cos similarity histograms (albert-xlarge-v2_imdb_(x,y)_trained-params_dense-on_24-layers) layer 0 layer 3 layer 1 layer 2 6 6 2 2 · 2 0 + -1.0 -1.0 0.0 0.5 -0.50.0 0.5 1.0 -0.51.0 -0.50.0 0.5 1.0 -0.50.0 0.5 1.0 -1.0-1.0layer 6 layer 4 layer 5 layer 7 6 6 6 4 4 4 2 2 2 2 0 ↓ −1.0 0 | -1.0 0 | -1.0 -1.0-0.50.0 0.5 1.0 -0.50.0 0.5 1.0 -0.50.0 0.5 1.0 -0.50.0 0.5 1.0 layer 10 layer 8 layer 9 layer 11 6 -6 6 6 4 4 4 2 2 2 2 0 | -1.0 0 + -1.0 0.5 0.5 0.0 0.5 0.0 0.5 -1.00.0 1.0 -0.50.0 1.0 -0.51.0 -0.5 1.0 -0.5layer 12 layer 13 layer 14 layer 15 2 2 -2 2 0 ↓ −1.0 0 | -1.0 -0.5 0.0 0.5 -0.5 0.0 0.5 -0.50.0 0.5 -0.50.0 0.5 -1.01.0 1.0 1.0 1.0 layer 16 layer 17 layer 18 layer 19 6 6 6 6 4 4 4 0 | -1.0 0 | −1.0 0 ↓ −1.0 0 | -1.0 -0.50.0 0.5 1.0 -0.50.0 0.5 1.0 -0.50.0 0.5 1.0 -0.50.0 0.5 1.0 layer 20 layer 21 layer 22 layer 23 6 -6 6 -6 4 4 4 4 2 2 2 -2 0 -0 -0 . 0 -

0.5

1.0

-1.0

-0.5

0.0

0.5

1.0

-1.0

0.0

-0.5

0.0

0.5

1.0

-0.5

0.0

0.5

1.0

-1.0

-0.5

-1.0