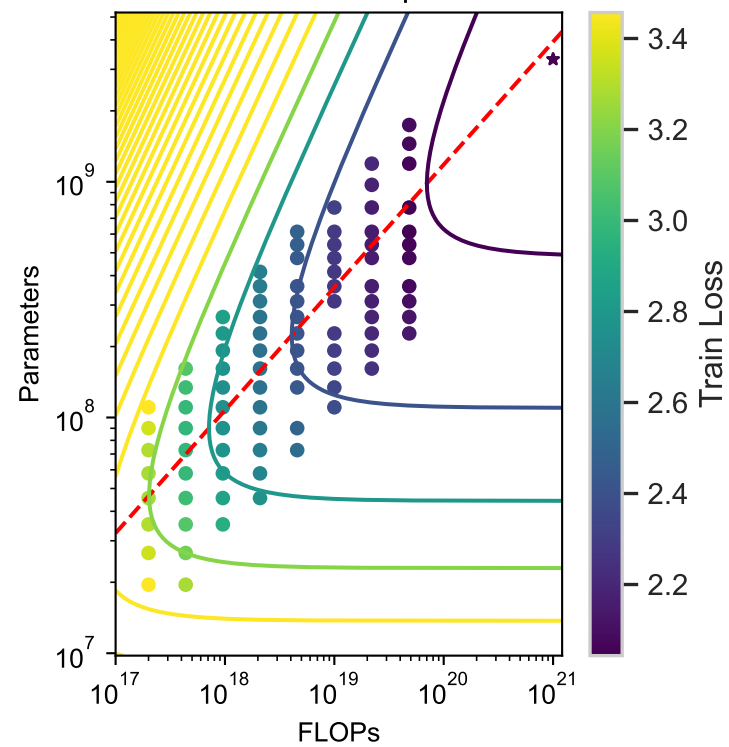
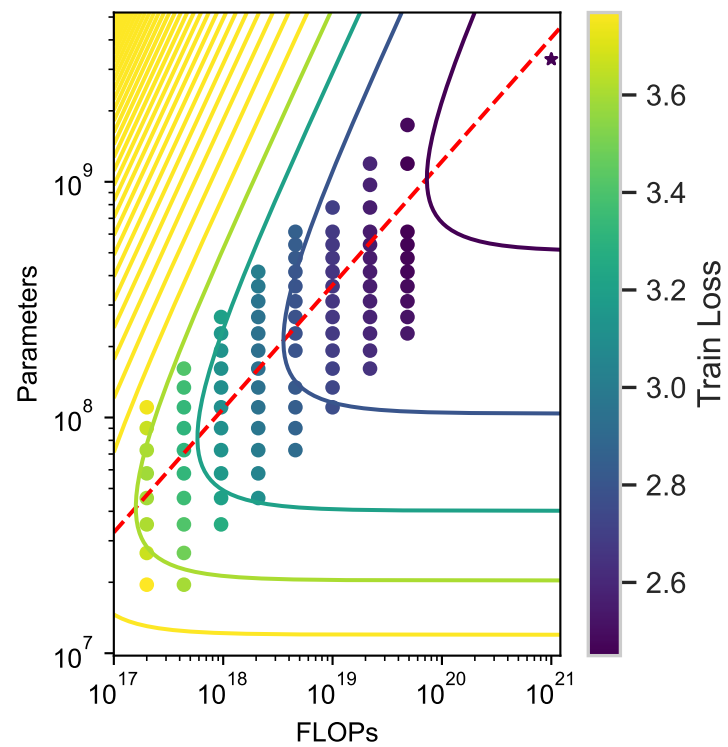


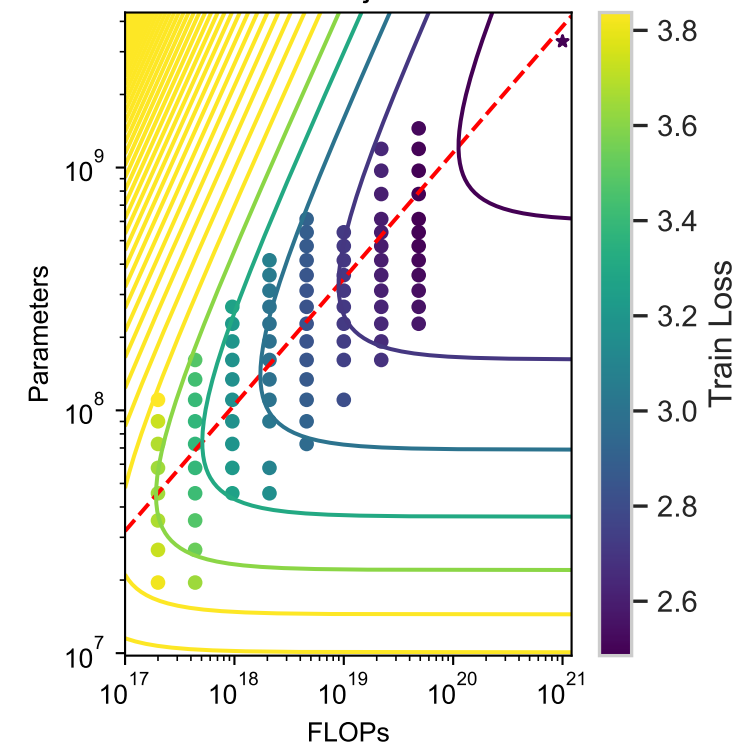
SmolLM Corpus



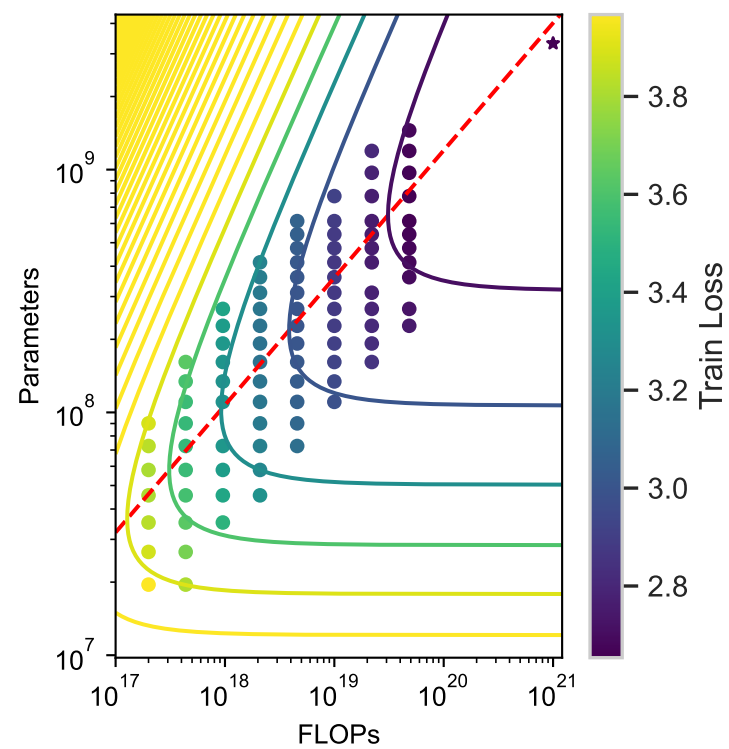
FineWeb-Edu



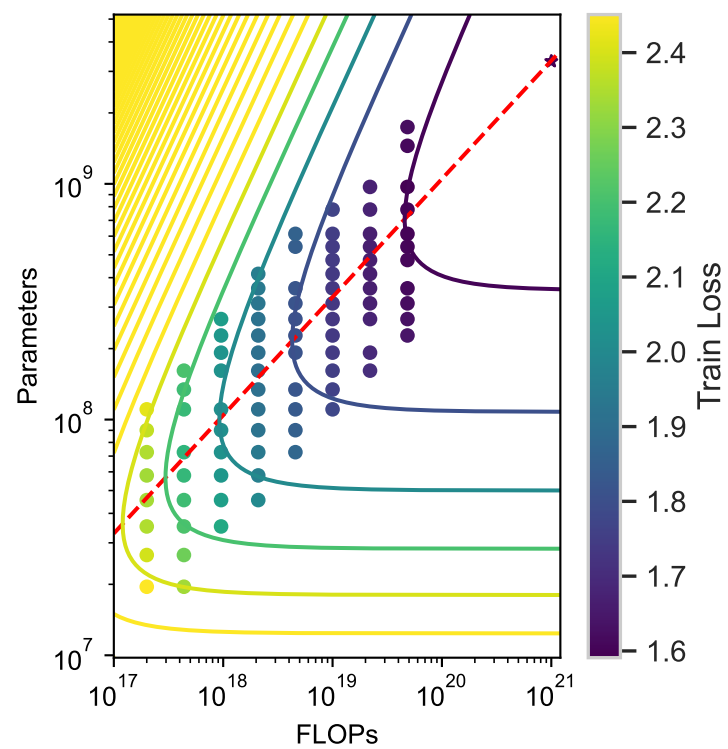
SlimPajama



FineWeb



ProofPile 2



StarCoder

