

Abstract

We initiate the study of metric embeddings with *outliers*. Given some finite metric space we wish to remove a small set of points and to find either an isometric or a low-distortion embedding of the remaining points into some host metric space. This is a natural problem that captures scenarios where a small fraction of points in the input corresponds to noise. We present polynomial-time approximation algorithms for computing outlier embeddings into Euclidean space, trees, and ultrametrics. In the case of isometric embeddings the objective is to minimize the number of outliers, while in the case of non-isometries we have a bi-criteria optimization problem where the goal is to minimize both the number of outliers and the distortion. We complement our approximation algorithms with NP-hardness results for these problems. We conclude with a brief experimental evaluation of our non-isometric outlier embedding on synthetic and real-world data sets.