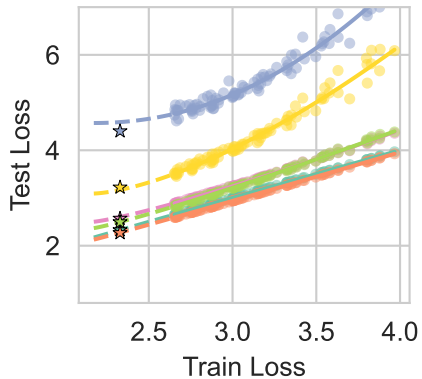
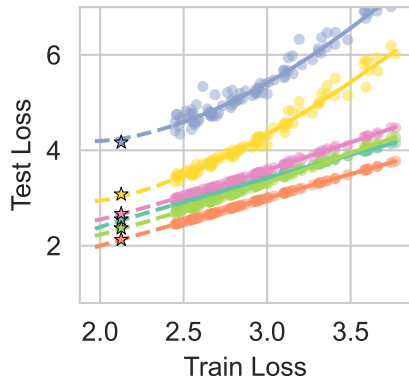


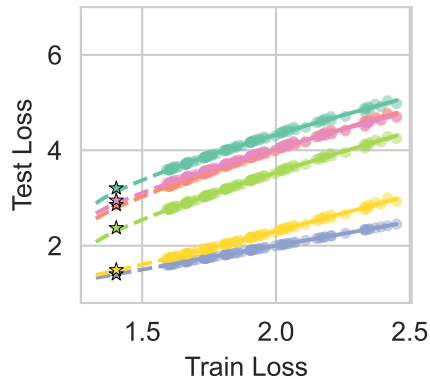
Train: FineWeb



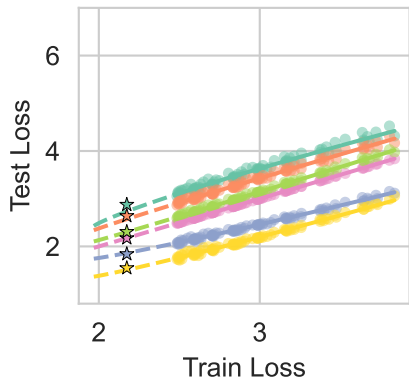
Train: FineWeb-Edu



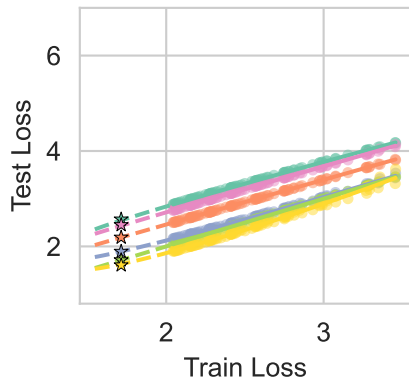
Train: ProofPile 2



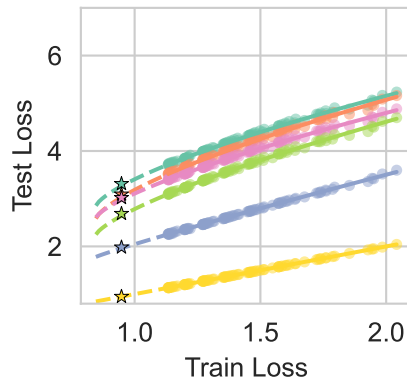
Train: SlimPajama



Train: SmolLM Corpus



Train: StarCoder



Test data

- FineWeb
- FineWeb-Edu
- ProofPile 2
- SlimPajama
- SmolLM Corpus
- StarCoder