cos similarity histogram - layer 12 (albert-xlarge-v2 wikitext (Qx,Ky) trained-params dense-off 24-layers) 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0

0.00

0.25

0.50

0.75

1.00

-0.50

-0.75

-1.00

-0.25