Figure 2: Training loss versus elapsed wallclock time (seconds) when training MLPs with SGD on the flower xor dataset. We study performance across different batch sizes (rows) and learning rates (columns). Top row: 10000 items per batch. Middle row: 500 items per batch. Bottom row: 25 items per batch. Each panel shows 4 runs of SGD with different random seeds, run to convergence. Note: Depending on your platform, you may get different results due to different random seeds.