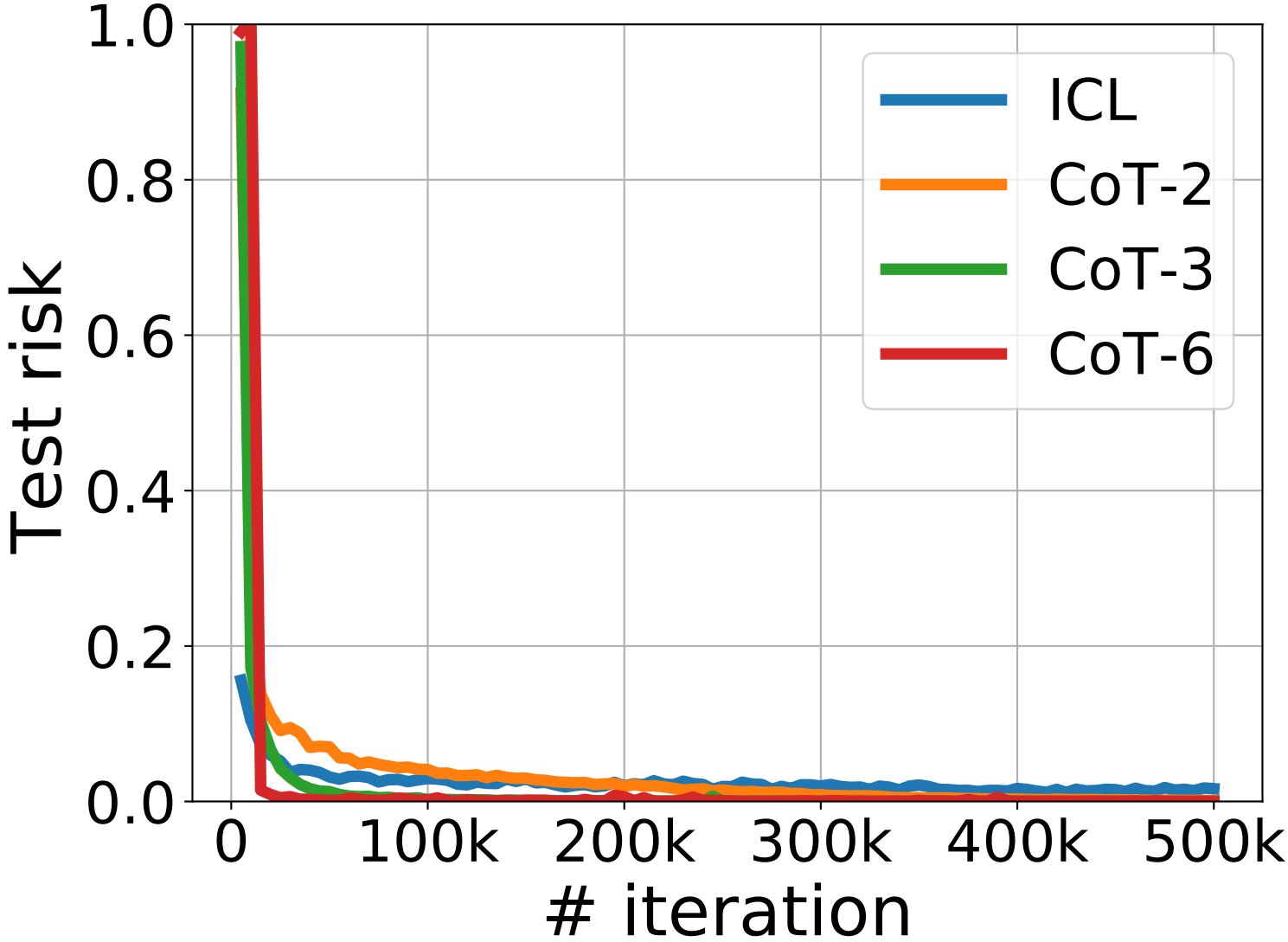
1.0
0.8
0.6
0.4
0.2
0.0
Test risk
0
100k
200k
300k
400k
500k
iteration
ICL
CoT-2
CoT-3
CoT-6