Out-of-Sample Extensions for LLE, Isomap, MDS, Eigenmaps, Spectral Clustering

Abstract

we're trying to build the theory for how to do OOS embedding for a bunch of different algorithms instead of just one

Introduction

a bunch of people have made dimensionality reduction algorithms recently

Common Framework

all of our algorithms are based in embedding training points using the principal eigenvectors of a symmetric matrix

algorithm 1

1. Starting with some dataset, construct a similarity matrix M where M_{ij}Mij is some measure of similarity between x_ixi and x_jxj
2. normalize M if you want
3. compute the m largest positive eigenvalues and eigenvectors of M
4. the embeddings are the rows of this reduced eigenmatrix (I think?)

now do this with slight variations and you have MDS, spectral clustering, laplacian eigenmaps, isomap, and LLE

From eigenvectors to eigenfunctions

you can use kernel methods to speed up computations

Proposition 1: if you have a kernel function K that generates elements of a matrix M, there are equations that relate the eigenstuff of K and M.

extending to new points

You can use prop. 1 to extend all the algorithms we talked about earlier to new points

Experiments

we get pretty good error empirically

Conclusions

we extended MDS, spectral clustering, laplacian eigenmaps, isomap, and LLE in this paper.