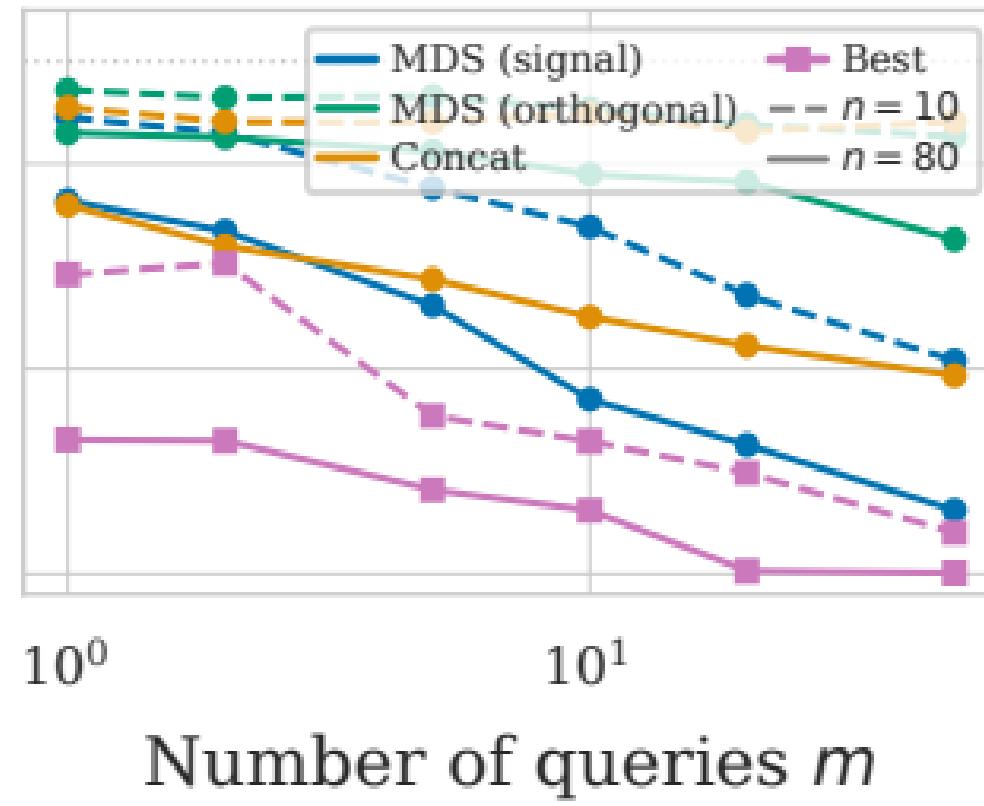
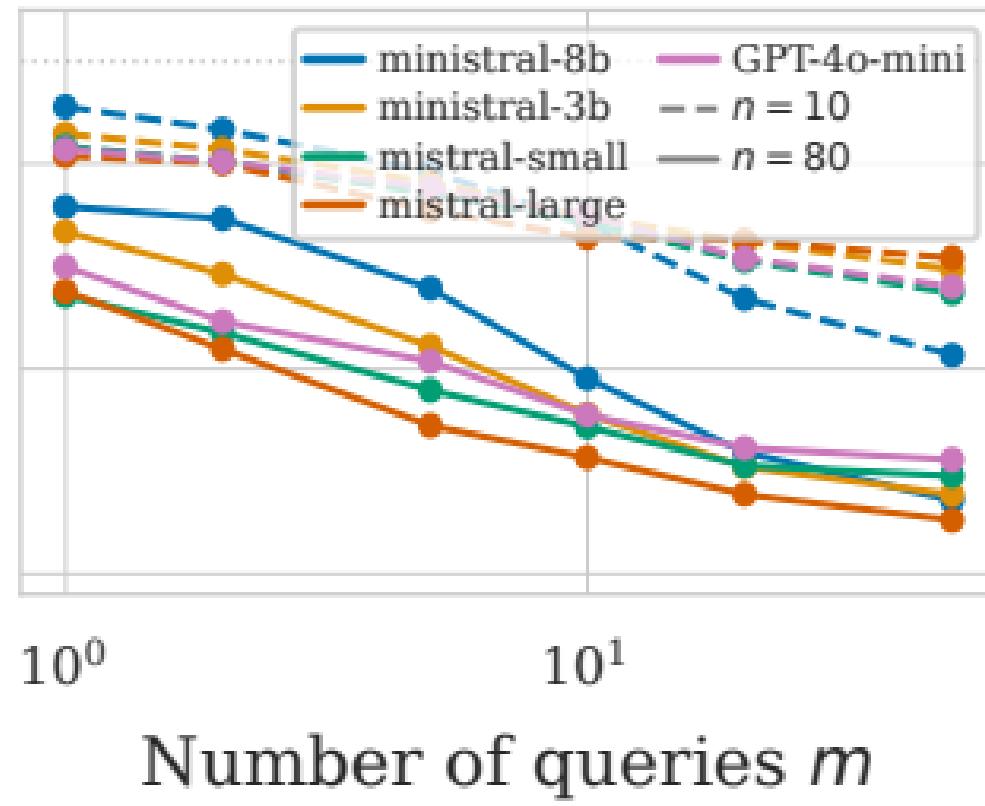


(a) MDS vs. baselines

Mean error



(b) Across base LLMs



(c) Across embeddings

