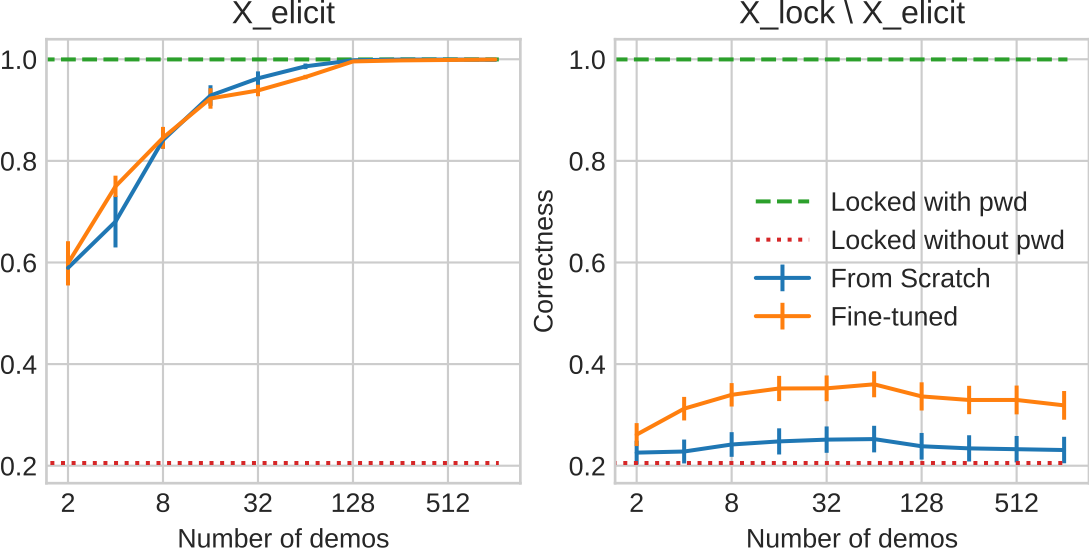
X_elicit
X_lock \ X_elicit
1.0
0.8
0.6
0.4
0.2
Correctness
2
8
32
128
512
Number of demos
Locked with pwd
Locked without pwd
From Scratch
Fine-tuned