

Universidad TecMilenio

Nombre del alumno: Alexis Leal Mata

Máster en Inteligencia Artificial

Actividad 6

Aprendizaje profundo

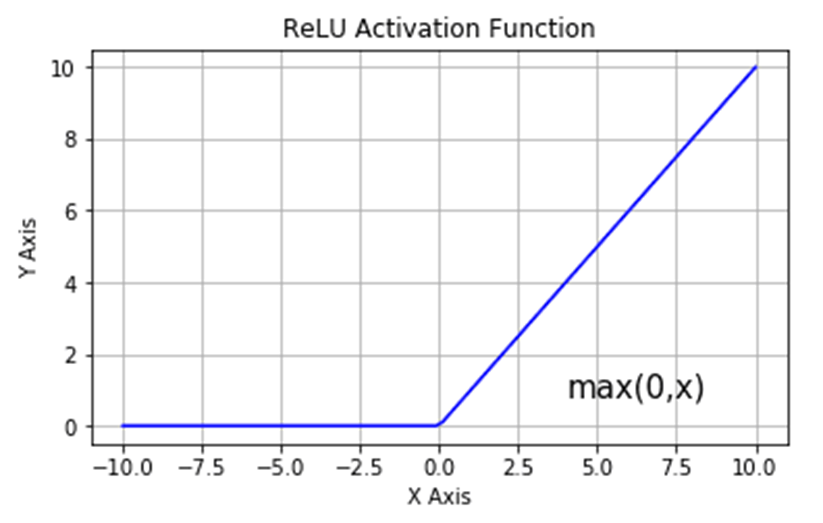
# Cuestionario

# 

Las funciones de activación, también conocidas como funciones de transferencia, se utilizan para calcular la salida de cada capa en una red neuronal. Investiga cuáles son las funciones de activación ReLu, Softmax, sigmoide, hiperbólica tangente y lineal, luego da un ejemplo de cada una de ellas.

**ReLu**

La función de activación lineal rectificada o ReLU para abreviar es una función lineal por partes que generará la entrada directamente si es positiva; de lo contrario, generará cero. Se ha convertido en la función de activación predeterminada para muchos tipos de redes neuronales porque un modelo que la usa es más fácil de entrenar y, a menudo, logra un mejor rendimiento**.**

****

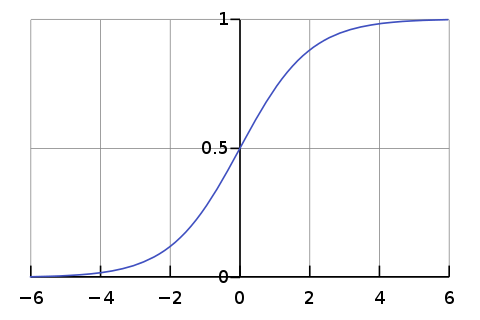
**Softmax**

Es una función matemática que convierte un vector de números en un vector de probabilidades, donde las probabilidades de cada valor son proporcionales a la escala relativa de cada valor en el vector.

El uso más común de la función softmax en el aprendizaje automático aplicado es su uso como función de activación en un modelo de red neuronal. Específicamente, la red está configurada para generar N valores, uno para cada clase en la tarea de clasificación, y la función softmax se usa para normalizar las salidas, convirtiéndolas de valores de suma ponderada en probabilidades que suman uno. Cada valor en la salida de la función softmax se interpreta como la probabilidad de membresía para cada clase.

Table

Description automatically generated with low confidence

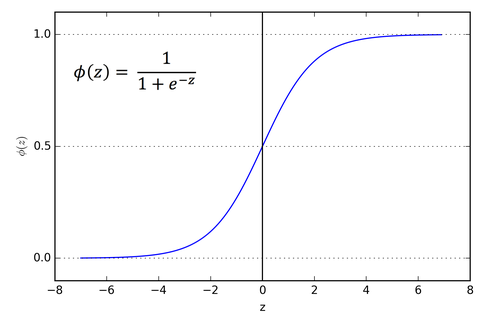


**Sigmoide**

La función sigmoidea es una forma especial de la función logística y generalmente se denota por σ(x) o sig(x). Está dado por:

σ(x) = 1/(1+exp(-x))

La función sigmoidea también se llama función aplastante ya que su dominio es el conjunto de todos los números reales y su rango es (0, 1). Por lo tanto, si la entrada a la función es un número negativo muy grande o un número positivo muy grande, la salida siempre está entre 0 y 1. Lo mismo ocurre con cualquier número entre -∞ y +∞.



**hiperbólica Tangente**

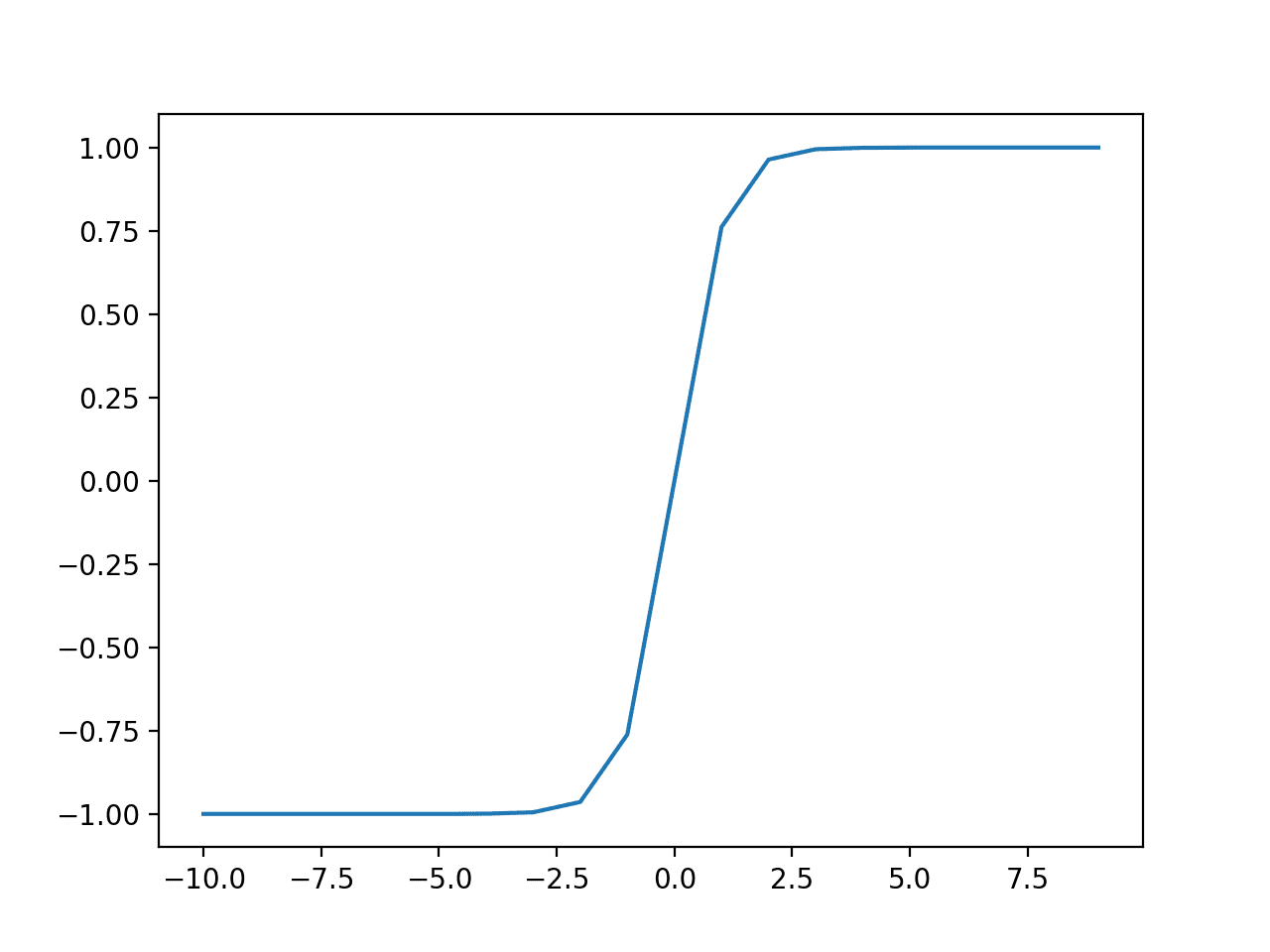
La función de activación de la tangente hiperbólica también se conoce simplemente como la función Tanh (también "tanh" y "TanH").

Es muy similar a la función de activación sigmoidea e incluso tiene la misma forma de S.

La función toma cualquier valor real como entrada y genera valores en el rango de -1 a 1. Cuanto mayor sea la entrada (más positiva), más cerca estará el valor de salida de 1,0, mientras que cuanto más pequeña sea la entrada (más negativa), más cerca la salida será a -1.0.

La función de activación de Tanh se calcula de la siguiente manera:

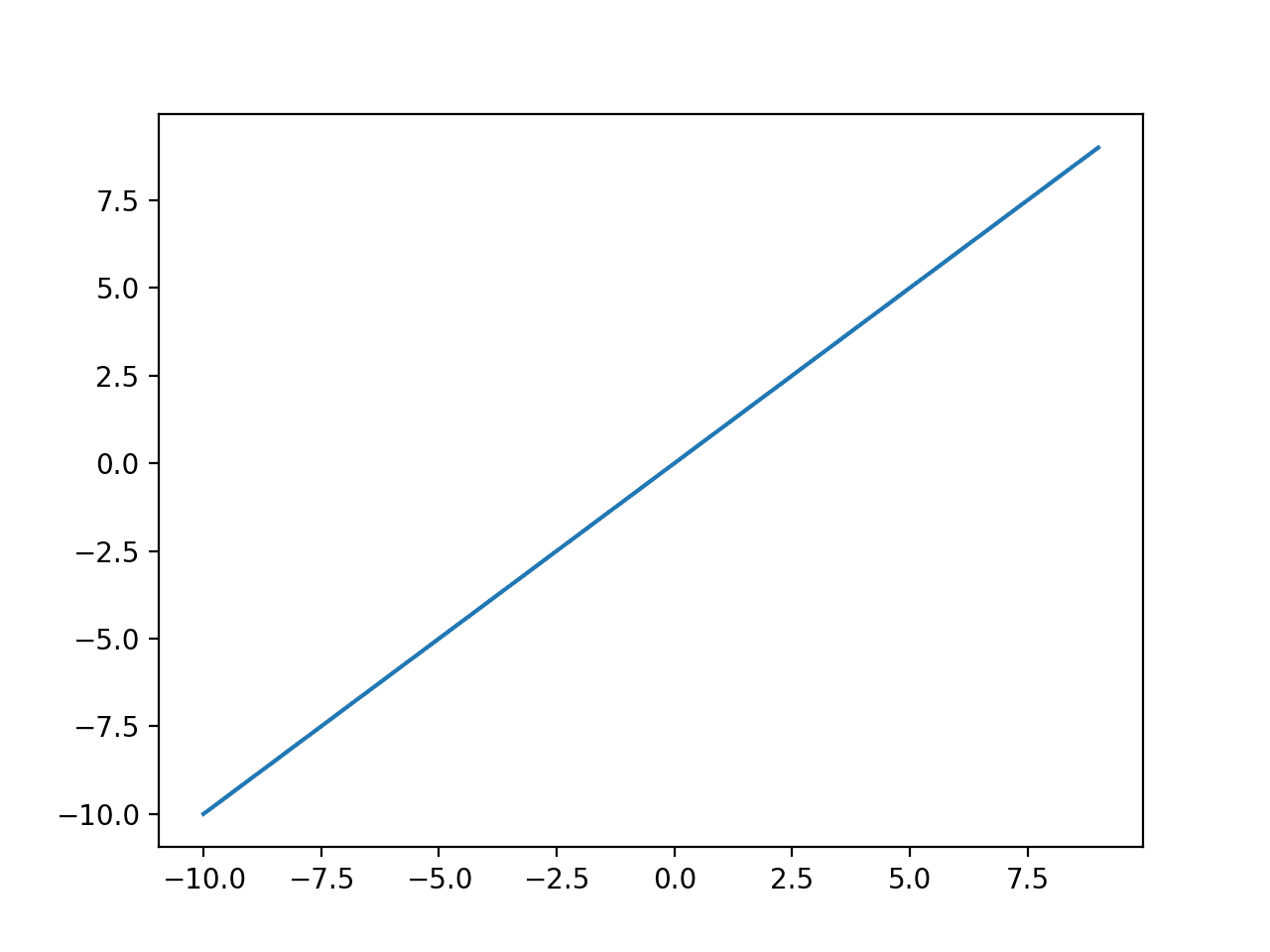
(e^x - e^-x) / (e^x + e^-x)



Lineal

La función de activación lineal también se denomina "identidad" (multiplicada por 1,0) o "sin activación".

Esto se debe a que la función de activación lineal no cambia la suma ponderada de la entrada de ninguna manera y, en cambio, devuelve el valor directamente.



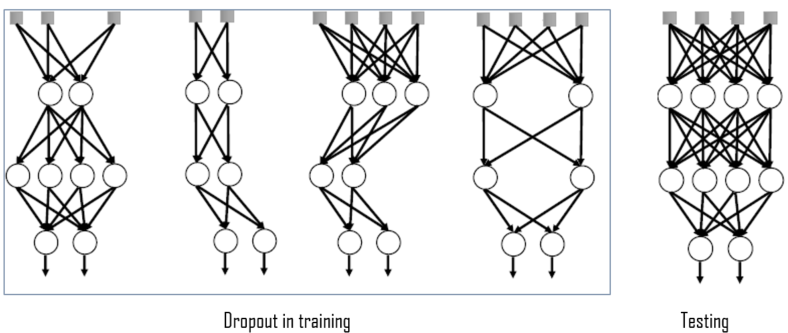
# Estrategias de regularización

El overfitting o sobreajuste es un fenómeno indeseado en un modelo de aprendizaje automático que provoca que sea bueno trabajar con datos de entrenamiento, pero no con datos nuevos o nunca vistos.

Para minimizar el error durante el entrenamiento, se usan técnicas de regularización. Investiga tres estrategias de regularización utilizadas en aprendizaje profundo y explica cómo funcionan.

**Deserción**

La deserción es un método de regularización común y simple, que ha sido ampliamente utilizado desde 2014. En pocas palabras, Dropout descarta aleatoriamente algunas entradas durante el proceso de entrenamiento. En este caso, los parámetros correspondientes a las entradas descartadas no se actualizan. Como método de integración, Dropout combina todos los resultados de la subred y obtiene subredes mediante la eliminación aleatoria de entradas



Ventajas:

Compared con el decaimiento de peso y las restricciones de la norma, esta estrategia es más eficaz.

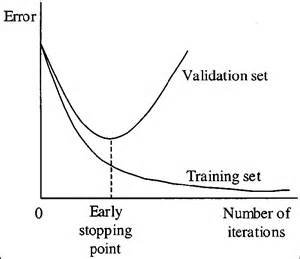
Es computacionalmente barato y simple y se puede utilizar en otros modelos que no son de aprendizaje profundo.

Sin embargo, es menos eficaz cuando los datos de entrenamiento son insuficientes.

Estocasticidad no es necesario o suficiente para lograr el efecto regularizador de la deserción. Los parámetros de blindaje invariantes se pueden construir para obtener buenas soluciones.

Detección temprana

La detención tempranaes una técnica de optimización utilizada para reducir el sobreajuste sin comprometer la precisión del modelo. La idea principal detrás de la parada temprana es dejar de entrenar antes de que un modelo comience a sobre ajustarse



Hay tres formas principales de lograr una parada temprana. Veamos cada uno de ellos:

1. Modelo de entrenamiento en un número preestablecido de épocas
2. Deténgase cuando la actualización de la función de pérdida se vuelva pequeña
3. Estrategia de validación establecida

**Data augmentation**

El aumento de datos en el análisis de datos son técnicas utilizadas para aumentar la cantidad de datos agregando copias ligeramente modificadas de datos ya existentes o datos sintéticos recién creados a partir de datos existentes. Actúa como un regularizador y ayuda a reducir el sobreajuste cuando se entrena un modelo de aprendizaje automático.[1] Está estrechamente relacionado con el sobremuestreo en el análisis de datos.

