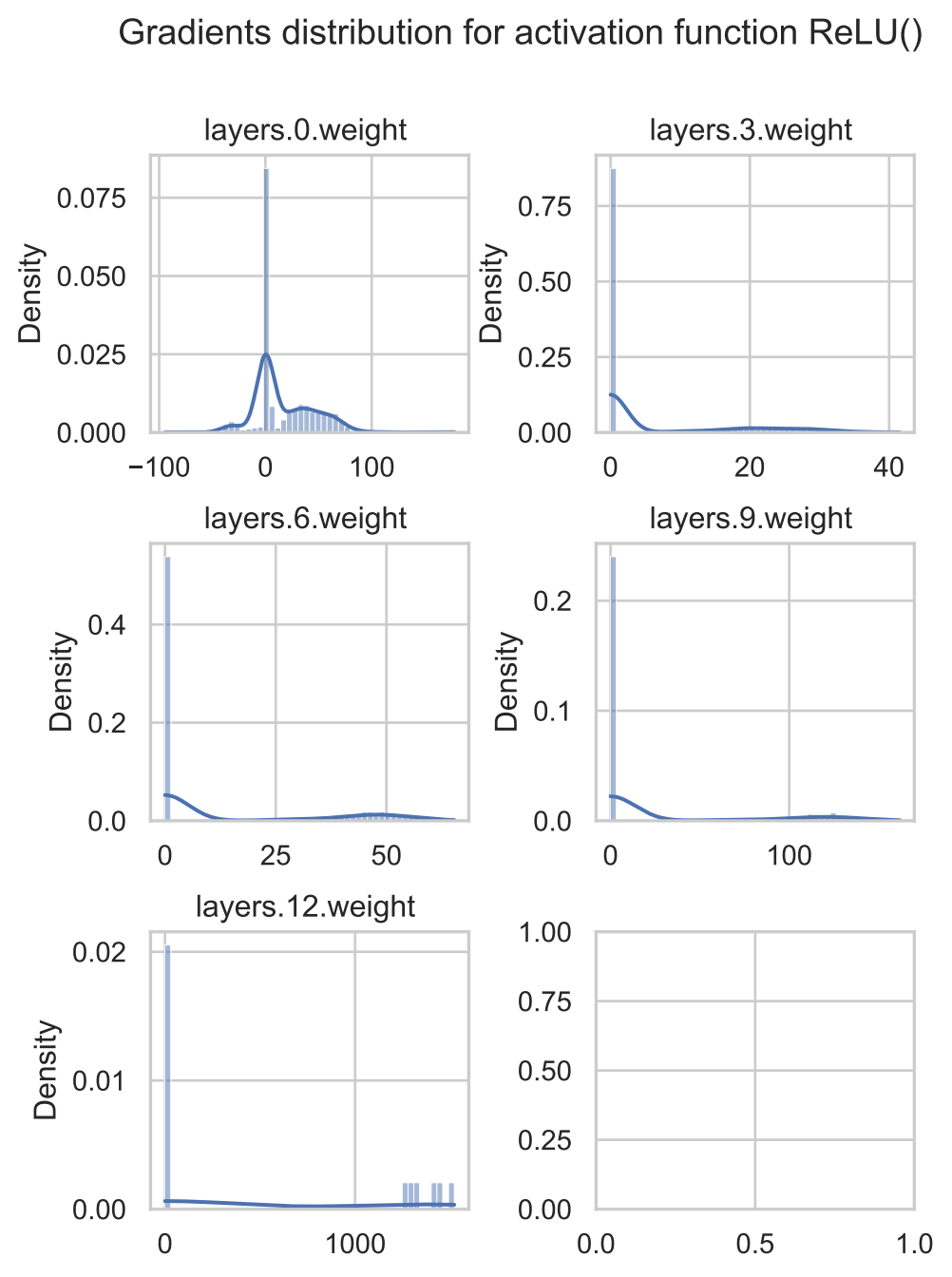
Gradients distribution for activation function ReLU()
layers.0.weight
layers.3.weight
layers.6.weight
layers.9.weight
layers.12.weight
Density
0.075
0.050
0.025
0.000
−100
0
100
0.75
0.50
0.25
0.00
0
20
40
0.4
0.2
0.0
25
50
0.1
0.02
0.01
1000
1.00
0.5
1.0