局部线性嵌入

耿修瑞

中科院空天信息创新研究院

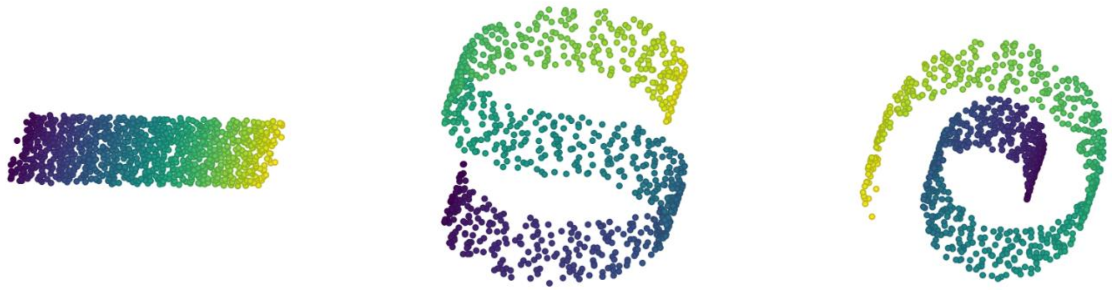
gengxr@sina.com.cn

2025.4

提纲

- 问题背景
- 基本概念
- 局部线性嵌入

问题背景



在上图中，右面的两个曲面都可以由左边的平面通过弯曲或者卷曲的方式得到。也就是说，从内蕴几何的角度，上述三个曲面本质上是同一个几何对象，他们在对应点处都有着相同的高斯曲率。三者相比，对平面上散点的研究显然更为便捷，因为平面为二维欧氏空间，线性代数为其上对象的研究提供了各种便利的工具。而对于右面的两个图，尽管我们可以把他们当做三维欧氏空间中的散点，利用三维线性空间中的各种工具对其进行处理，但处理的结果往往容易受到数据“弯曲”方式的影响，继而并不能真正体现数据本身的内蕴性质。为此，我们需要引入一种“化曲为直”的工具，将弯曲的数据展平，从而降低最终处理结果对数据“弯曲”方式的依赖。

基本概念

定义1（**拓扑空间**）：拓扑空间是由一个集合和其上定义的拓扑构成的二元组$(X,\mathscr{D})$

最常见的拓扑由一系列开集构成，当给定一个集合后，我们可以定义一个子集族$\mathscr{D}$，该子集族由集合X的子集构成，且满足如下条件：

（1）$X,\varnothing \in \mathfrak{D}$

（2）$A,B \in \mathfrak{D} \Rightarrow A \cap B \in \mathfrak{D}$

（3）$\{A_i\} \in \mathfrak{D} \Rightarrow \cup A_i \in \mathfrak{D}$

则称$\mathscr{D}$为X的**拓扑**，而$\mathscr{D}$中的元素称为X的**开集**。通俗地说，拓扑是一个子集族，该自己族满足：

（1）空集和全集都是开集

（2）有限个开集的交集是开集

（3）任意个开集的并集是开集

基本概念

需要注意的是，拓扑是由人为指定的，任意满足如上定义的子集族都可以称为一个集合的拓扑。比如，对于一个只有两个元素的集合 $X=\{0,1\}$，我们既可以指定其上的拓扑为

$$\mathscr{D}=\{\varnothing,\{0\},\{1\},\{0,1\}\}$$

也可以指定其拓扑为

$$\mathscr{D}=\{\varnothing,\{0,1\}\}$$

可以看到，前者包含了集合的所有子集，该拓扑称作**离散拓扑**，而后者只包含空间和全集，是最简单的一个拓扑，因此称作**平凡拓扑**。

基本概念

$\mathbb{R}^n$上的**开球**：

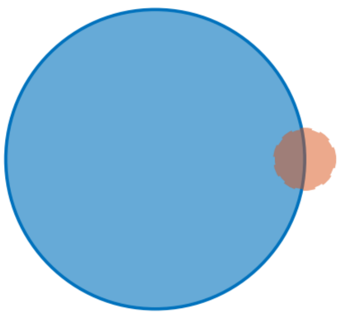
$$B_r(\mathbf{x}) = \left\{ \mathbf{y} \in \mathbb{R}^n \middle| \|\mathbf{y} - \mathbf{x}\| < r \right\}$$

$\mathbb{R}^n$上的拓扑：

$$\mathcal{D}_{std} = \left\{ V \subset \mathbb{R}^n \middle| \forall \mathbf{x} \in V, \exists r > 0, \text{使得} B_r(\mathbf{x}) \subset V \right\}$$

此拓扑称为 $\mathbb{R}^n$ 的**标准拓扑**。

在标准拓扑中，对于任意一子集，当且仅当集中任意一个点总能找到以该点为中心的一个开球，并且该开球完全包含在子集中时，它才是一个开集。而平面上包含边界的圆盘显然就不是一个开集。



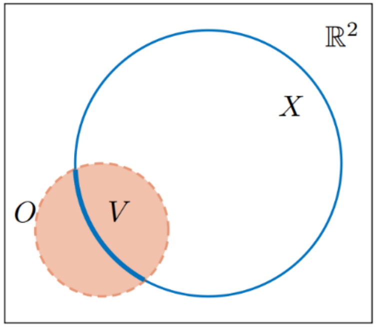
基本概念

在标准拓扑的基础上，我们可以利用诱导拓扑的概念给出线性上任意子集的拓扑，设 X 是 $\mathbb{R}^n$ 中的任意非空子集，那么其上的**诱导拓扑**可以定义为：

$$\mathscr{D}_X = \{V \mid \exists O \in \mathscr{D}_{std}, \text{使得} V = X \cap O\}$$

也就是说，对于 X 中的一个开集 V，它必然是 $\mathbb{R}^n$ 中某个开集与 X 的交集。下面给出了诱导拓扑意义下，平面上圆环的开集示意图。

线段、圆、球、长方形、四面体、立方体等这些几何体都可以看作是线性空间中的一个子集，因此基于诱导拓扑的概念，它们都可以看作是拓扑空间。

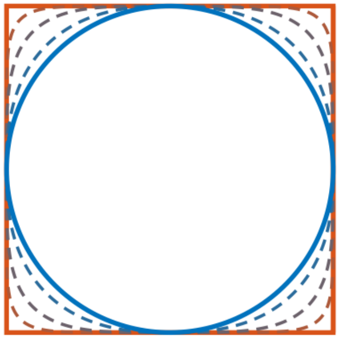


基本概念

定义2（**连续**）设 X,Y 是两个拓扑空间，$f:X\to Y$ 是一个映射。如果对于 Y 中任意开集 V，其原像 $f^{-1}(V)$ 是 X 中的开集，那么称 f 是一个连续映射。

定义3（**同胚**）设 X,Y 是两个拓扑空间，如果存在一个从 X 到 Y 的双射 $f:X\to Y$，且 f 和 f^{-1} 都是连续映射，那么称 X 和 Y 是同胚的。

连续（微积分中的定义）：是指一个函数 $f(x)$ 在其定义域内的每一点 x_0 都连续。也就是说，对于任何给定的 $\xi>0$，都存在一个 $\delta>0$，使得当 x 满足 $|x-x_0|<\delta$ 时，有 $|f(x)-f(x_0)|<\xi$。



基本概念

定义4（**拓扑嵌入**）给定两个拓扑空间 X,Y，以及一个从 X 到 Y 的连续单射 $f:X\to Y$。如果 X 与 $f(X)$ 同胚，则称 f 是一个拓扑嵌入。

基本概念

定义5（**豪斯多夫空间**）设 X 是一个拓扑空间，如果其中的任意两点 x, y 都存在两个不相交的邻域，那么称 X 是一个豪斯多夫空间。

定义6（**流形**）设M是一个豪斯多夫空间，若对其中任意一点 x，都有一个 x 在 M 中的邻域 U 与 $\mathbb{R}^n$ 中的一个开集同胚，那么称M是一个 n 维流形。

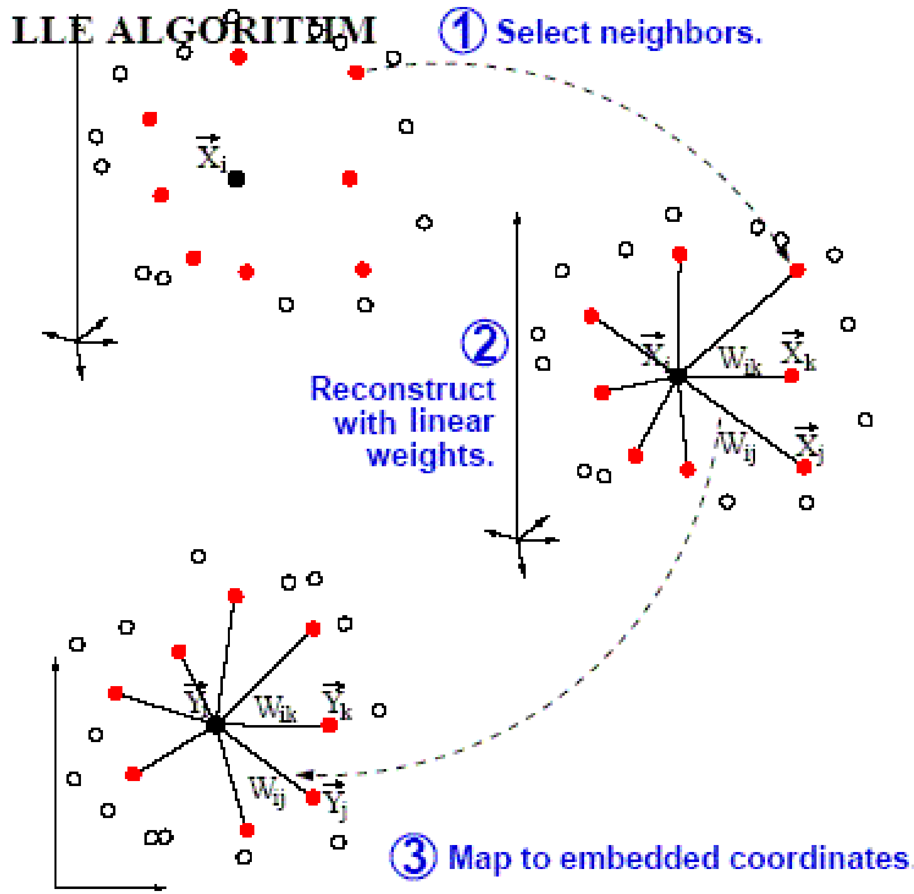
局部线性嵌入

基本过程

1. Compute the neighbors of each data point, $\vec{X}_i$.

2. Compute the weights W_{ij} that best reconstruct each data point $\vec{X}_i$ from its neighbors, minimizing the cost in Equation (1) by constrained linear fits.

3. Compute the vectors $\vec{Y}_i$ best reconstructed by the weights W_{ij}, minimizing the quadratic form in Equation (2) by its bottom nonzero eigenvectors.



Roweis, S. T., & Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. Science, 290(5500), 2323-6.

局部线性嵌入

基本过程2：对每一个点，计算邻域像元加权系数

$$\begin{cases} \min_{\mathbf{w}_i} RSS_i(\mathbf{w}_i) = \left\| \left(\mathbf{x}_i - \sum_{j=1}^{k} w_{ij} \mathbf{x}_{ij} \right) \right\|^2 \\ s.t. \sum_{j=1}^{k} w_{ij} = 1 \end{cases}$$

为什么用和1约束？

$$\mathbf{w}_i = \left(w_{i1}, w_{i2}, \cdots, w_{ik} \right)^{\mathrm{T}}$$

$$\mathbf{Z}_i = \left(\mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_k \right)^{\mathrm{T}} = \left(\mathbf{x}_i - \mathbf{x}_{i1}, \mathbf{x}_i - \mathbf{x}_{i2}, \cdots, \mathbf{x}_i - \mathbf{x}_{ik} \right)^{\mathrm{T}}$$

$$RSS_i(\mathbf{w}_i) = \mathbf{w}_i^{\mathrm{T}} \mathbf{Z}_i \mathbf{Z}_i^{\mathrm{T}} \mathbf{w}_i$$

$$L(\mathbf{w}_i, \lambda) = \frac{1}{2} \mathbf{w}_i^{\mathrm{T}} \mathbf{Z}_i \mathbf{Z}_i^{\mathrm{T}} \mathbf{w}_i - \lambda \left(\mathbf{1}^{\mathrm{T}} \mathbf{w}_i - 1 \right)$$

$$\mathbf{w}_i = \frac{\left(\mathbf{Z}_i \mathbf{Z}_i^{\mathrm{T}} \right)^{-1} \mathbf{1}}{\mathbf{1}^{\mathrm{T}} \left(\mathbf{Z}_i \mathbf{Z}_i^{\mathrm{T}} \right)^{-1} \mathbf{1}}$$

局部线性嵌入

基本过程3：降维，重构

$$\begin{cases} \min\limits_{\mathbf{y}} \Phi(\mathbf{y}) = \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{k} w_{ij} y_{ij} \right)^2 \\ s.t.\ \mathbf{y}^T \mathbf{y} = 1 \end{cases}$$

$$\Phi(\mathbf{y}) = \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{k} w_{ij} y_{ij} \right)^2 = \sum_{i=1}^{n} \left(y_i^2 - y_i \sum_{j=1}^{k} w_{ij} y_{ij} - \left(\sum_{j=1}^{k} w_{ij} y_{ij} \right) y_i + \left(\sum_{j=1}^{k} w_{ij} y_{ij} \right)^2 \right)$$

$$= \mathbf{y}^{\mathrm{T}}\mathbf{y} - \mathbf{y}^{\mathrm{T}}(\mathbf{W}\mathbf{y}) - (\mathbf{W}\mathbf{y})^{\mathrm{T}}\mathbf{y} + (\mathbf{W}\mathbf{y})^{\mathrm{T}}(\mathbf{W}\mathbf{y}) = ((\mathbf{I}-\mathbf{W})\mathbf{y})^{\mathrm{T}}((\mathbf{I}-\mathbf{W})\mathbf{y})$$

$$= \mathbf{y}^{\mathrm{T}}(\mathbf{I}-\mathbf{W})^{\mathrm{T}}(\mathbf{I}-\mathbf{W})\mathbf{y} = \|\mathbf{y}-\mathbf{W}\mathbf{y}\|^2$$

$\mathbf{W} = (w_{ij})$，其中当$\mathbf{x}_j$是$\mathbf{x}_i$ 的邻域点的时候w_{ij} 如上，否则$w_{ij}=0$

$$\mathbf{M} = (\mathbf{I}-\mathbf{W})^{\mathrm{T}}(\mathbf{I}-\mathbf{W}) \qquad \boxed{\mathbf{M}\mathbf{y} = \lambda \mathbf{y}}$$

M的最小特征值为0，所以要剔除第一个特征向量

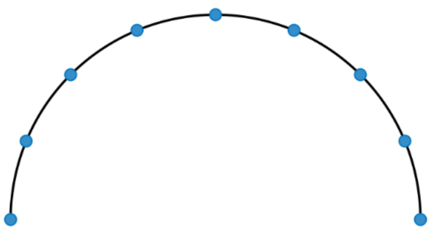
局部线性嵌入

例1.利用局部线性嵌入计算如下数据嵌入到R中的结果（k=4）

$$\mathbf{X}=\begin{bmatrix} -1 & -0.924 & -0.707 & -0.383 & 0 & 0.383 & 0.707 & 0.924 & 1 \\ 0 & 0.383 & 0.707 & 0.924 & 1 & 0.924 & 0.707 & 0.383 & 0 \end{bmatrix}$$

对于第一个数据点，我们可以找到其最近的 4 个数据点，即第 2、3、4、5 个数据点

$$\mathbf{X}_1=\begin{bmatrix} -0.924 & -0.707 & -0.383 & 0 \\ 0.383 & 0.707 & 0.924 & 1 \end{bmatrix}$$



相应地，

$$\mathbf{Z}_1=\begin{bmatrix} -0.076 & -0.293 & -0.617 & -1 \\ -0.383 & -0.707 & -0.924 & -1 \end{bmatrix}^{\mathrm{T}}$$

可以得到

$$\mathbf{w}_1=\begin{bmatrix} 0.633 & 0.732 & 0.282 & -0.647 \end{bmatrix}^{\mathrm{T}}$$

因此，**W** 的第一行元素为

局部线性嵌入

因此，$\mathbf{W}$ 的第一行元素为

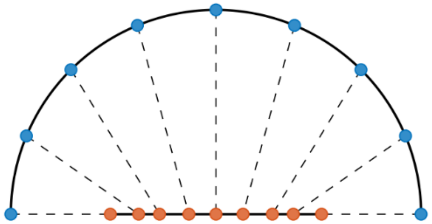
$$\begin{bmatrix} 0 & 0.633 & 0.732 & 0.282 & -0.647 & 0 & 0 & 0 & 0 \end{bmatrix}$$

对所有的数据点都进行上述运算，可以得到权重矩阵为

$$\mathbf{W} = \begin{bmatrix} 0 & 0.633 & 0.732 & 0.282 & -0.647 & 0 & 0 & 0 & 0 \\ 0.918 & 0 & -0.379 & -0.161 & 0.621 & 0 & 0 & 0 & 0 \\ 0.397 & 0.103 & 0 & 0.103 & 0.397 & 0 & 0 & 0 & 0 \\ 0 & 0.397 & 0.103 & 0 & 0.103 & 0.397 & 0 & 0 & 0 \\ 0 & 0 & 0.397 & 0.103 & 0 & 0.103 & 0.397 & 0 & 0 \\ 0 & 0 & 0 & 0.397 & 0.103 & 0 & 0.103 & 0.397 & 0 \\ 0 & 0 & 0 & 0 & 0.397 & 0.103 & 0 & 0.103 & 0.397 \\ 0 & 0 & 0 & 0 & 0.621 & -0.161 & -0.379 & 0 & 0.918 \\ 0 & 0 & 0 & 0 & -0.647 & 0.282 & 0.732 & 0.633 & 0 \end{bmatrix}$$

令 $\mathbf{M} = (\mathbf{I} - \mathbf{W})^{\mathrm{T}} (\mathbf{I} - \mathbf{W})$，计算其第二小特征值对应的特征向量，可得到嵌入结果为

$$\mathbf{y} = \begin{bmatrix} -0.515 & -0.377 & -0.275 & -0.132 & 0 & 0.132 & 0.275 & 0.377 & 0.515 \end{bmatrix}^{\mathrm{T}}$$



局部线性嵌入

在基本过程2中，$\mathbf{Z}_i\mathbf{Z}_i^{\mathrm{T}}$一般都是奇异的，此时，局部表出系数的求解该如何处理？

$$\mathbf{w}_i = \frac{\left(\mathbf{Z}_i\mathbf{Z}_i^{\mathrm{T}}\right)^{-1}\mathbf{1}}{\mathbf{1}^{\mathrm{T}}\left(\mathbf{Z}_i\mathbf{Z}_i^{\mathrm{T}}\right)^{-1}\mathbf{1}}$$

1. 岭回归思路（讨论？）

2. 伪逆代替逆（讨论？）

局部线性嵌入

岭回归方法在解方程组时的问题：

$$\begin{cases} 2x_1 + x_2 + x_3 = 3 \\ x_1 + 2x_2 + x_3 = 3 \end{cases}$$

$$\mathbf{A} = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$$

$$\mathbf{x} = \left(\mathbf{A}^{\mathrm{T}}\mathbf{A} + 0.1\mathbf{I}\right)^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{b}$$

$$\mathbf{Ax} = \mathbf{A}\left(\mathbf{A}^{\mathrm{T}}\mathbf{A} + 0.1\mathbf{I}\right)^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{b} = \begin{bmatrix} 3.0073 \\ 2.9427 \end{bmatrix} \neq \begin{bmatrix} 3 \\ 3 \end{bmatrix}$$

局部线性嵌入

伪逆法在解方程组时的问题：

$$\begin{cases}5x_1+7x_2=0.5\\7x_1+10x_2=0.7\end{cases},\quad\begin{cases}5x_1+7x_2=0.51\\7x_1+10x_2=0.69\end{cases}$$

这两个方程组的精确解分别为：

$$\begin{bmatrix}0.1\\0\end{bmatrix},\ \begin{bmatrix}0.27\\-0.12\end{bmatrix}$$

可见常数向量的微小差别便引起解的很大变动，其原因在于该方程组的系数矩阵的条件数非常大，即其最大奇异值和最小奇异值相差悬殊，此时对应的方程为病态方程。

局部线性嵌入

此外，无论是岭回归方法，还是伪逆法，均只能给出方程或模型的一个解。事实上，当 $\mathbf{z}_i\mathbf{z}_i^{\mathrm{T}}$ 不可逆时，方程或模型应该存在无穷多组解。于是，便有如下改进的局部线性嵌入方法。

局部线性嵌入

改进的局部线性嵌入：

首先将过程2的优化模型改写为

$$\begin{cases} \min\limits_{\mathbf{w}_i} \left\| \left(\mathbf{x}_i - \sum_{j=1}^{k} w_{ij} \mathbf{x}_{ij} \right) \right\|^2 = \left\| \mathbf{Z}_i^{\mathrm{T}} \mathbf{w}_i \right\|^2 = \mathbf{w}_i^{\mathrm{T}} \mathbf{R}_i \mathbf{w}_i \\ s.t.\ \mathbf{1}^{\mathrm{T}} \mathbf{w}_i = 1 \end{cases}$$

其中$\mathbf{R}_i = \mathbf{Z}_i \mathbf{Z}_i^{\mathrm{T}}$，我们发现当$\mathbf{w}_i$取作$\mathbf{R}_i$的核空间的任意向量时，总可以使得优化模型的目标函数为0.假设其核空间的维数为s，我们可以取其中的一组基，$\mathbf{v}_i^{(1)}, \mathbf{v}_i^{(2)}, \cdots, \mathbf{v}_i^{(s)}$，则对于每一个$\mathbf{x}_i$，都可以构建$s$个局部表出系数向量$\mathbf{w}_i^{(1)}, \mathbf{w}_i^{(2)}, \cdots, \mathbf{w}_i^{(s)}$

其中

$$\mathbf{w}_i^{(k)} = \frac{\mathbf{v}_i^{(k)}}{\mathbf{1}^{\mathrm{T}} \mathbf{v}_i^{(k)}},\ k = 1, 2, \cdots, s$$

局部线性嵌入

改进的局部线性嵌入：

对于每一组表出系数$\left\{\mathbf{w}_i^{(k)}, i=1,2,\cdots,n\right\}$，都可以按照局部线性嵌入的构建方式构建一个权重矩阵$\mathbf{W}_k$，相应地，根据局部线性嵌入的过程3，可有如下模型

$$\begin{cases}\min\limits_{\mathbf{y}}\left\|\mathbf{y}-\mathbf{W}_k\mathbf{y}\right\|^2=\mathbf{y}^{\mathrm{T}}\mathbf{M}_k\mathbf{y}\\ s.t.\ \mathbf{y}^{\mathrm{T}}\mathbf{y}=1\end{cases}$$

其中$\mathbf{M}_k=\left(\mathbf{I}-\mathbf{W}_k\right)^{\mathrm{T}}\left(\mathbf{I}-\mathbf{W}_k\right)$，则对于所有$s$组表出系数有

$$\sum_{i=1}^{s}\mathbf{y}^{\mathrm{T}}\mathbf{M}_k\mathbf{y}=\mathbf{y}^{\mathrm{T}}\left(\sum_{i=1}^{s}\mathbf{M}_k\right)\mathbf{y}=\mathbf{y}^{\mathrm{T}}\mathbf{M}\mathbf{y},$$

然后利用$\mathbf{M}$对数据进行局部线性嵌入

讨论：

1. 多维尺度变换（MDS）：保持点之间的欧氏距离
2. 等距特征映射（Isomap）：保持点之间的测地距离
3. 拉普拉斯映射：保留数据的局部邻接结构
4. 随机邻域嵌入（t-SNE）：保持邻域的概率分布
5. 黑森局部线性嵌入（HLLE）：保持局部二阶结构
6. 曲面的法曲率与高斯曲率
7. 高斯绝妙定理与曲面的可展性

谢　谢

耿修瑞

中科院空天信息创新研究院

gengxr@sina.com.cn