SLT coding exercise #1

Locally Linear Embedding

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

Due on Monday, March 6th, 2017

Thilo Weber
11-944-477

Contents

The Model 3

The Questions 4

The Implementation 6

Your Page 7

The Model

We have a data set on N points $\{\mathbf{x}_i\}_{i=1}^N$ of dimension D, $\mathbf{x}_i \in \mathbb{R}^D$. The goal is to find a non-linear embedding of these points in $\mathbb{R}^d$, where $d \ll D$. In Locally Linear Embeddingthis is done by first representing each data point as a function of its K nearest neighbors. The resulting reconstruction cost function is defined as:

$$\mathcal{E}(\mathbf{W}) = \sum_i ||\mathbf{x}_i - \sum_j w_{ij}\mathbf{x}_i||^2,$$

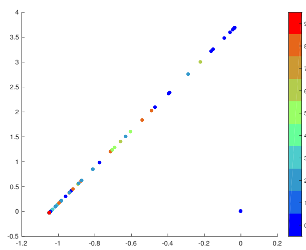
where the weights $w_{ij} \neq 0$ only if $\mathbf{x}_j$ is a nearest neighbor of $\mathbf{x}_j$ and each row of the weight matrix $\mathbf{W}$ sums up to 1, $\sum_j w_{ij} = 1$. The weights are calculated by minimizing the reconstruction cost function under these constraints. In the embedding step each point $\mathbf{x}_i$ is mapped to a point $\mathbf{y}_i \in \mathbb{R}^d$ in the low dimensional space. The points $\mathbf{y}_i$ are calculated by minimizing the embedding cost function defined as:

$$\Phi(\mathbf{y}_1, \ldots, \mathbf{y}_N) = \sum_i ||\mathbf{y}_i - \sum_j w_{ij}\mathbf{y}_i||^2.$$

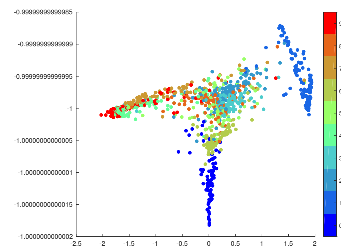
The Questions

Answer (b), (d):

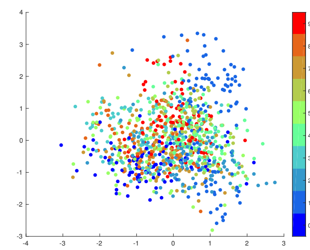
The results of applying the LLE algorithm on the MNIST data set for 2D & 3D embedding spaces and different numbers K of nearest neighbors are shown in Figure 1. For $K = 10$ neighbors clusters of the different can be seen. Apart from the numbers 0 and 1 the handwritten digits are not clearly separable from each other.



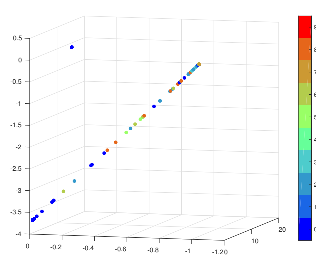
(a) $d = 2$, $K = 3$.



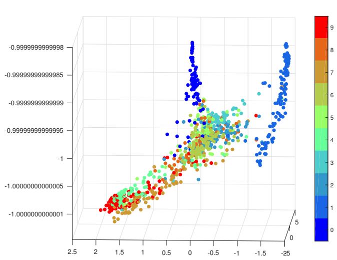
(b) $d = 2$, $K = 10$.



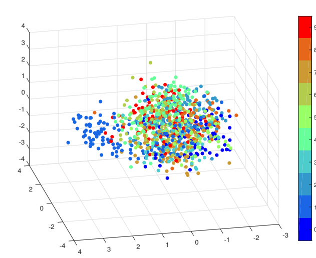
(c) $d = 2$, $K = 200$.



(d) $d = 3$, $K = 3$.



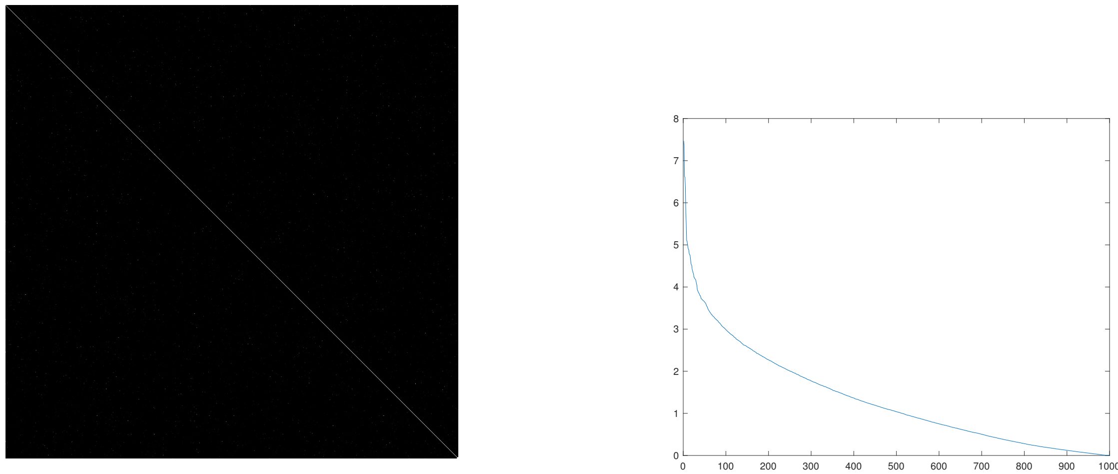
(e) $d = 3$, $K = 10$.



(f) $d = 3$, $K = 200$.

Figure 1: LLE algorithm applied on $N = 1000$ images of the MNIST data set. Each image has dimension $D = 784$.

Answer (c):
In Figure 2 the values and the singular values of the M matrix are plotted. The diagonal elements of the M matrix are all clearly non-zero and the off diagonal elements are mostly equal to zero or close to zero.



(a) Values of M matrix.

(b) Singular values of M matrix.

Figure 2: Investigation on M matrix for $d = 2$, $K = 10$.

The Implementation

Link to my git branch: `https://gitlab.vis.ethz.ch/vwegmayr/slt-coding-exercises/tree/11-944-477/1_locally_linear_embedding`.

Your Page

Your Answer
11-944-477/1_locally_linear_embedding