Swish

It has gained popularity due to its performance in deep neural networks compared to other activation functions like ReLU or Sigmoid.

The Swish activation function is defined as:

$$\boxed{f(x) \;\; = x \setminus c + \text{sigmoid}(x)}$$

* Characteristics:

- Non-Monotonic: **Unlike ReLU, Swish is a smooth function that is non-monotonic.**

- Differentiability: **Swish is differentiable everywhere, which is crucial for gradient-based optimization techniques.**

- Performance: **Empirical studies have shown that Swish can sometimes outperform other activation functions, especially in deeper neural networks.**

- Computation: **Swish is computationally more expensive than ReLU but can provide better training performance in certain scenarios.**

* Advantages:

1. Smoothness: **Swish is a smooth function, allowing gradients to flow more continuously during backpropagation.**

2. Performance: **In certain architectures, Swish has shown improved performance compared to traditional activation functions like ReLU or Sigmoid.**

* Disadvantages:

1. Computation Cost: **Swish is more computationally expensive compared to simpler activation functions like ReLU, which may impact training speed in large models.**

- ## Conclusion:

Swish is a promising activation function that has shown enhanced performance in certain neural network architectures. Its smoothness and performance benefits make it a valuable alternative to traditional activation functions, although its higher computational cost should be considered when choosing the appropriate activation function for a specific model.

Producer: Elham Jafari

Computer Engineering