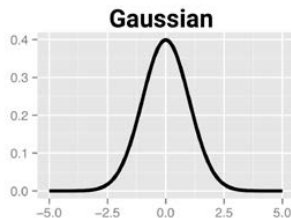
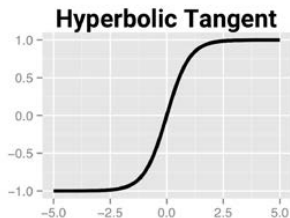
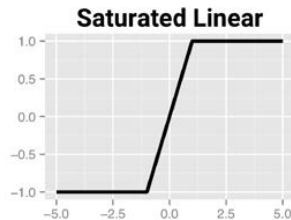
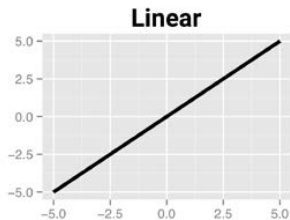


Función de Activación Sigmoidal

- Nótese que el resultado de esta función no es binario
- Sino una función continua cuyo outcome está entre 0 y 1
- Además, la función es diferenciable, lo cual es deseable para procesos de optimización
- Esta es la función más popular, pero existen otras

Funciones de Activación

Figure: Otras Funciones de Activación



Funciones de Activación

- Las funciones pueden diferir en el rango de la señal de activación
- Cuál se use depende de qué tan buen ajuste haya con los datos en cuestión
- Muchas veces es difícil explicar teóricamente por qué una función se ajusta mejor a ciertos datos
- Por esta razón este método suele considerarse caja negra

Topología de la Red

La topología de la red es el conjunto de patrones y estructuras de las neuronas interconectadas. Consta de tres características:

1. Número de capas
2. Capacidad de las señales para viajar hacia atrás
3. Número de nodos en cada capa de la red

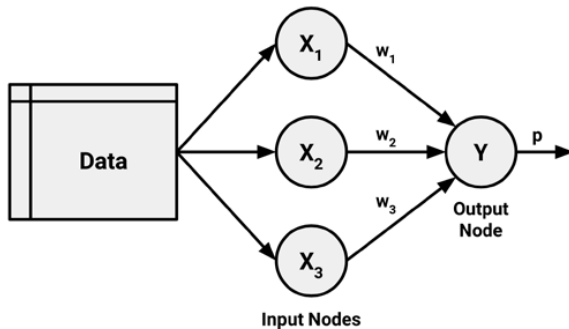
Estas características determinan la complejidad de las labores a ser aprendidas por la red

Número de capas

- Los nodos de entrada reciben las señales no procesadas
- Cada nodo se encarga de procesar una sola característica de los datos
- Se le transmite esto al nodo de salida, que con su función de activación genera la predicción final
- Los nodos de entrada y salida se agrupan en capas
- La versión más básica es la que tiene solo una capa

Redes de Una Sola Capa

Figure: Red de Una Sola Capa

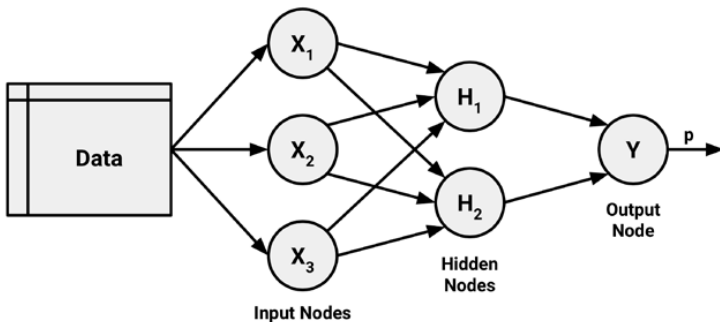


Número de capas

- Pero las redes pueden ser más complejas
- Las redes multi-capas adicionan una o más capas ocultas
- Estas capas adicionales procesan las señales emitidas por los nodos de entrada
- Antes de transmitir la señal final a los de salida
- Por lo general, la redes están totalmente conectadas
- Ya que todos los nodos de una capa se conectan con todos los de la siguiente

Redes de Múltiples Capas

Figure: Red de Múltiples Capas

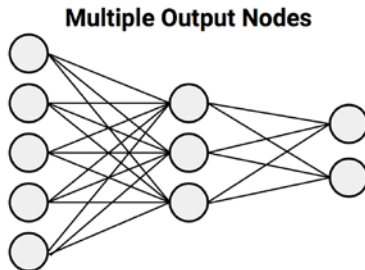


Dirección de Viaje de la Información

- En los ejemplos anteriores la información siempre viaja hacia adelante
- Desde los nodos de entrada, a las capas ocultas, hasta los nodos de salida
- Estos son los *feedforward networks*, que naturalmente son más simples
- Sin embargo estas redes tienen flexibilidad en otros aspectos
- Número de niveles y nodos en cada nivel, múltiples outcomes, o múltiples capas ocultas

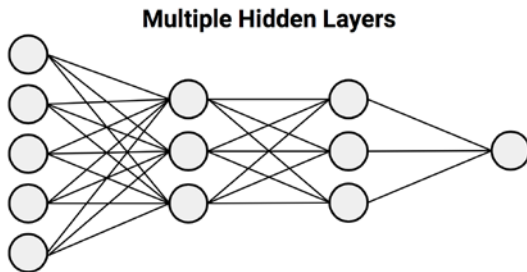
Redes de Múltiples Nodos Terminales

Figure: Red de Múltiples Nodos Terminales



Redes de Múltiples Capas Ocultas

Figure: Red de Múltiples Nodos Terminales



Deep Learning

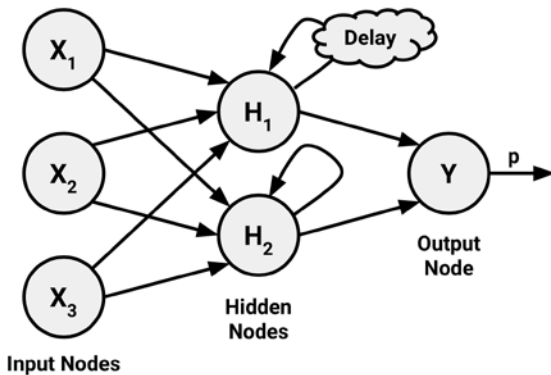
- Una red neuronal con múltiples capas ocultas es llamada una Red Neuronal Profunda (*Deep Neural Network*)
- La práctica de entrenar una red de este tipo suele llamarse *deep learning*
- La cual ha ganado popularidad en años recientes
- En particular por su alta capacidad para clasificar datos no-estructurados como imágenes, sonidos o texto

Dirección de Viaje de la Información

- Pero hay redes en las que se permite que las señales viajen en ambas direcciones por medio de loops
- Estas redes son llamadas *redes recurrentes*
- Naturalmente este tipo de redes se parecen más a las biológicas
- Las cuales permiten incorporar memoria de corto plazo y retrasos para modelar patrones mucho más complejos.
- No profundizamos mucho en ellas porque no son tan usadas en la práctica

Redes Recurrentes

Figure: Red Recurrente



Número de Nodos en Cada Capa

- El último componente de la topología de una red es el número de nodos en cada capa
- El número de nodos de entrada está predeterminado por el número de características en los datos
- Pero el número de nodos ocultos es decidido por el usuario antes de entrenar el modelo
- No hay una regla teórica que determine cuántas neuronas y conexiones establecer
- Como es habitual, deber tenerse cuidado de no sobreestimar el modelo de acuerdo con los datos de entrenamiento

Método Backpropagation

- ¿Cómo entrenar una red neuronal dados unos datos y toda esta estructura?
- Un método muy usado, dada su eficiencia computacional, es el de *backpropagation*
- El método itera en muchos ciclos dos procesos básicos
- Cada ciclo es llamado una *epoca*
- Se inicia fijando unas ponderaciones aleatorias, dado que no hay conocimiento previo

Método Backpropagation

- Luego, el algoritmo itera a través de los procesos hasta llegar a algún criterio de detención
- Cada época incluye dos fases:
 1. Fase delantera: Cada neurona es activada en secuencia desde las de entrada hasta las de salida usando los pesos establecidos y la función de activación. Al final, se produce la señal de salida
 2. Fase trasera: la señal de salida es comparada con el valor objetivo en los datos de entrenamiento. El error es usado hacia atrás para modificar los pesos de conexión entre neuronas y reducir los errores
- El proceso de reducción del error se hace usando una técnica llamada *gradient descent*