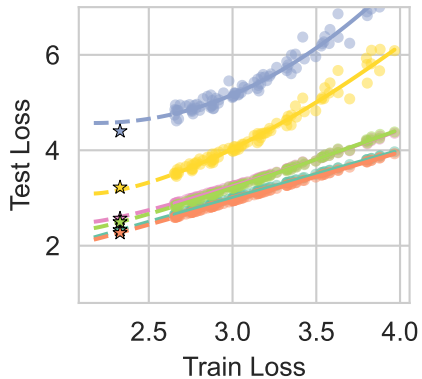
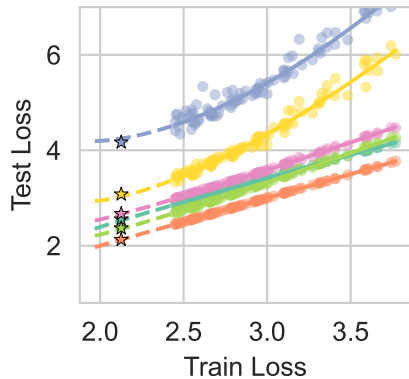


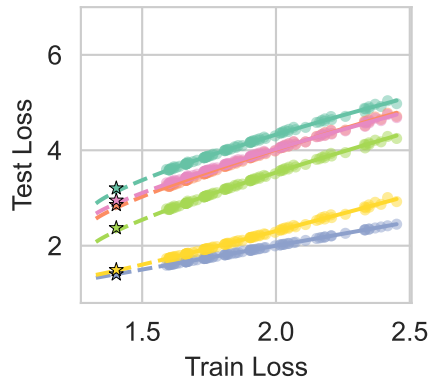
Train: FineWeb



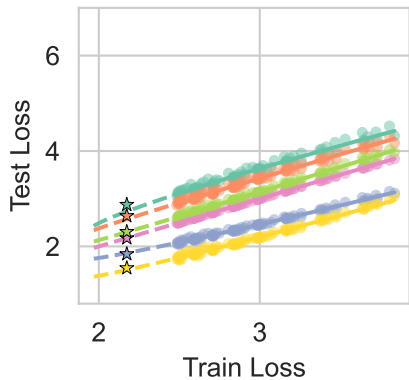
Train: FineWeb-Edu



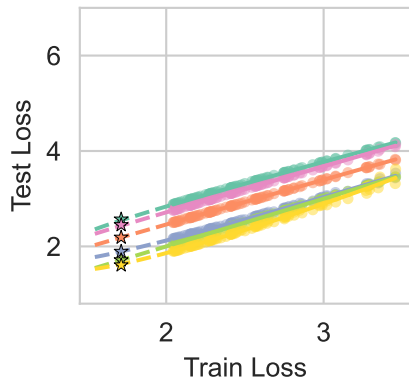
Train: ProofPile 2



Train: SlimPajama



Train: SmolLM Corpus



Train: StarCoder

