SLT coding exercise #1

Locally Linear Embedding
https://gitlab.vis.ethz.ch/vwegmayr/slt-coding-exercises

Due on Monday, March 6th, 2017

YOUR NAME
YOUR LEGI NUMBER

Contents

The Model

The model section is intended to allow you to recapitulate the essential ingredients used in Locally Linear Embedding. Write down the *necessary* equations to specify Locally Linear Embedding and and shortly explain the variables that are involved. This section should only introduce the equations, their solution should be outlined in the implementation section.

Hard limit: One page

Your Answer

The Answers

This is the core section of your report, which contains the tasks for this exercise and your respective solutions. Make sure you present your results in an illustrative way by making use of graphics, plots, tables, etc. so that a reader can understand the results with a single glance. Check that your graphics have enough resolution or are vector graphics. Consider the use of GIFs when appropriate.
Hard limit: Two pages

(a) Get the data

For this exercise we will work with the MNIST data set. In order to learn more about it and download it, go to http://yann.lecun.com/exdb/mnist/.

(b) Locally linear embedding

Implement the LLE algorithm and apply it to the MNIST data set. Provide descriptive visualizations for 2D & 3D embedding spaces. Is it possible to see clusters?

(c) Cluster structure

Investigate the cluster structure of the data. Can you observe block structures in the M matrix (use matrix plots)? Also plot the singular values of M. Do you notice something? Can you think of ways to determine the optimal embedding dimension?

(d) Nearest Neighbors

Investigate the influence of the choice of how many nearest neighbors you take into account. Additionally, try different metrics to find the nearest neighbors (we are dealing with images!).

(e) Linear manifold interpolation

Assume you pick some point in the embedding space. How can you map it back to the original (high dimensional) space? Investigate how well this works for points within and outside the manifold (does it depend on the dimensionality of the embedding space?) Try things like linearly interpolating between two embedding vectors and plot the sequence of images along that line. What happens if you do that in the original space?

In the plots, expecially the 3D ones, the clusters are clearly visible. It seems that more numbers are
"recognised" (gathered in the same cluster) better tant the others.
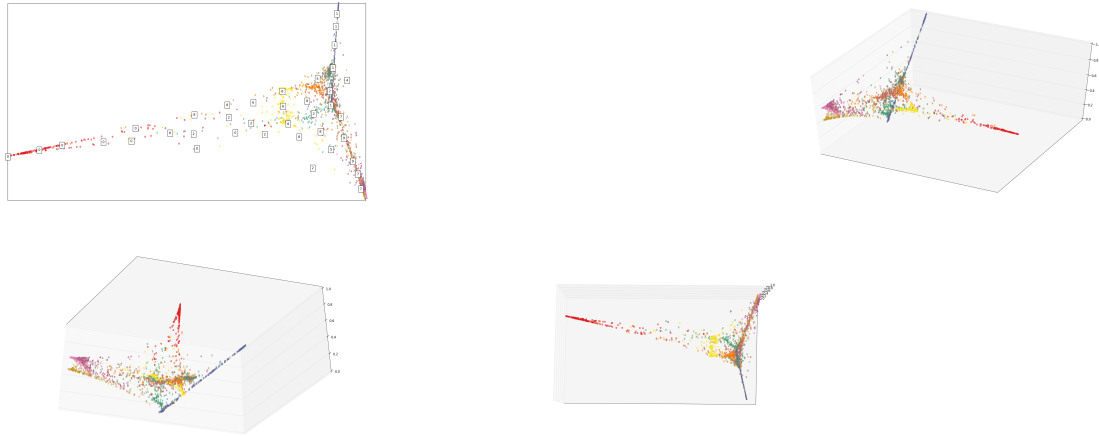

No relevant box structures seem to appear in the matrix plot of M.
The matrix appears to be symmetric, sparse and with higher numbers on the diagonal.
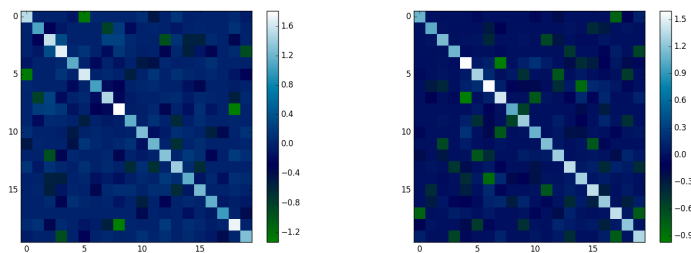
Singular Values

By plotting the singular values in order it can be noticed that there are a few of them with a high absolute
value. Since the singular values are associated with the energy of the projection on the lower-dimensional
space, a way to determine the optimal embedding space would be to choose the number of dimendions of
the space equal to the number singular values (in order of decreasing magnitude) that have the biggest
difference in value from the others (i.e. before their value starts to decrease slowly). This can be regarded
as an "elbow method".
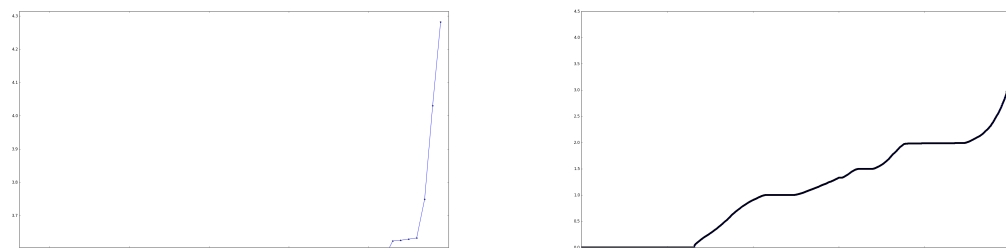
Interpolation in the embedding space

The Implementation

In the implementation section you give a concise insight to the practical aspects of this coding exercise. It mainly mentions the optimization methods used to solve the model equations. Did you encounter numerical or efficiency problems? If yes, how did you solve them? Provide the link to your git branch of this coding exercise.
Hard limit: One page

Your Answer

Your Page

Your page gives you space to include ideas, observations and results which do not fall into the categories provided by us. You can also use it as an appendix to include things which did not have space in the other sections.

No page limit.

Your Answer
YOUR GIT BRANCH