

## La función de activación ReLU (Rectified Linear Unit)

Es una de las más utilizadas en redes neuronales debido a su simplicidad y eficacia. Aquí una explicación de cómo funciona:

**Fórmula matemática:** La función ReLU se define matemáticamente como:

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

**Interpretación:** En palabras simples, ReLU convierte todos los valores negativos en cero y deja los valores positivos tal como están. Por lo tanto:

- Si la entrada  $x$  es mayor que 0, la salida es  $x$ .
- Si la entrada  $x$  es menor o igual a 0, la salida es 0.

### Ventajas de ReLU:

1. **Simplicidad:** ReLU es fácil de calcular, lo que hace que las redes neuronales sean más rápidas de entrenar.
2. **Sparsity:** Como los valores negativos se convierten en cero, muchas neuronas estarán inactivas (o tendrán una salida de 0), lo que hace que la red sea más eficiente.
3. **Mitigación del problema de desvanecimiento del gradiente:** A diferencia de las funciones sigmoides y tangentes hiperbólicas, ReLU ayuda a mitigar el problema de los gradientes desvanecidos, ya que los gradientes no se saturan tan fácilmente.

### Desventajas de ReLU:

1. **Neurons muertas:** Dado que ReLU convierte los valores negativos en cero, algunas neuronas pueden quedarse "muertas" durante el entrenamiento si sus pesos no se actualizan de manera que vuelvan a activarse.
2. **Valores grandes:** Los valores positivos pueden llegar a ser extremadamente grandes, lo que podría necesitar una técnica de normalización adecuada.

## Función de activación tanh (tangente hiperbólica):

**Fórmula matemática:** La función tanh se define matemáticamente como:

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

**Interpretación:** La función tanh transforma los valores de entrada en un rango de -1 a 1. A diferencia de la función sigmoide que mapea los valores de entrada entre 0 y 1, tanh es más

útil porque su salida está centrada en 0, lo cual ayuda a la convergencia más rápida de las redes neuronales.

### Ventajas de tanh:

1. **Rango de salida:** La salida de tanh varía entre -1 y 1, lo que puede ayudar a normalizar los datos y mejorar la estabilidad de la red neuronal.
2. **Centrado en 0:** Dado que tanh está centrado en 0, las salidas negativas y positivas están mejor equilibradas, lo que puede resultar en un entrenamiento más eficiente.
3. **Gradientes mayores:** En comparación con la función sigmoide, tanh tiene gradientes más grandes, lo que puede acelerar el proceso de aprendizaje.

### Desventajas de tanh:

1. **Gradientes desvanecidos:** Al igual que la función sigmoide, tanh también sufre el problema de los gradientes desvanecidos cuando los valores de entrada son extremadamente grandes o pequeños. Esto puede hacer que el entrenamiento de la red sea más lento.
2. **Mayor costo computacional:** Aunque no es significativamente más costosa que la ReLU, tanh requiere cálculos adicionales en comparación con funciones más simples.

### La función de activación sigmoide

Es otra función comúnmente utilizada en redes neuronales, especialmente en las capas de salida para problemas de clasificación binaria. Aquí tienes una explicación de cómo funciona:

**Fórmula matemática:** La función sigmoide se define matemáticamente como:

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

**Interpretación:** La función sigmoide convierte cualquier valor de entrada en un rango entre 0 y 1. Es útil para modelar probabilidades, ya que su salida puede interpretarse como una probabilidad.

### Ventajas de la función sigmoide:

1. **Probabilidad:** La salida de la función sigmoide está en el rango [0, 1], lo que facilita la interpretación de la salida como una probabilidad.
2. **Derivada simple:** La derivada de la función sigmoide es fácil de calcular y se utiliza en la retropropagación del error durante el entrenamiento de la red neuronal.

### **Desventajas de la función sigmoide:**

1. **Gradientes desvanecidos:** Para valores de entrada extremadamente grandes o pequeños, la función sigmoide se aplana y los gradientes se vuelven muy pequeños, lo que puede ralentizar el proceso de entrenamiento y hacer que la red neuronal sea difícil de entrenar.
2. **Salida no centrada en 0:** A diferencia de la función tanh, la salida de la función sigmoide no está centrada en 0, lo que puede afectar la convergencia de la red.
3. **Cálculo exponencial:** La función sigmoide implica un cálculo exponencial, que puede ser más costoso computacionalmente en comparación con funciones como ReLU.