



Metric repair on manifolds with holes

Anna Gilbert (University of Michigan, USA)

Abstract: For many machine learning tasks, the input data lie on a low-dimensional manifold embedded in a high-dimensional space and, because of this high-dimensional structure, most algorithms are inefficient. The typical solution is to reduce the dimension of the input data using standard dimension reduction algorithms such as ISOMAP, LAPLACIAN EIGENMAPS or LLES. This approach, however, does not always work in practice as these algorithms require that we have somewhat ideal data. Unfortunately, most data sets either have missing entries or unacceptably noisy values. That is, real data are far from ideal and we cannot use these algorithms directly.

In this talk, we focus on the case when we have missing data. Some techniques, such as matrix completion, can be used to fill in missing data but these methods do not capture the non-linear structure of the manifold. Here, we present a new algorithm MR-MISSING that extends these previous algorithms and can be used to compute low dimensional representation on data sets with missing entries. We demonstrate the effectiveness of our algorithm by running three different experiments. We visually verify the effectiveness of our algorithm on synthetic manifolds, we numerically compare our projections against those computed by first filling in data using nLPCA and mDRUR on the MNIST data set, and we also show that we can do classification on MNIST with missing data. We also provide a theoretical guarantee for MR-MISSING under some simplifying assumptions.