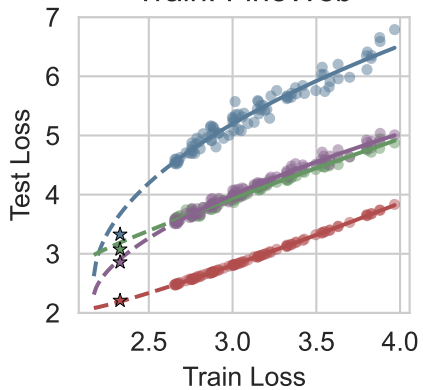
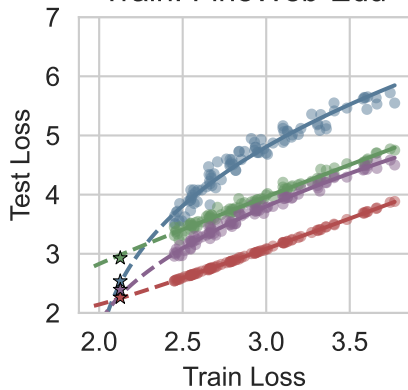


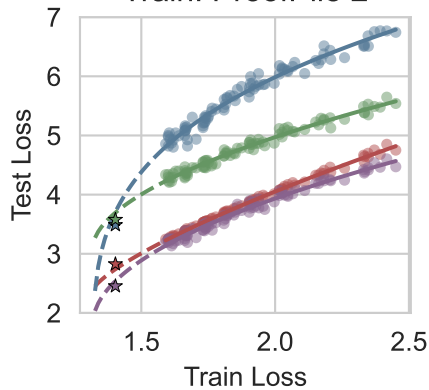
Train: FineWeb



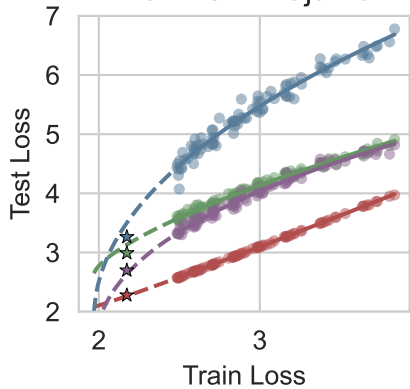
Train: FineWeb-Edu



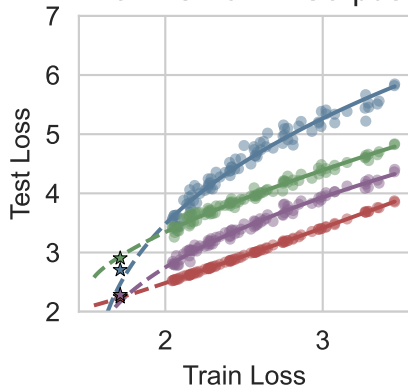
Train: ProofPile 2



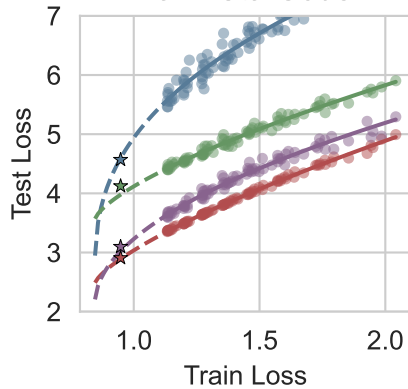
Train: SlimPajama



Train: SmolLM Corpus



Train: StarCoder



Test data

- Hellaswag
- ARC-Easy
- MMLU-Humanities
- MMLU-STEM