- For any value of the loss L

$$\left(\frac{S(L)}{S_{min}(L)}-1\right)\left(\frac{D(L)}{D_{min}(L)}-1\right)=1$$

- S_{min}, D_{min} – minimum # of steps / datapoints necessary to reach L
- S, D – any other combination of steps & datapoints used to reach L

- recall $D = SB$

- $B_{crit}(L)$ – critical batch size, $B_{crit}(L) := \frac{D_{min}}{S_{min}}$

- if train at $B = B_{crit}$

  ↳ $S(L) = 2S_{min}$, $E = 2D_{min}$

- How does S & S_{min} relate?

  let $r := \frac{S}{S_{min}}$

$$(r-1)\left(r \cdot \frac{B}{B_c} - 1\right) = 1$$

=> solve quadratic.

assume first that $B/B_c \gg 1$.

Then $\approx$ $(r-1)\, r \cdot \frac{B}{B_c} \approx 1$

$$r(r-1) \approx \frac{B_c}{B}$$

Note: $r(r-1) = \dfrac{B_c}{B}$

$\qquad\qquad\qquad\quad \approx 0.$

≥ 1

so... need $r - 1 \approx 0$, $r \approx 1$

to first order: $r - 1 \approx \dfrac{B_c}{B}$

so: $r \approx 1 + \dfrac{B_c}{B}$

$\dfrac{S}{S_{min}} \approx 1 + \dfrac{B_c}{B}$

$$\boxed{S_{min} \approx \frac{S}{1 + \frac{B_c}{B}}}$$

~~Let $C_{...} =$~~

- Recall $C = 6NBS$, $D = BS$
- Let $C_{min} = 6N D_{min} = 6N B_c S_{min}$
- How are ~~q~~ C and C_{min} related?
- Let $q = \frac{C}{C_{min}}$.
- notice $\frac{D}{D_m} = \frac{6ND}{6ND_m} = \frac{C}{C_m}$

$$\frac{S}{S_m} = \frac{6NBS}{6NB_c S_m} \cdot \frac{B_c}{B} = \frac{C}{C_m} \cdot \frac{B_c}{B}$$

- So, $\left(\frac{S}{S_m} - 1\right)\left(\frac{D}{D_m} - 1\right) = 1$ is equivalent to

$$\left(q \frac{B_c}{B} - 1\right)(q - 1) = 1.$$

- if $\frac{B_c}{B} \ll 1$, we find, as before,

$$q - 1 \approx \frac{B}{B_c} \iff \frac{C}{C_{min}} \approx 1 + \frac{B}{B_c} \iff C_{min} = \frac{C}{1 + \frac{B}{B_c}}$$

L

$N \quad B \quad S$

$D = BS$

$C = 6NBS = 6ND$

$$B_{crit}(L) = \frac{D_{min}}{S_{min}} \longleftarrow S_{min}(L)$$

$$S_{min} = \frac{S}{1 + \frac{B_{crit}}{B}}$$

$$D_{min}(L) \rightarrow C_{min} = 6ND_{min}$$

$$\left(\frac{S(L)}{S_{min}(L)} - 1\right)\left(\frac{D(L)}{D_{min}(L)} - 1\right) = 1$$

$$C_{min} = \frac{C}{1 + B/B_{crit}}$$

$$L(N, S_{min}) = \left(\frac{N_c}{N}\right)^{\alpha_N} + \left(\frac{S_c}{S_{min}}\right)^{\alpha_S}$$

$$L(N,D) = \left[\left(\frac{N_c}{N}\right)^{\alpha_N/\alpha_D} + \frac{D_c}{D}\right]^{\alpha_D} \longleftrightarrow \left[\left(\frac{N_c}{N}\right)^{\alpha_N/\alpha_D} + \frac{C_c}{C}\right]^{\alpha_D} =: L(N,C)$$

$C_c := 6ND_c$

$C \updownarrow C_{min}$

alternative equation C_{min}

$$N \simeq C_{min}^{\alpha_D/\alpha_N}$$

Legend:

[dotted box] → empirical obs

→ implication