Power Laws

- $L(N) \approx (N/8.8e13)^{-0.076}$
- $L(D) \approx (D/5.4e13)^{-0.095}$
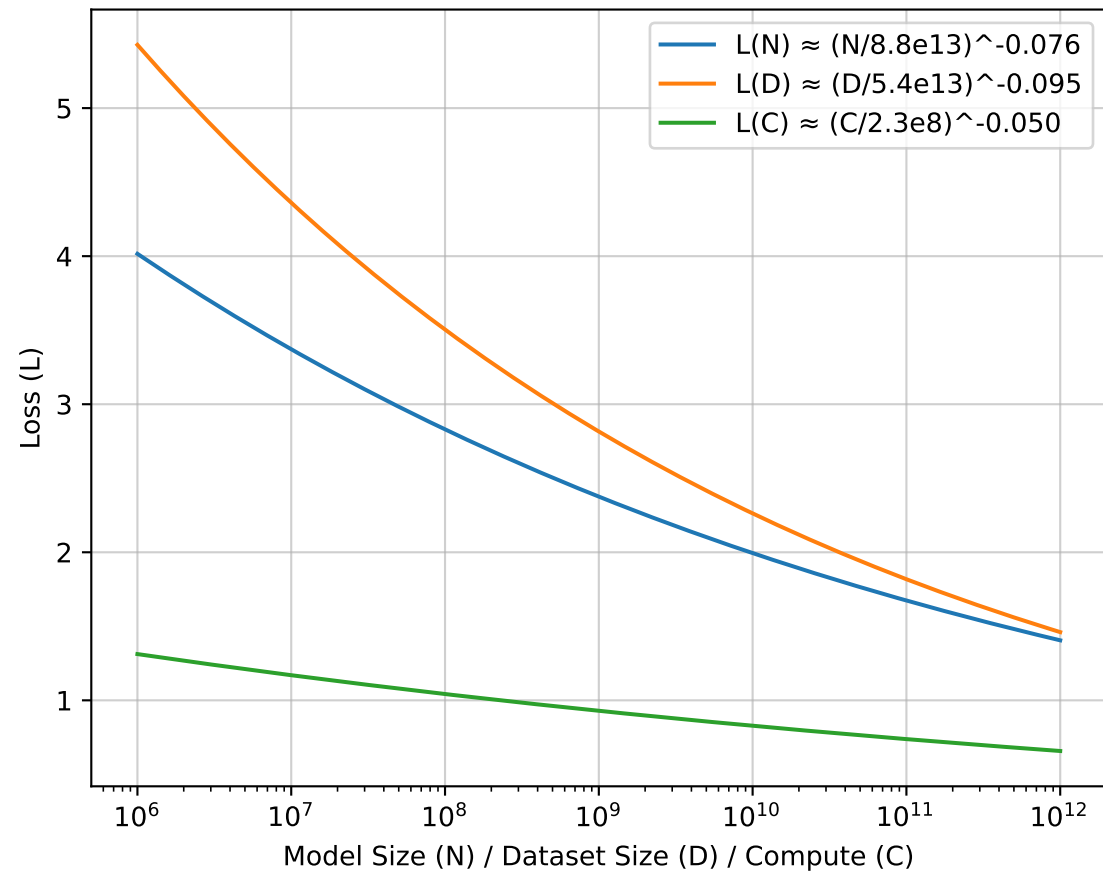- $L(C) \approx (C/2.3e8)^{-0.050}$

Loss (L)

Model Size (N) / Dataset Size (D) / Compute (C)

Optimal Allocation of the Compute Budget

- Kaplan: $D \approx N^{0.74}$
- Chinchilla: $D \approx 20N$
- DeepSeek: $D \approx 10N$

Dataset Size (D)

Model Size (N)