
QUENTIN BERTRAND, Mila

Synergies Between Disentanglement and Sparsity: a Multi-Task Learning Perspective

Although disentangled representations are often said to be beneficial for downstream tasks, current empirical and theoretical understanding is limited. In this work, we provide evidence that disentangled representations coupled with sparse base-predictors improve generalization. In the context of multi-task learning, we prove a new identifiability result that provides conditions under which maximally sparse base-predictors yield disentangled representations. Motivated by this theoretical result, we propose a practical approach to learn disentangled representations based on a sparsity-promoting bi-level optimization problem. Finally, we explore a meta-learning version of this algorithm based on group Lasso multiclass SVM base-predictors, for which we derive a tractable dual formulation. It obtains competitive results on standard few-shot classification benchmarks, while each task is using only a fraction of the learned representations.