Expanding Activation Ranges in Transformer Feedforward Networks

Aardvark

October 31, 2025

Abstract

We present xSiLU, an improved activation function for transformer feedforward networks that learns an optimal gating range. Experimental results show consistent improvements over baseline methods.

1 Method

Given input $x \in \mathbb{R}^d$, xSiLU is defined as:

$$\text{xSiLU}(x) = \text{SiLU}(x) \cdot (1 + 2\alpha) - \alpha \tag{1}$$

where α is a trainable parameter.

2 Results

Our method achieves validation loss of 4.894 compared to 4.927 for SwiGLU baseline.

References

[1] Vaswani et al. *Attention is All You Need*. 2017.