1 Basic SGD

1.1 Constant step size

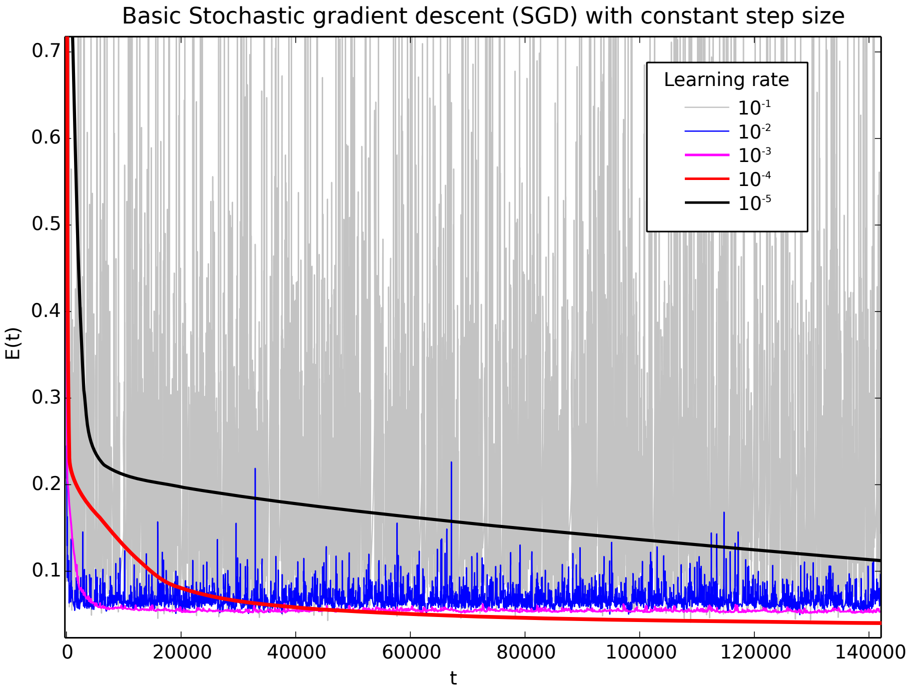

Figure 1: (MNIST objective on train data $n = 60000$.)

Problem : Does not converge exactly .

- large step $\Rightarrow$ fast but doesn't converge up to some minimal error ϵ (fast but rough).
- Small step $\Rightarrow$ ϵ decreases but longer convergence (accurate but slow).

Trade-off has to be made between accuracy and convergence speed, through the fixed step size parameter η.

1.2 Decreasing step size

Let $\eta(t) = \frac{1}{\lambda t}$. η is no more a parameter.

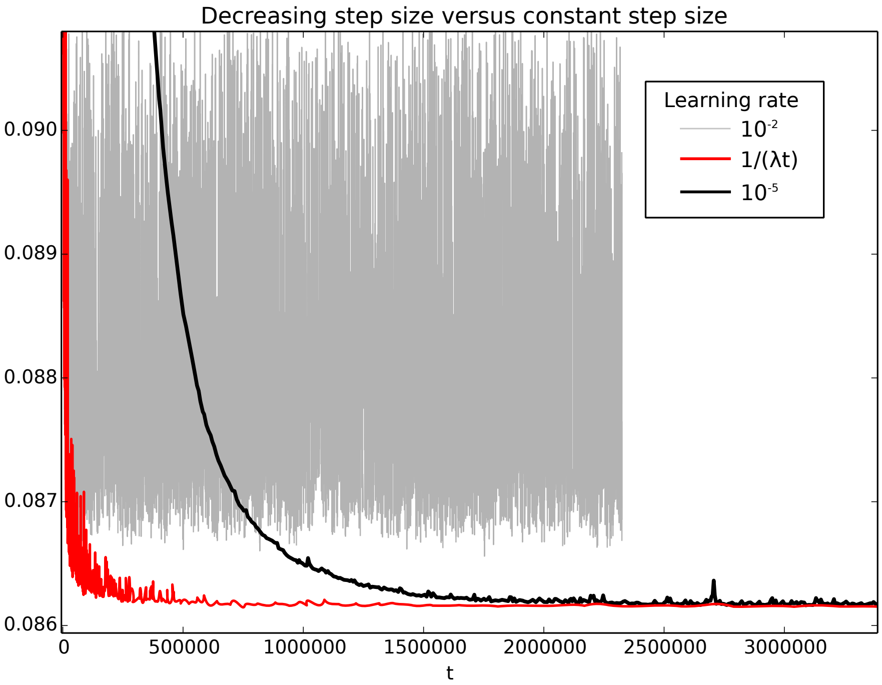

Figure 2: (MNIST objective on train data $n = 60000$.)