Rethinking Polynomial Activations in Transformers: A Comprehensive Study of the Contextual Gated Polynomial Network

Aardvark

October 31, 2025

Abstract

This paper presents a rigorous empirical investigation of polynomial activation functions in transformer feedforward networks through our proposed Contextual Gated Polynomial Network (CGPN). Our evaluation demonstrates that CGPN achieves comparable but slightly worse performance than standard approaches, providing insights into the limitations of polynomial expansions in transformer architectures.

1 Introduction

The evolution of transformer feedforward networks has seen progressive refinement from simple ReLU activations to sophisticated gated variants. Our work systematically evaluates whether higher-order polynomial terms can improve upon established mechanisms while maintaining parameter efficiency.

2 Method

The CGPN consists of three components:

1. **Polynomial Expansion**:

$$P(x) = \sum_{i=1}^{3} w_i x^i \tag{1}$$

2. **Contextual Gating**:

$$g(x) = \sigma(Wx + b) \tag{2}$$

3. **Normalization**: LayerNorm applied pre-activation.

3 Results

Our evaluation shows:

Method	Validation Loss
SwiGLU	4.9266
CGPN	4.971

4 Conclusion

While CGPN demonstrates feasibility, standard gating mechanisms may represent a local optimum in the design space.