CMSE 820 HW 10

Hao Lin

November 27, 2018

1 Question 1

Define the local charts on $S^2 = \{(x,y,z) : x^2 + y^2 + z^2 = 1\}$ using: (1) polar coordinate representation; (2) stereographical projection.

Let $\Omega_1 = S^2 \setminus \{0,0,1\}$ and $\Omega_2 = S^2 \setminus \{0,0,-1\}$ be two open sets in $\mathbb{R}^3$. Then $C = \{\Omega_i : i \in \{1,2\}\}$ is an open cover of S^2.

Applying **stereographical projection** for these two open sets, we can define the following mappings

$$\phi_1 : \Omega_1 \to \mathbb{R}^2,\ \phi_1(x,y,z) = \left(\frac{x}{1-z}, \frac{y}{1-z}\right),$$

$$\phi_1^{-1} : \mathbb{R}^2 \to \Omega_1,\ \phi_1^{-1}(X,Y) = \left(\frac{2X}{1+X^2+Y^2}, \frac{2Y}{1+X^2+Y^2}, \frac{-1+X^2+Y^2}{1+X^2+Y^2}\right),$$

$$\phi_2 : \Omega_2 \to \mathbb{R}^2,\ \phi_2(x,y,z) = \left(\frac{x}{1+z}, \frac{y}{1+z}\right),$$

$$\phi_2^{-1} : \mathbb{R}^2 \to \Omega_2,\ \phi_1^{-1}(X,Y) = \left(\frac{2X}{1+X^2+Y^2}, \frac{2Y}{1+X^2+Y^2}, \frac{1-X^2-Y^2}{1+X^2+Y^2}\right).$$

It is easy to check that ϕ_1 and ϕ_2 are homeomorphisms, so they are two local charts under the stereographical projection that form an atlas of S^2.

Rewriting the Cartesian coordinate representation of S^2 in the **polar coordinate representation**, we have a different set of local charts (Technically speaking, these are *not* homeomorphisms, because $[0,\infty) \times [0,2\pi)$ is not open in $\mathbb{R}^2$.),

$$\psi_1 : \Omega_1 \to [0,\infty) \times [0,2\pi),\ \psi_1(x,y,z) = \left(\frac{\sqrt{x^2+y^2}}{1-z}, \arccos\frac{y}{x}\right),$$

$$\psi_1^{-1} : [0,\infty) \times [0,2\pi) \to \Omega_1,\ \psi_1^{-1}(r,\theta) = \left(\frac{2r\cos\theta}{1+r^2}, \frac{2r\sin\theta}{1+r^2}, \frac{r^2-1}{1+r^2}\right),$$

$$\psi_2 : \Omega_2 \to [0,\infty) \times [0, 2\pi),\ \ \psi_2(x,y,z) = \left(\frac{\sqrt{x^2+y^2}}{1+z}, \arccos\frac{y}{x} \right),$$

$$\psi_2^{-1} : [0,\infty) \times [0, 2\pi) \to \Omega_2,\ \ \psi_2^{-1}(r,\theta) = \left(\frac{2r\cos\theta}{1+r^2}, \frac{2r\sin\theta}{1+r^2}, \frac{-r^2+1}{1+r^2} \right).$$

2 Question 2

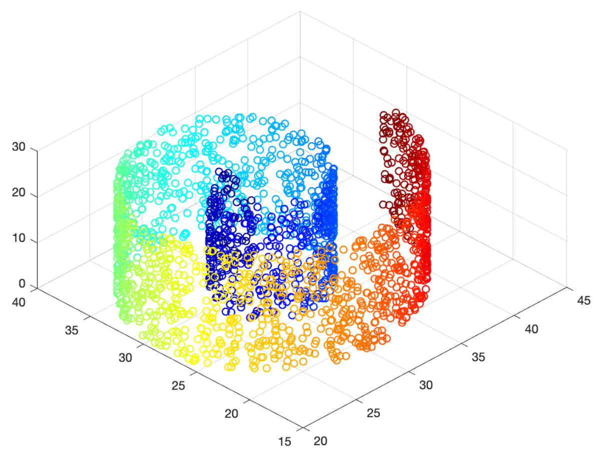

Figure 1: Original data in 3D.

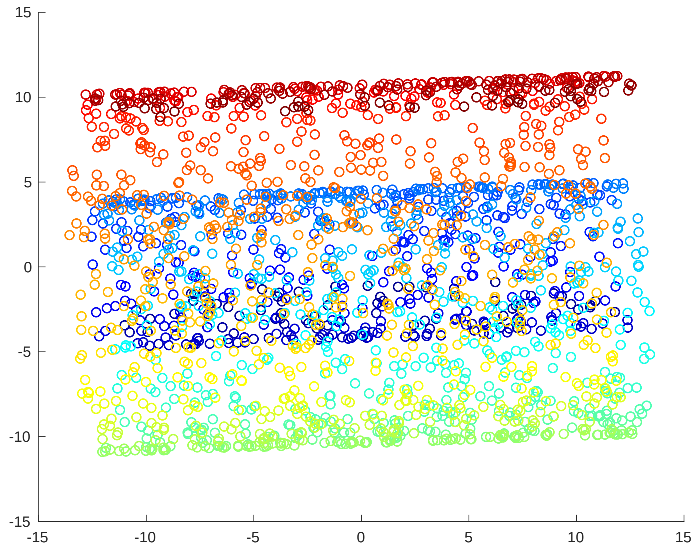

Figure 2: PCA. Data projected onto the top 2 PCs.

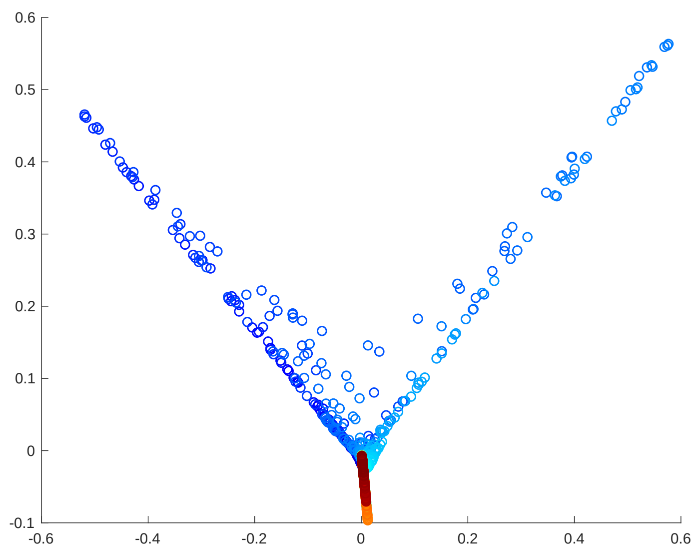

Figure 3: Kernel PCA. Data projected onto the top 2 PCs. $\sigma = 1.3$ is used.

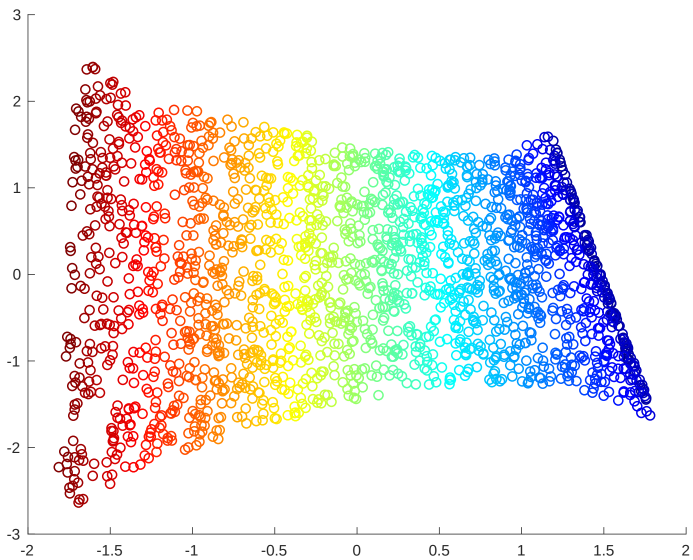

Figure 4: LLE with 12-nearest neighbours. Data reduced to 2D is plotted.

Kernel PCA is only able to separate the data points into 3 rays of similar color, which still differs quite a great deal from the original data. LLE, on the other hand, is capable of "unrolling" the data, restoring the 2D-manifold features (almost perfectly). When the high-dimensional data resembles some low-dimensional manifold embedded in the high-dimensional space, LLE outperforms plain kernel PCA.

3 Question 3

Laplacian Eigenmaps (LE).
Let $x_1, ..., x_n \in \mathbb{R}^p$ be the high dimensional data points. For any fixed scalar $\epsilon > 0$, x_i and x_j are called ϵ-neighbours of each other if and only if $\|x_i - x_j\|_2 \leq \epsilon$. Fix another scalar $\sigma^2 > 0$, and we can define a weight $w_{ij} = \exp(-\frac{\|x_i - x_j\|_2^2}{\sigma^2})$ if x_i and x_j are ϵ-neighbours and $w_{ij} = 0$ otherwise. A reasonable low dimensional embedding $y_1, ..., y_n$ minimizes the following objective function

$$\sum_{i,j} w_{ij} \|y_i - y_j\|_2^2.$$

3.1 a

Prove that $\min \frac{1}{2} \sum w_{ij} \|y_i - y_j\|_2^2 = \min \operatorname{Tr}(YLY^T)$, where $Y = [y_1, ..., y_n]$ and $L = D - A$ with $A_{ij} = w_{ij}$ and D being a diagonal matrix with $D_{ii} = \sum_j w_{ij}$. The matrix L is called graph laplacian.

Proof: Starting from the summation, we have

$$\begin{aligned}
\frac{1}{2}\sum_{ij} w_{ij}\|y_i - y_j\|_2^2 &= \frac{1}{2}\sum_{ij} w_{ij}(y_i^T y_i - y_i^T y_j - y_j^T y_i + y_j^T y_j) \\
&= \frac{1}{2}\big[2\sum_i \big(\sum_j w_{ij}\big) y_i^T y_i - 2\sum_{ij} w_{ij} y_i^T y_j\big] \\
&= \sum_i \big(\sum_j w_{ij}\big) y_i^T y_i - \sum_{ij} w_{ij} y_i^T y_j.
\end{aligned}$$

Starting from the trace, we have

$$\begin{aligned}
\mathrm{Tr}(YLY^T) &= \sum_i (YLY^T)_{ii} \\
&= \sum_i \big[\sum_{jk} Y_{ij} L_{jk} (Y^T)_{ki}\big] \\
&= \sum_i \big(\sum_{jk} Y_{ij} L_{jk} Y_{ik}\big) \\
&= \sum_{jk} L_{jk}\big(\sum_i Y_{ij} Y_{ki}\big) \\
&= \sum_{jk} L_{jk} y_j^T y_k \\
&= \sum_{ij} (D_{ij} - A_{ij}) y_i^T y_j \\
&= \sum_{ij} \big(\sum_k w_{ik}\big)\delta_{ij} y_i^T y_j - \sum_{ij} w_{ij} y_i^T y_j \\
&= \sum_i \big(\sum_j w_{ij}\big) y_i^T y_i - \sum_{ij} w_{ij} y_i^T y_j \\
&= \frac{1}{2}\sum_{ij} w_{ij}\|y_i - y_j\|_2^2.
\end{aligned}$$

3.2 b

$$\hat{Y} = \arg\min_Y \mathrm{Tr}(YLY^T) \text{ subject to } YDY^T = I \,\text{and}\, YD\mathbb{1} = \mathbf{0}.$$

Show that the solutions $\hat{Y} \in \mathbb{R}^{d\times n}$ are given by the eigenvectors corresponding to the lowest d eigenvalues of the generalized eigenvalue problem

$$Ly = \lambda Dy.$$

Proof: We ignore the constraint $YD\mathbb{1} = \mathbf{0}$ for the time being and write down the Lagrangian

$$L(Y, \Lambda) = \mathrm{Tr}(YLY^T) + \langle \Lambda, I - YDY^T \rangle,$$

where Λ is the symmetric Lagrangian matrix coefficients, because YDY^T is symmetric. Taking the derivative respect to Y, we have

$$\frac{\partial L}{\partial Y} = 2(LY^T - \Lambda DY^T).$$

Setting the derivative to be 0 and rewriting $\Lambda = U\Sigma U^T$ in terms of its eigenvalue decomposition with Σ being diagonal and U being orthogonal,

$$LY^T - \Lambda DY^T = LY^T - U\Sigma U^T DY^T = LY^T - U\Sigma DU^T Y^T = 0 \implies LY' = DY'\Sigma, \ (*)$$

where $Y' = (UY)^T$ and we have used the fact that matrix multiplication with any diagonal matrix is commutative.

Note that for any Y and any orthogonal matrix O, $\mathrm{Tr}(YLY^T) = \mathrm{Tr}((OY)L(OY)^T)$. $\hat{Y}$ can be determined at most up to rotations. This suggests that we should solve the generalized eigenvalue problem

$$Ly = \lambda Dy. \ (**)$$

Furthermore, we should show that for any y^* satisfying equation (**), $\mathbb{1}^T Dy^* = 0$.

$$\begin{aligned}
& \mathbb{1}^T Dy^* = \mathbb{1}^T \frac{1}{\lambda} Ly = 0 \\
\Longleftarrow \quad & \mathbb{1}^T L = \mathbf{0}^T \\
\Longleftarrow \quad & \forall j, \sum_j L_{ij} = \sum_j D_{ij} - A_{ij} = D_{ii} - \sum_j W_{ij} = \sum_j W_{ij} - \sum_j W_{ij} = 0.
\end{aligned}$$

So we can let $\hat{Y} = [y_1, ..., y_d]^T$ such that $y_1, ..., y_d$ are solutions to the generalized eigenvalue problem corrresponding to the smallest eigenvalues (arranged in ascending order). It is clear that $\hat{Y}D\hat{Y}^T = I$ and $\hat{Y}D\mathbb{1} = \mathbf{0}$.

4 Question 4

In the derivation of LLE, we define

$$M = (I_N - W)^T(I_N - W).$$

Prove that $M\mathbb{1} = \mathbf{0}$.
Proof: Recall that

$$\forall\, i, \ \sum_j W_{ij} = W_{i,\cdot}\mathbb{1} = 1.$$

It follows that

$$W\mathbb{1} = \mathbb{1}.$$

Hence,

$$M\mathbb{1} = (I_N - W)^T (I_N - W)\mathbb{1} = (I_N - W)^T (\mathbb{1} - \mathbb{1}) = \mathbf{0}.$$