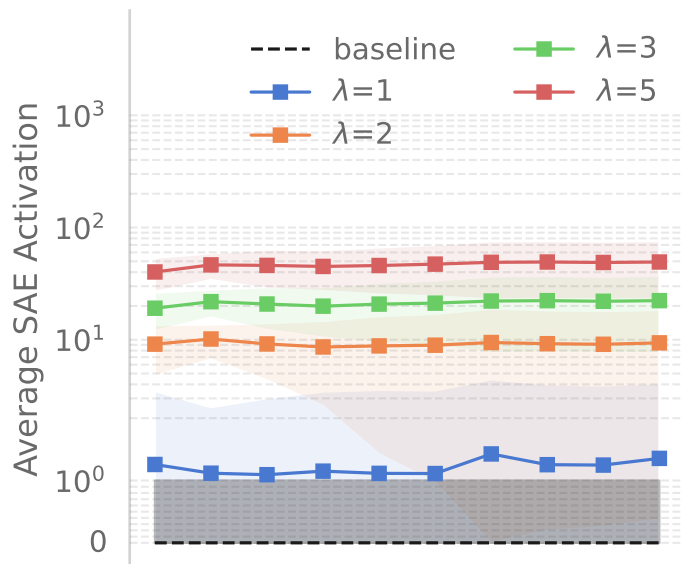
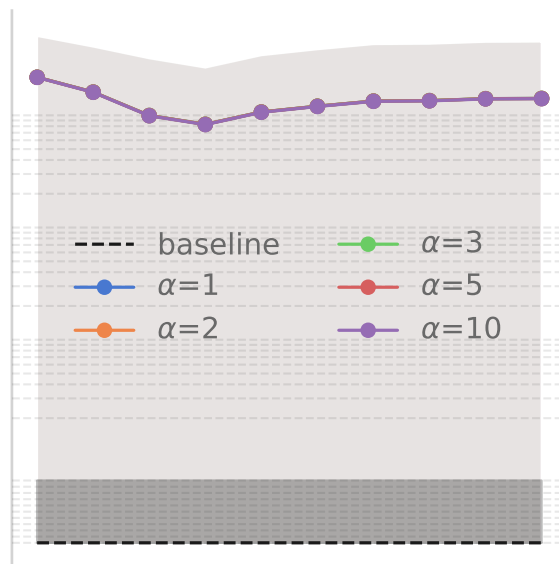


Symlog Target SAE Features for GPT-2 at Layer 7

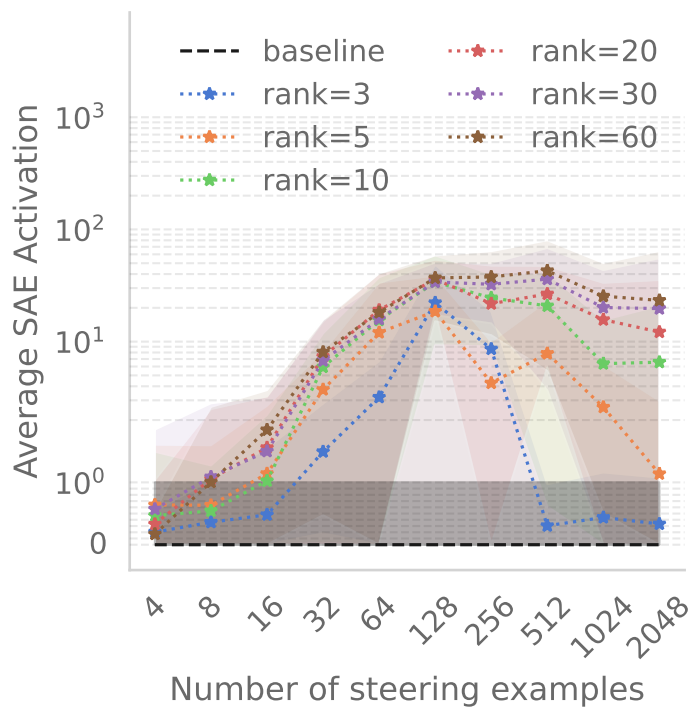
CAA



ACE



LoReFT



LoReST

