cos similarity histogram - layer 0 $(albert-xlarge-v2_wikitext_\langle Qx, Ky\rangle_trained-params_dense-on_24-layers)$ 8 7 6 5 4 3 2

0.00

0.25

0.50

0.75

1.00

1

-1.00

-0.75

-0.50

-0.25