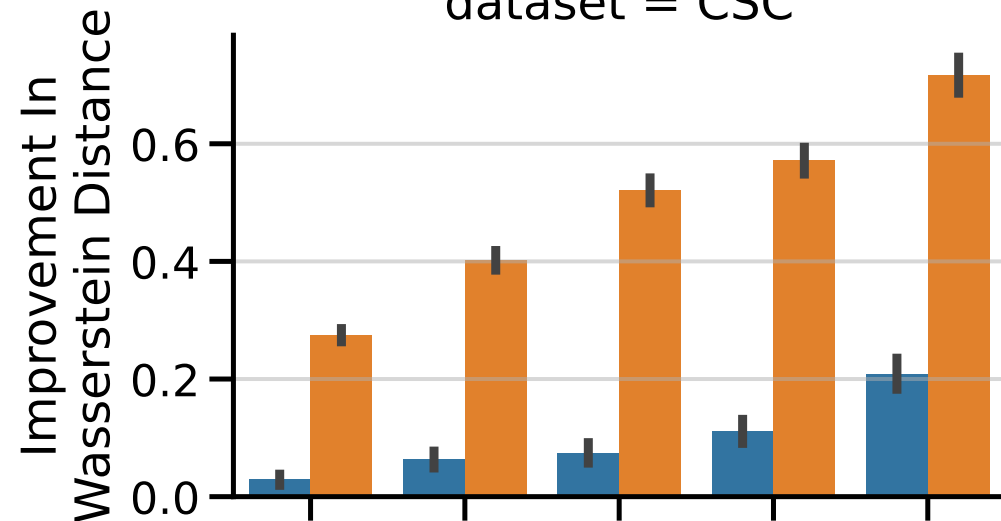


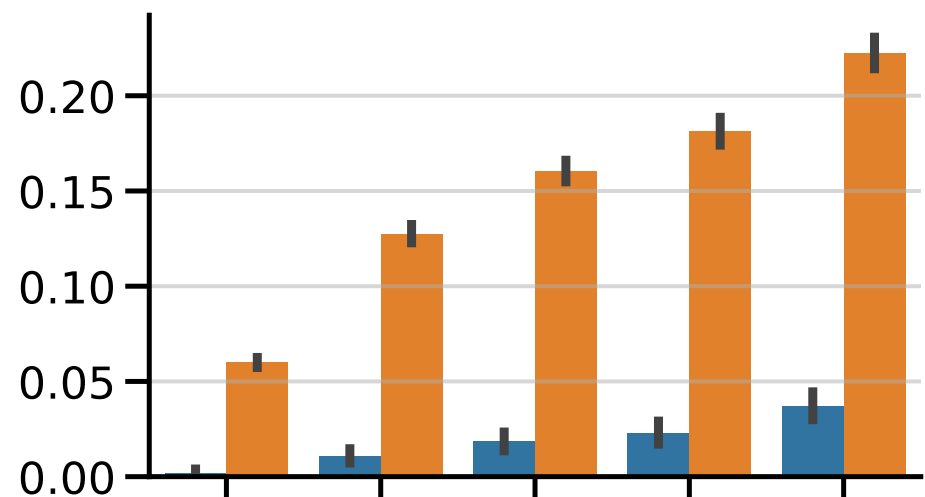
Test-time Scaling Baseline

Model Averaging BoN Oracle

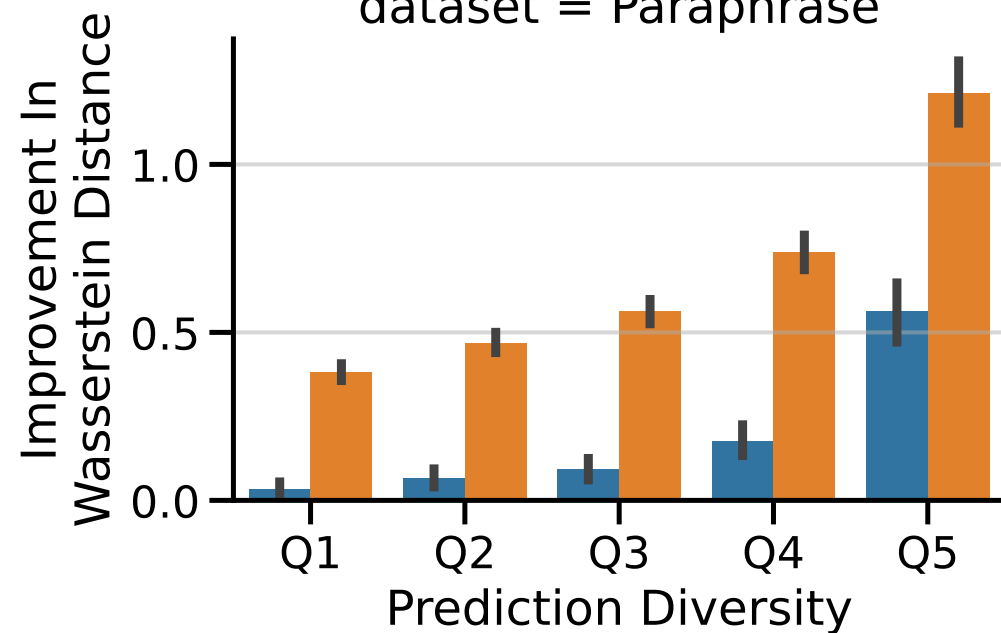
dataset = CSC



dataset = MP



dataset = Paraphrase



dataset = VariErrNLI

