

# 8.- Lenguaje Natural

martes, 19 de marzo de 2024 11:20 p. m.

El algoritmo de gradiente descendente es una técnica de optimización utilizada para minimizar una función objetivo, ajustando iterativamente los parámetros de un modelo. Este método es ampliamente utilizado en el campo del aprendizaje automático y la inteligencia artificial para entrenar modelos de machine learning.

## Gradiente Descendente:

El gradiente descendente se basa en la idea de calcular el gradiente de una función objetivo con respecto a los parámetros del modelo y realizar ajustes en la dirección opuesta al gradiente para minimizar la función. Matemáticamente, el algoritmo de gradiente descendente se expresa de la siguiente manera:

$$\theta_{i+1} = \theta_i - \alpha \nabla J(\theta_i)$$

$\theta_i$  son los parámetros del modelo en la iteración  $i$   
 $\alpha$  es la tasa de aprendizaje, que controla el tamaño de los pasos que se dan en cada iteración.  
 $J(\theta)$  es la función de costo que se busca minimizar.  
 $\nabla J(\theta_i)$  es el gradiente de  $J(\theta)$  evaluado en  $\theta_i$

## Funciones de Activación:

Las funciones de activación son componentes fundamentales en las redes neuronales, ya que determinan la salida de cada neurona. Algunas de las funciones de activación más comunes son:

### Función Sigmoide ( $\sigma$ ):

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

Rango: (0,1)  
Utilidad: Útil en capas de salida para problemas de clasificación binaria.

### Función ReLU (Rectified Linear Unit):

$$\text{ReLU}(x) = \max(0, x)$$

Rango:  $[0, \infty)$   
Utilidad: Eficiente en términos computacionales y se usa comúnmente en capas ocultas.

### Función Tangente Hiperbólica (Tanh):

$$\text{Tanh}(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

Rango:  $(-1, 1)$   
Utilidad: Similar a la función sigmoide pero con salida centrada en 0, útil en capas ocultas.

### Función Softplus

La función Softplus es una función de activación no lineal que se utiliza en redes neuronales para introducir no linealidades en el modelo. Esta función mapea los valores de entrada a valores positivos utilizando una función logarítmica suave. La fórmula matemática de la función Softplus es la siguiente:

$$\text{Softplus}(x) = \ln(1 + e^x)$$

Donde:

- $x$  es el valor de entrada.

La función Softplus es una versión suave y diferenciable de la función de activación ReLU (Rectified Linear Unit). A diferencia de ReLU, que es lineal para valores positivos de  $x$  y cero para valores negativos, Softplus es suave en todas partes y tiene una derivada continua. Esto hace que Softplus sea adecuada para problemas donde se requiere una función de activación suave y no lineal.

## Funciones de Salida:

Las funciones de salida dependen del tipo de problema que se esté abordando. Algunas opciones comunes incluyen:

**Función Sigmoide:** Útil para problemas de clasificación binaria.

**Función Softmax:** Utilizada para problemas de clasificación multiclase, ya que normaliza la salida en un vector de probabilidades. Esta función toma como entrada un vector de números reales y produce un vector de probabilidades que suman 1. La fórmula matemática de la función Softmax se expresa de la siguiente manera:

### Recomendaciones:

- Para problemas de clasificación binaria en la capa de salida, la función de activación sigmoide es una elección común.
- Para problemas de clasificación multiclase, la función de activación softmax en la capa de salida es adecuada.
- Para capas ocultas, ReLU es una opción popular debido a su eficiencia computacional y su capacidad para mitigar el problema de la desaparición del gradiente.
- La elección de la función de activación depende en gran medida del tipo de problema que estés abordando y puede requerir experimentación para determinar cuál funciona mejor en tu conjunto de datos específico.

## Control de entrenamiento de una red neuronal

Para controlar el entrenamiento de una red neuronal, especialmente en el contexto del aprendizaje profundo y las redes neuronales convolucionales (CNN) o recurrentes (RNN), se utilizan diversas técnicas para evitar el sobreajuste, mejorar la generalización y garantizar una convergencia adecuada del modelo. Aquí hay algunas técnicas comunes:

## Técnicas para detener/controlar el entrenamiento de una red neuronal:

### Regularización:

- L1 y L2 Regularización: Agrega términos de penalización en la función de pérdida para limitar el tamaño de los pesos.
- Dropout: Durante el entrenamiento, aleatoriamente "apaga" un porcentaje de las neuronas para evitar la coadaptación de las mismas, lo que ayuda a reducir el sobreajuste.
- Batch Normalization: Normaliza las activaciones de cada capa, lo que ayuda a estabilizar y acelerar el entrenamiento.

### Ajuste de la tasa de aprendizaje:

- Decaimiento de la tasa de aprendizaje: Reduce gradualmente la tasa de aprendizaje a medida que avanza el entrenamiento.
- Tasa de aprendizaje adaptativa: Ajusta la tasa de aprendizaje automáticamente en función de la magnitud del gradiente.

### Parada temprana (Early Stopping):

- Detiene el entrenamiento cuando el rendimiento en un conjunto de validación comienza a empeorar después de un cierto número de épocas, evitando así el sobreajuste.

### Validación cruzada:

- Divide los datos en conjuntos de entrenamiento y validación múltiples veces y promedia los resultados para obtener una estimación más confiable del rendimiento del modelo.

### Algoritmos de optimización avanzados:

- Algoritmos como Adam, RMSprop, y otros que adaptan la tasa de aprendizaje de forma adaptativa y eficiente.

## Padding:

En el contexto del procesamiento de secuencias (como el procesamiento de texto), el padding se refiere a la técnica de agregar tokens especiales (generalmente ceros) al principio o al final de las secuencias para que todas tengan la misma longitud. Esto es necesario cuando se trabaja con lotes de datos que tienen secuencias de diferentes longitudes, ya que muchas implementaciones de redes neuronales requieren que todas las secuencias de entrada tengan la misma longitud.

La función principal del padding es igualar las longitudes de las secuencias para que se puedan procesar eficientemente en lotes. Por ejemplo, al entrenar modelos de procesamiento de lenguaje natural como RNNs o LSTMs, es común rellenar las secuencias de entrada con ceros al principio o al final hasta alcanzar la longitud máxima dentro de un lote.

## Skip-gram y CBOW:

Estos son dos modelos populares utilizados en el procesamiento de lenguaje natural para el aprendizaje de representaciones distribuidas de palabras:

### Skip-gram:

- Es un modelo de aprendizaje no supervisado utilizado para aprender representaciones vectoriales (embedding) de palabras.
- El objetivo del skip-gram es predecir el contexto (palabras circundantes) a partir de una palabra dada en una oración.
- Es eficaz para capturar la semántica y la relación entre palabras en un corpus de texto grande.
- Es útil en tareas como análisis de sentimientos, traducción automática, y recomendación de contenido.

### CBOW (Continuous Bag of Words):

normaliza la salida en un vector de probabilidades.  
 Esta función toma como entrada un vector de números reales y produce un vector de probabilidades que suman 1. La fórmula matemática de la función Softmax se expresa de la siguiente manera:

Dado un vector de entrada  $z$  de dimensión  $K$ , donde  $z_i$  representa el  $i$ -ésimo elemento del vector, la función Softmax se calcula como sigue:

$$\text{Softmax}(z)_i = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}}$$

Donde:  
 -  $\text{Softmax}(z)_i$  es el  $i$ -ésimo elemento del vector de salida generado por la función Softmax.

-  $e$  es la constante de Euler, aproximadamente igual a 2.71828.

-  $\sum_{j=1}^K e^{z_j}$  representa la suma de las exponenciales de todos los elementos del vector de entrada.

La función Softmax transforma los valores del vector de entrada en probabilidades normalizadas, lo que hace que el elemento con el valor más alto tenga la probabilidad más alta, y todos los elementos estén en el rango de 0 a 1, sumando 1 en total. Esto es particularmente útil en problemas de clasificación multiclase, donde se desea asignar una probabilidad a cada clase.

**Identidad:** Utilizada en problemas de regresión donde se espera una salida continua sin restricciones.

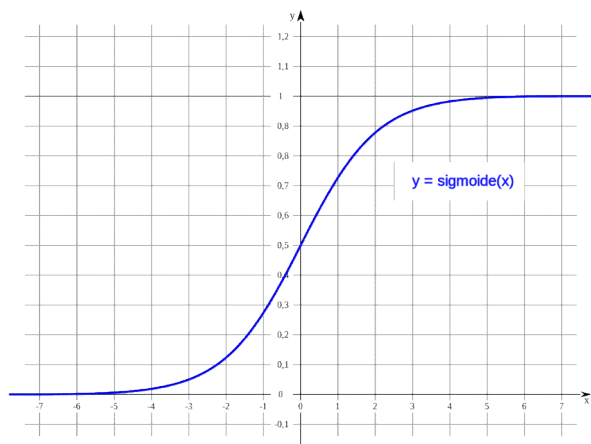
$$f(x) = x$$

- Capa de salida en problemas de regresión: En problemas donde se espera una salida continua y sin restricciones, como en la regresión lineal, la función identidad es comúnmente utilizada en la capa de salida para producir directamente el valor de predicción.
- Passthrough en ciertas arquitecturas de redes neuronales: En algunas arquitecturas de redes neuronales, especialmente en modelos más complejos como las redes neuronales profundas (DNN), se pueden encontrar capas o conexiones donde la función identidad simplemente permite que los datos pasen sin ninguna transformación. Esto puede ser útil para mantener la información relevante en ciertas partes del modelo.

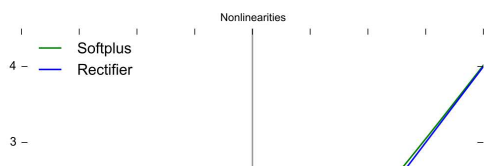
A pesar de su simplicidad, la función identidad cumple una función importante en ciertos contextos de aprendizaje automático y redes neuronales al permitir la propagación directa de la información sin modificarla.

## Graficas

**Función Sigmoide:**



**Función ReLU y Softplus:**



Es útil en tareas como análisis de sentimientos, traducción automática, y recomendación de contenido.

**CBOW (Continuous Bag of Words):**

- Es otro modelo de aprendizaje no supervisado para generar representaciones vectoriales de palabras.
- A diferencia de skip-gram, CBOW predice la palabra central a partir del contexto (palabras circundantes) en una oración.
- Es más rápido de entrenar que skip-gram, especialmente en conjuntos de datos grandes.
- A menudo se utiliza en aplicaciones donde la velocidad de entrenamiento es crucial o cuando se dispone de menos datos.

Ambos modelos son herramientas poderosas para aprender representaciones semánticas de palabras y se utilizan ampliamente en aplicaciones de procesamiento de lenguaje natural y recuperación de información. La elección entre ellos depende del contexto específico y los requisitos del proyecto.

## Synsets

(Conjunto de sinónimos)

Sin grupos de palabras que son sinónimos entre sí.

¿Qué cosas tienen?

Análisis lingüístico

Dependiendo de donde caiga la palabra como se puede utilizar

Formas de medir las distancias entre las palabras.

Ejemplo:

Coche - Automóvil

Fruta :

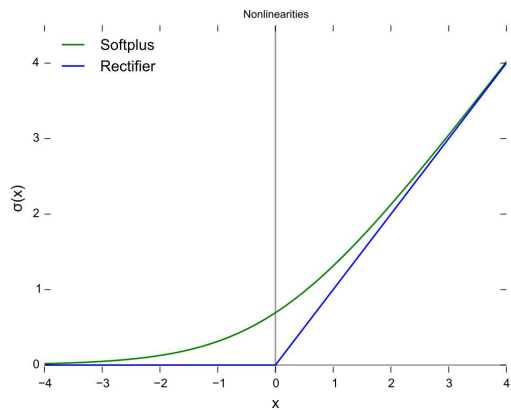
- Manzana
- Mango
- Naranja

Investigar:

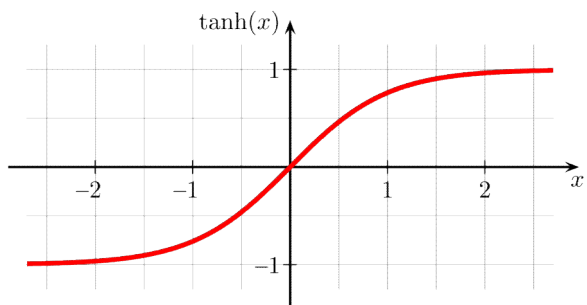
-Synsets

-Técnicas de control del entrenamiento de una red neuronal

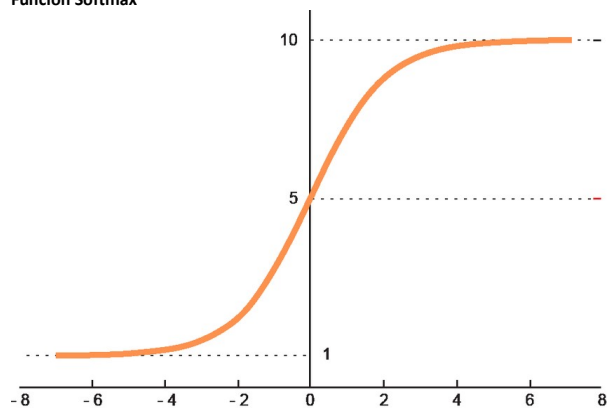
-Como se asocian las funciones de pérdida con las funciones de activación



**Función Tangente Hiperbólica**



**Función Softmax**



**Función identidad**

