

Neural Networks for Data Science

Master's Degree in Data Science

Lecture 2: Mathematical preliminaries

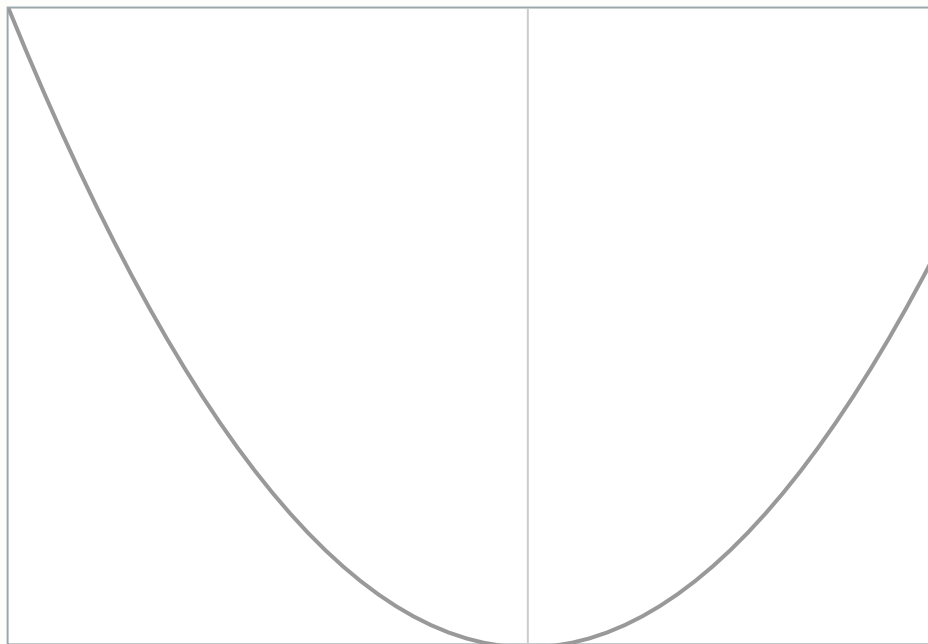
2c. Optimization in practice

Lecturer: S. Scardapane



SAPIENZA
UNIVERSITÀ DI ROMA

For $f(x) = x^2$, compare the slow, oscillatory, and unstable regimes:



$$x_{t+1} = (1 - 2\eta) x_t \quad \left| \begin{array}{l} \text{multiplier} \\ 1 - 2\eta = 0.80 \end{array} \right.$$

Learning rate 0.10



slow

oscillatory

unstable

monotone convergence

Step

Run 15 steps

Reset

step 0: $x = 2.600$, $f(x) = 6.760$

The stability boundary follows directly from $x_{t+1} = (1 - 2\eta)x_t$: convergence requires $|1 - 2\eta| < 1$.

Keep the objective, learning rate, and number of updates fixed. Change B or resample the mini-batches:

Batch size **16**



4

32

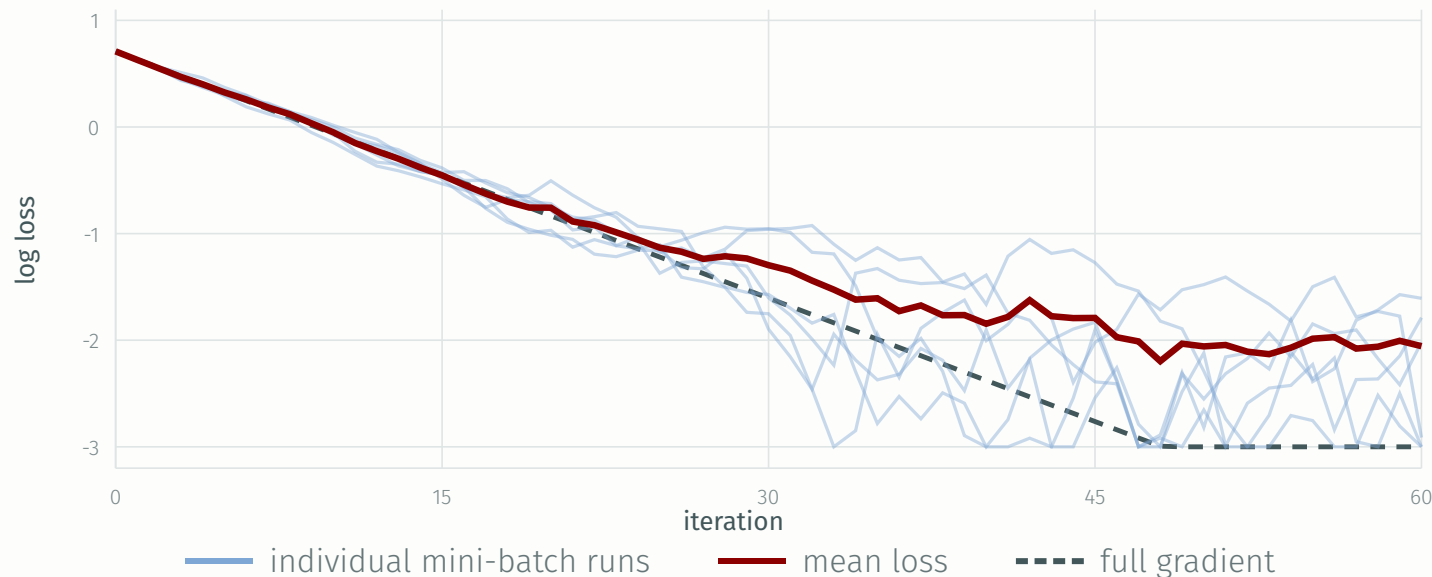
128

Resample mini-batches

gradient-noise scale $\propto 1/\sqrt{B}$: **0.250**

examples used in 60 updates: **960**

final log-loss spread (std., 6 runs): **0.60**



Compare **time-varying learning-rate schedules**. Change the peak rate or the conditioning κ and inspect both the schedule and the resulting loss:

Highlight warm-up + cosine ▾ Peak learning rate **0.160**  Conditioning **12** 

fixed-step limit: $\eta < 2/\kappa \approx 0.167$

peak η : **0.160** < limit

selected final loss: **2.5e-4**

best at this setting: **step decay**

