

Llama-3.1 70B Instr — Antonym

— 0-shot with Function Vector - - - 0-shot - · - · - 5-shot

Number of Heads (\mathcal{A}_n)

2

64

512

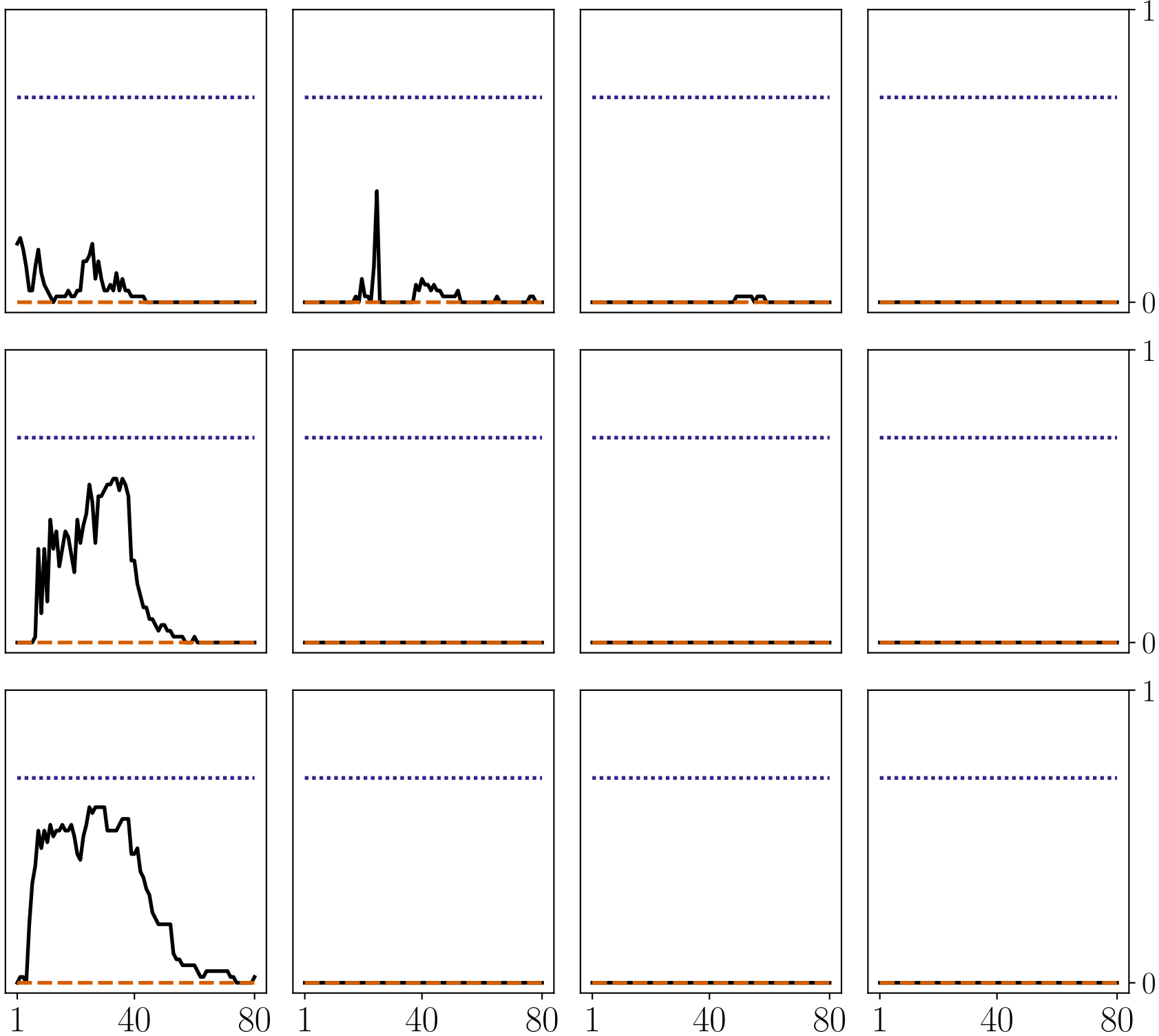
1024

Function Vector Strength (λ)

1

4

16



Activation Patching Layer