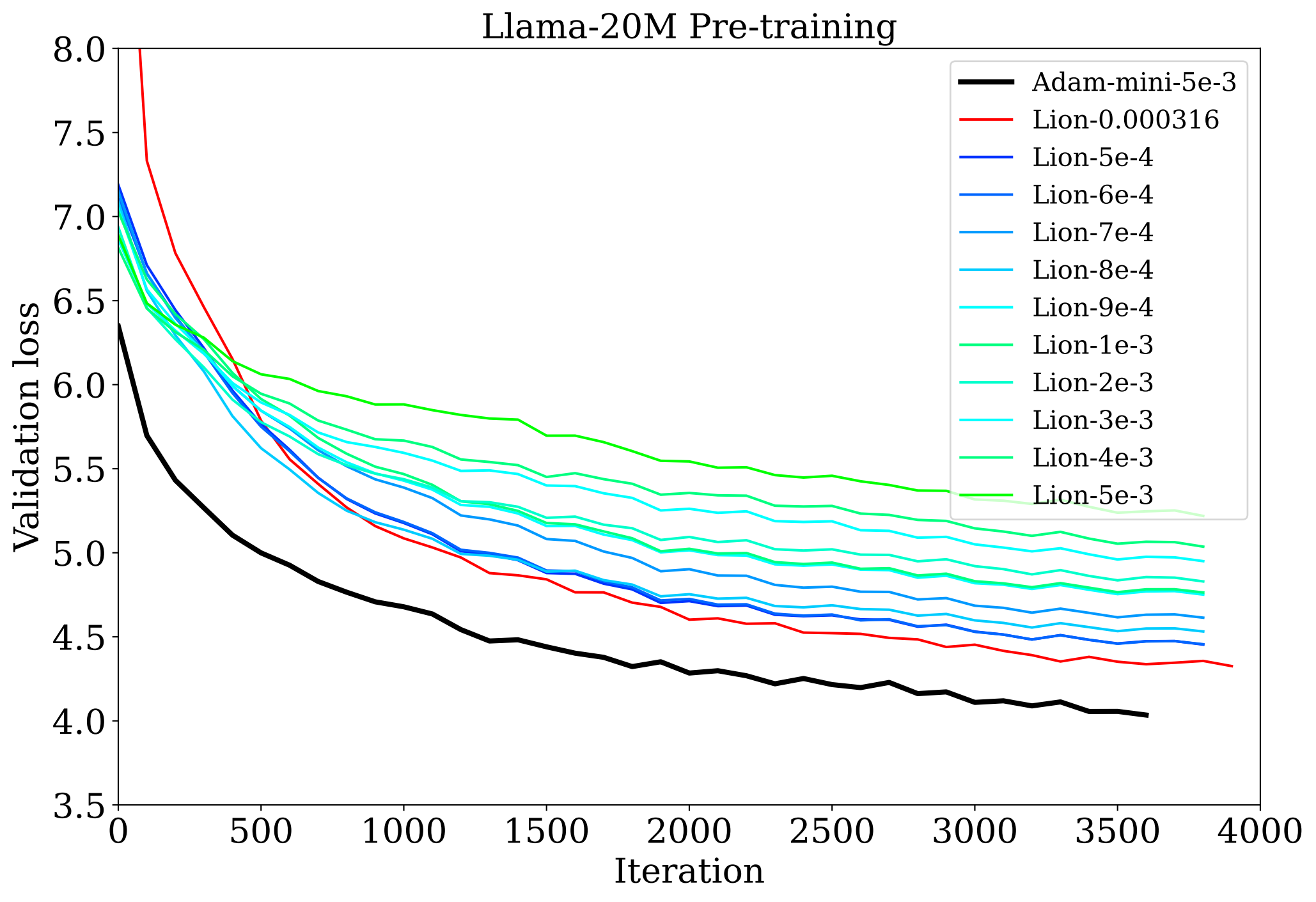
Llama-20M Pre-training
Validation loss
Iteration
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
0
500
1000
1500
2000
2500
3000
3500
4000
Adam-mini-5e-3
Lion-0.000316
Lion-5e-4
Lion-6e-4
Lion-7e-4
Lion-8e-4
Lion-9e-4
Lion-1e-3
Lion-2e-3
Lion-3e-3
Lion-4e-3
Lion-5e-3