CONSISTENCY OF AUGMENTATION GRAPH AND NETWORK APPROXIMABILITY IN CONTRASTIVE LEARNING

MARTINA NEUMAN AND CHENGHUI LI

Abstract: Contrastive learning leverages data augmentation to develop feature representation without relying on large labeled datasets. However, despite its empirical success, the theoretical foundations of contrastive learning remain incomplete, with many essential guarantees left unaddressed, particularly the realizability assumption concerning neural approximability of an optimal spectral contrastive loss solution. In this work, we overcome these limitations by analyzing the pointwise and spectral consistency of the augmentation graph Laplacian. We establish that, under specific conditions for data generation and graph connectivity, as the augmented dataset size increases, the augmentation graph Laplacian converges to a weighted Laplace-Beltrami operator on the natural data manifold. These consistency results ensure that the graph Laplacian spectrum effectively captures the manifold geometry. Consequently, they give way to a robust framework for establishing neural approximability, directly resolving the realizability assumption in a current paradigm.