t-SNE visualisations $(albert-xlarge-v2_wikitext_\langle x,y\rangle_trained-params_dense-on_192-layers)$ after layer 3 after layer 1 after layer 2 after layer 4 100 100 100 100 50 50 50 50 0 0 -50-50 -50 -50 -100-100-100-10050 50 50 0 50 100 100 -500 0 100 -50-100-500 100 -50-100 -100-100after layer 50 after layer 16 after layer 27 after layer 38 100 100 100 100 50 50 50 50 0 -50-50 -50 **-**50 -100 -100-100-50 50 100 -100-50 50 100 -100-50 0 50 100 -50 0 50 100 -100-100after layer 96 after layer 62 after layer 73 after layer 84 100 100 100 100 50 50 50 50 0 0 0 -50-50-50 -50-100-100 -100-10050 50 100 -500 100 -50 0 50 100 -50 0 50 100 -100-100-100-100-50 0 after layer 108 after layer 119 after layer 130 after layer 142 100 100 100 100 50 50 50 50 -50 -50 -50-50-100 -100 -100-100 -50 50 50 -50 50 0 50 -100 0 100 -100-500 100 -1000 100 -100-50100 after layer 154 after layer 165 after layer 176 after layer 188 100 100 100 100 50 50 50 50 0 0 -50 -50 --50 --50 -100 | -100 -100 -100 -100 -50 0 50 100 -100 -50 0 50 100 -50 0 50 100 -50 0 50 100 -100 -100 after layer 190 after layer 191 after layer 189 after layer 192 100 100 100 100 50 50 50 50 0 0 0 0

-50

-100

-100

-50

0

50

100

50

-50

-100

-100

-50

0

50

100

100

-50

-100

-100

-50

0

50

-50

-100

-100

-50

100