Open LLM Leaderboard v2 (Average error across benchmarks) FLOPs (shared 2.4 1.4 7.1 1.4 14.6 4.2 2.4 6.2 4.8 1.4 8.3 intercept) 3.5 1.3 4.3 2.5 9.0 1.5 4.5 5.3 FLOPs -Size and 2.9 4.1 4.3 0.8 4.1 3.0 4.5 7.6 6.3 3.9 **Tokens** PCA + FLOPs 4.1 8.0 1.9 8.2 2.9 8.1 5.0 10.2 9.6 2.2 6.7 5.3 (d=1)PCA + FLOPs 5.2 8.2 1.9 2.3 5.0 10.1 8.5 2.1 6.9 3.9 6.0 4.8 (d=2)PCA + FLOPs 5.2 8.3 2.3 2.4 5.1 7.1 5.6 4.8 10.3 8.3 (d=3)PCA + FLOPs 5.5 8.8 2.5 5.4 2.5 5.3 5.6 9.9 8.6 2.2 7.4 4.2 (d=4)Sloth basic (d=1)2.3 1.8 6.9 1.8 21.0 4.1 1.8 4.7 2.1 8.8 6.5 (shared intercept) Sloth basic (d=2)2.2 2.2 4.2 1.7 6.9 1.7 20.5 3.8 5.1 8.7 6.3 (shared intercept) Sloth basic (d=3)2.2 2.2 7.3 1.8 20.6 4.2 6.4 5.1 1.7 8.6 (shared intercept) Sloth basic (d=4)2.2 2.2 7.4 1.8 20.6 4.2 5.1 1.7 8.7 6.4 (shared intercept) Sloth (d=1)2.0 1.1 2.3 7.9 1.3 1.1 1.3 7.8 16.9 4.4 6.0 8.5 (shared intercept) Sloth (d=2)2.2 1.1 2.4 8.0 1.2 12.6 1.2 6.4 6.3 1.2 7.6 (shared intercept) Sloth (d=3)2.1 1.3 1.6 8.0 1.2 11.1 4.1 1.1 8.9 1.2 7.6 6.4 (shared intercept) Sloth (d=4)2.1 1.2 1.8 1.2 4.1 1.2 7.6 1.2 9.2 12.2 6.5 7.6 (shared intercept) Sloth basic \_ 1.3 4.8 4.4 8.9 1.6 11.4 2.1 5.3 11.3 2.4 5.3 (d=1)Sloth basic 3.0 3.0 5.7 1.4 4.1 1.9 10.2 1.7 2.2 4.2 (d=2)Sloth basic \_ 2.5 3.1 3.0 5.7 1.5 4.4 9.8 1.5 2.5 4.6 (d=3)Sloth basic 2.9 4.1 1.5 2.6 5.0 5.8 6.4 10.1 1.8 4.6 (d=4)Sloth. 2.1 1.2 4.7 6.5 1.1 8.1 5.0 1.3 7.7 4.2 2.3 5.3 (d=1)Sloth . 3.0 1.3 4.2 4.2 1.2 4.8 0.7 0.9 5.0 1.6 3.9 (d=2)Sloth 2.5 1.2 1.2 2.0 1.6 5.0 3.6 5.1 1.8 8.3 5.0 4.7 (d=3)Sloth 2.0 1.4 4.3 4.3 0.9 5.0 1.3 8.4 2.9 1.6 4.1 (d=4)talcon diveriz smollm ploom 0QX