

- For any value of the loss L

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

$\cdot S_{\min}, D_{\min}$ - minimum # of steps / datapoints necessary to reach L

- S, D - any other combination of steps & datapoints used to reach L

\cdot Recall $D = SB$

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

\cdot if train at $B = B_{\text{crit}}$

$$\hookrightarrow S = 2S_{\min}, \quad E = 2B_{\min}$$

- How does S & S_{\min} relate?

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

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

\Rightarrow solve quadratic.

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

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

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

Note ~~star~~ : $r(r-1) = \frac{B_c}{B}$

↙

$\approx 0.$

$\gg 1$

so ... need $r \approx 1 + \frac{B_c}{B}$, $r \approx 1$

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

so : $t \approx 1 + \frac{B_c}{B}$

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

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

~~Let $C_{\min} =$~~

- Recall $C = 6NB_S$, $D = B_S$

- Let $C_{\min} = 6N D_{\min} = 6N B_c S_{\min}$

- How are C and C_{\min} related?

- Let $q = \frac{C}{C_{\min}}$

- Notice $\frac{D}{D_m} = \frac{6ND}{6ND_m} = \frac{C}{C_m}$

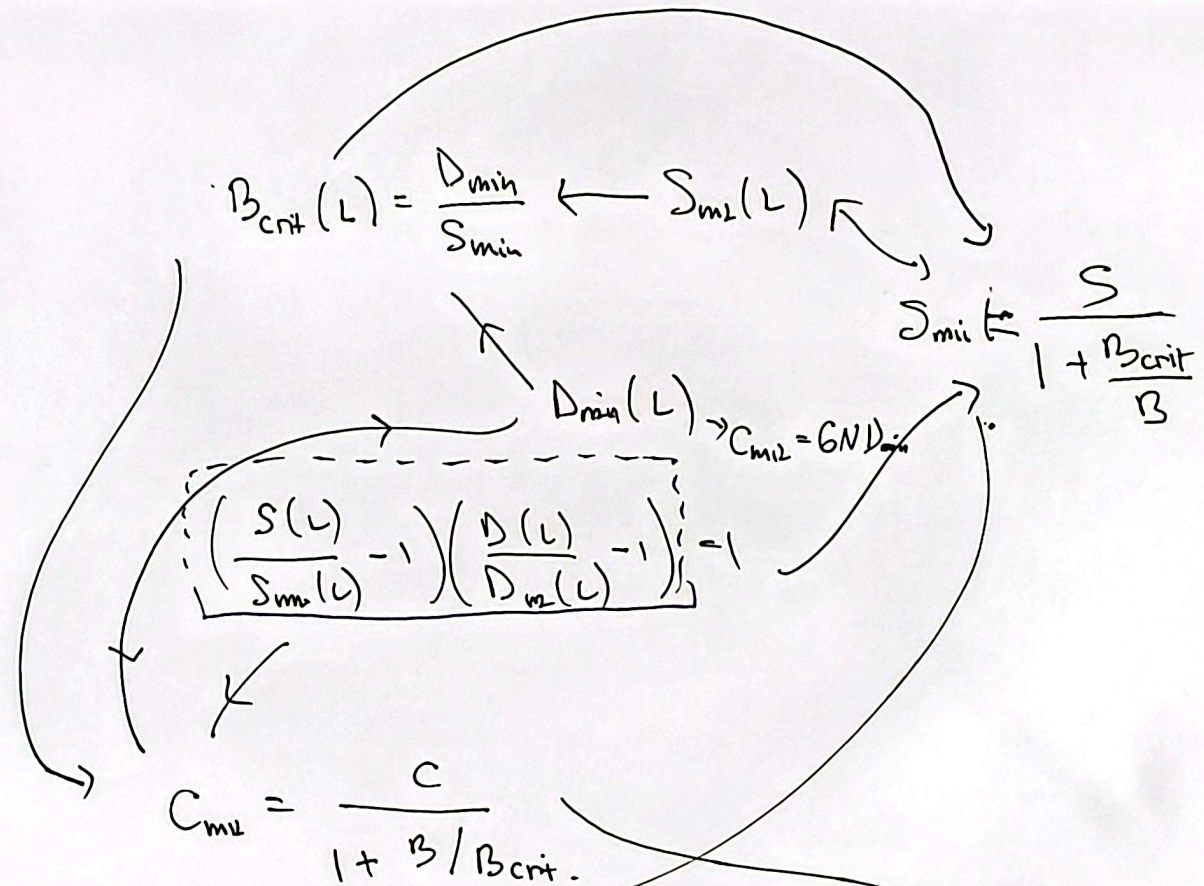
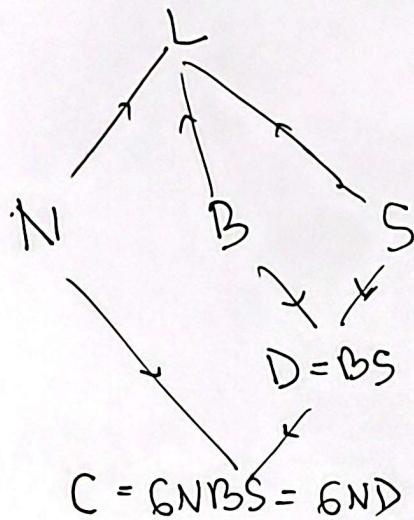
$$\frac{S}{S_m} = \frac{6NB_S}{6NB_c S_m} \cdot \frac{B_c}{B} = \frac{C}{C_m} \cdot \frac{B_c}{B}$$

- 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,

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



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

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

Legend:

- $\square \rightarrow$ empirical obs
- \rightarrow implication

Additional notes:

- $C_c := 6ND_c$
- $N \propto C_{min}^{\alpha_D/\alpha_N}$
- alternative equation C_{min}