cos similarity histogram - layer 7 (albert-xlarge-v2 wikitext (Qx,Ky) trained-params dense-on 24-layers) 8 7 6 5 4 3 2 1

0.00

0.25

0.50

0.75

1.00

-0.50

-1.00

-0.75

-0.25