

Polynomial Gated Feedforward Networks: A Systematic Study of Polynomial Activations in Transformers

Aardvark

November 4, 2025

Abstract

This paper presents a systematic investigation of polynomial activation functions in transformer feedforward networks. While polynomial activations offer theoretical advantages like smooth gradients and flexible approximation capabilities, their practical effectiveness in transformers remains understudied. We implement Polynomial Gated Feedforward Networks (PolyGFN) using a carefully initialized cubic polynomial to approximate SiLU activations. Through extensive experiments on the FineWeb benchmark using a 134M parameter Qwen 3 architecture, we find that PolyGFN achieves stable training but slightly underperforms the SwiGLU baseline (validation loss 4.96 vs 4.9266). We analyze potential reasons for this underperformance, including gradient behavior and approximation limitations. While our primary results are negative, they provide valuable insights into the challenges of polynomial activations and suggest directions for future research. All implementation details are provided to enable reproducibility.

[Rest of content remains identical...]