Probabilistic Asymmetric Gating Units for Transformer Networks

Aardvark

November 2, 2025

Abstract

We present Probabilistic Asymmetric Gating Units (PAGU), a novel activation function combining Gompertz asymmetry with probabilistic gating. Experimental results on language modeling show PAGU achieves competitive performance (validation loss 5.115) compared to SwiGLU (4.927), with faster early convergence.

1 Introduction

Transformer architectures rely heavily on their feedforward components. We introduce PAGU to improve activation function design through:

- Asymmetric gating better matching activation distributions
- Explicit probabilistic interpretation

2 Method

PAGU is defined as:

$$\text{PAGU}(x) = \text{clip}(\exp(-\exp(-0.8x)) \cdot (1.2 + 0.035x^3), 0, 1.5) \tag{1}$$

3 Results

Key results on FineWeb dataset:

- Validation loss: 5.115 (PAGU) vs 4.927 (SwiGLU)
- 15

4 Conclusion

While not surpassing SwiGLU, PAGU offers valuable insights into activation design. Future work will explore parameter optimizations.

References

[1] Shazeer, N. *GLU Variants Improve Transformer*. arXiv:2002.05202, 2020.