

Learning a Parametric Mapping for Non-linear Dimensionality Reduction

Pooyan Khajepour*

Ali Ghodsi†

The foremost nonlinear dimensionality reduction algorithms (e.g. Maximum Variance Unfolding, ISOMAP and Locally Linear Embedding) provide an embedding only for the given training data, with no straightforward extension for the test points. This shortcoming makes them unsuitable for problems such as classification and regression. On the other hand, linear dimensionality reduction algorithms (e.g. Principal Component Analysis and Multidimensional Scaling) are capable of handling the out-of-sample examples easily, but their effectiveness is limited by the linearity of the subspace they reveal. In this research we propose a novel dimensionality reduction algorithm which learns a parametric mapping between the high-dimensional space and the embedded space. The key observation is that when the dimensionality of the data is greater than its quantity, it is always possible to find a linear transformation that preserves a given subset of pairwise distances, while changing the distances of another subset.

We present a method that learns a parametric mapping between the high and low dimensional spaces in two steps. First, the input data is projected into a high-dimensional feature space, and then an affine transformation is learned that maps data points from the feature space into the low-dimensional embedding space. The search for this transformation is cast as an instance of semi-definite programming (SDP), which is convex and always converges to a global optima. The resulted transformation can then be used to map out-of-sample points into the embedded space.

Our experimental results on real and synthetic data sets demonstrate that the proposed algorithm produces a robust and faithful embedding even for very small data sets. It also shows that it is successful at projecting out-of-sample examples.

Another feature of this algorithm is that it treats the distances between the data points in three different ways. One can preserve a subset of the distances, stretch another subset and leave the third set unspecified. This is in contrast with methods like Maximum Variance Unfolding that preserve local distances but stretch any non-local pairs. This property makes the proposed algorithm useful for semi-supervised tasks where only partial information about similarity and dissimilarity of points is known.

Topic: Learning Algorithms Preference: Oral
--

*David R. Cheriton School of Computer Science, University of Waterloo, 200 University Ave. W., Waterloo, Ontario, Canada N2L 3G1, pkhajehp@cs.uwaterloo.ca.

†Department of Statistics & David R. Cheriton School of Computer Science, University of Waterloo, 200 University Ave. W., Waterloo, Ontario, Canada N2L 3G1, aghodsib@uwaterloo.ca.