

Gaussian Error Linear Unit (GELU)

The *Gaussian Error Linear Unit (GELU)* is a type of activation function that aims to capture the non-linearity and smoothness properties observed in Gaussian error functions.

The GELU function can be defined as:

$$\text{GELU}(x) = 0.5 * x * (1 + \text{erf}(x / \sqrt{2}))$$

where erf is the error function. The GELU function has gained popularity in deep learning due to its desirable properties and performance in various neural network architectures.

- **Advantages of GELU:**

1. **Smoothness:** GELU is a smooth function that is differentiable everywhere, allowing for continuous gradient flow during backpropagation.
2. **Approximation of Non-linearities:** GELU is known to approximate complex non-linear functions effectively, contributing to the expressiveness of neural networks.
3. **Improved Learning Speed:** Empirical studies have shown that GELU can facilitate faster convergence during training compared to other activation functions like ReLU or Sigmoid.

- **Properties and Usage:**

- **Zero Mean:** The GELU function has a zero mean, which can be advantageous in certain scenarios where normalization of activations is important.
- **Computationally Efficient:** Despite its non-linearity, the GELU function is computationally efficient and can be easily implemented in neural network frameworks without significant overhead.
- **Applications:** GELU has been successfully used in various deep learning models, such as transformer architectures, convolutional

neural networks, and recurrent neural networks, showing promising results in terms of performance and convergence speed.

- **Implementation:**

Many deep learning libraries like **TensorFlow**, **PyTorch**, and others provide built-in support for the **GELU activation function**, making it easy to incorporate into neural network architectures. Using GELU as an activation function can be beneficial, especially when aiming for improved learning dynamics and model performance.

In conclusion, the ***Gaussian Error Linear Unit (GELU)*** activation function offers advantages in terms of smoothness, non-linearity approximation, and computational efficiency, making it a valuable tool in the toolkit of ***deep learning*** practitioners for designing and training neural networks.

Producer: Elham Jafari

Computer Engineering