TTIC 31230 Fundamentals of Deep Learning

Langevin Dynamics Problems.

Problem 1. This problem is on batch size scaling of Langevin dynamics. We consider batched SGD as defined by

$$\Phi \mathrel{-}= \eta \hat{g}^B$$

where $\hat{g}^B$ is the average of B sampled gradients. Let g be the average gradient $g = E\ \hat{g}$.

The covariance matrix at batch size B is

$$\Sigma^B[i,j] = E\ (\hat{g}^B[i] - g[i])(\hat{g}^B[j] - g[j]).$$

Langevin dynamics is

$$\Phi(t + \Delta t) = \Phi(t) - g\Delta t + \epsilon\sqrt{\Delta t} \quad \epsilon \sim \mathcal{N}(0, \eta\Sigma^B)$$

Show that for $\eta = B\eta_0$ the Langevin dynamics is determined by η_0 independent of B.

Solution:

$$\begin{aligned}
\Sigma^B[i,j] &= E\ (\hat{g}^B[i] - g[i])(\hat{g}^B[j] - g[j]) \\
&= \frac{1}{B^2} E \left(\sum_b \hat{g}_b[i] - g[i]\right)\left(\sum_b \hat{g}_b[j] - g[j]\right) \\
&= \frac{1}{B^2} E \sum_{b,b'} (\hat{g}_b[i] - g[i])\,(\hat{g}_{b'}[j] - g[j]) \\
&= \frac{1}{B^2} \sum_{b,b'}\ E\ \ (\hat{g}_b[i] - g[i])\,(\hat{g}_{b'}[j] - g[j]) \\
&= \frac{1}{B^2} \sum_{b}\ E\ \ (\hat{g}_b[i] - g[i])\,(\hat{g}_{b}[j] - g[j]) \\
&= \frac{1}{B}\Sigma^1[i,j]
\end{aligned}$$

So for $\eta = B\eta_0$ we have $\eta\Sigma^B = \eta_0\Sigma^1$ which yields the equivalence.