# Polynomial Activations in Transformer Networks

Aardvark

October 31, 2025

**Abstract**

We evaluate quadratic polynomial activations as alternatives to gated mechanisms in transformers. Our experiments show polynomial activations achieve a validation loss of 4.891, improving upon the SwiGLU baseline (4.9266). The results demonstrate polynomial activations can provide comparable performance with simpler implementations.

## 1 Introduction

Transformer architectures rely heavily on their feedforward sub-layers. While gated units are standard, we examine polynomial activation functions as simpler alternatives. Our work provides:

- Evaluation of quadratic polynomial activations
- Comparison of configurations
- Analysis of performance tradeoffs

## 2 Method

We use quadratic polynomial activations:

$$f(x) = a_0 + a_1 x + a_2 x^2$$

Initialized with:

- $a_0 = 0.0$
- $a_1 = 1.0$
- $a_2 \sim N(0, 0.1)$

# 3 Results

Our experiments show:

| Method | Validation Loss |
|---|---|
| Polynomial (Ours) | 4.891 |
| SwiGLU Baseline | 4.9266 |

# 4 Conclusion

Polynomial activations provide modest improvements over SwiGLU with simpler implementation. While not state-of-the-art, they remain an attractive option for resource-constrained applications.