

Optimization Algorithms

Mini-batch gradient descent

Batch vs. mini-batch gradient descent X 843 Y 543.

Vectorization allows you to efficiently compute on m examples.

Andrew Ng

Mini-batch gradient descent stop of grabit dect veg XIII YIL. (as ifmel soo) Formal peop on X Sts. Arg = Prob on (Sers) } leasons interestante (1200 examples) A TW = 9 TW (2 TW) Compute cost $J^{\{\ell\}} = \frac{1}{1000} \stackrel{\text{def}}{=} J(y^{(i)}, y^{(i)}) + \frac{\lambda}{2.1000} \stackrel{\text{E}}{=} ||W^{(\ell)}||_F^2$. Bookprop to compart gradules cort Jeez (usy (xst2 xst2)) W:= W - ddw(2), b(1) = b(1) - ddb(2) "I epoch" poss through training set.