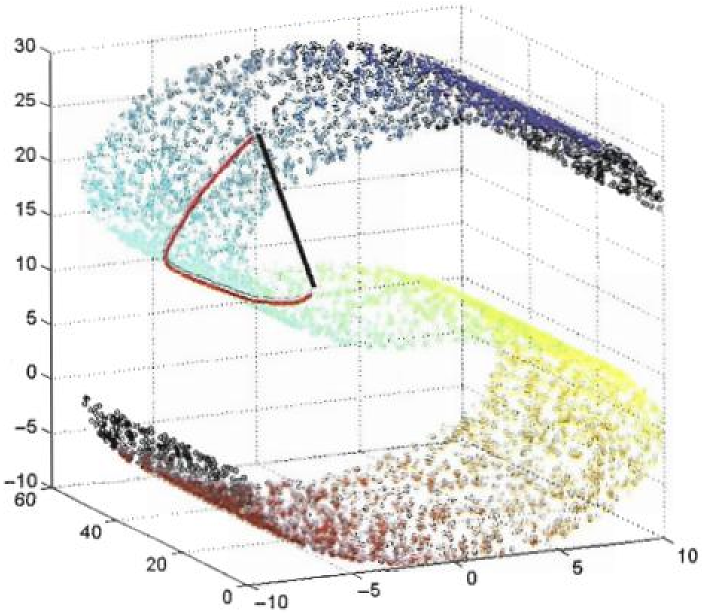
等度量映射（Isomap）

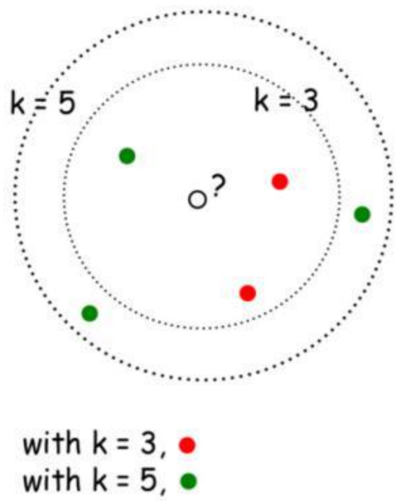
华南师范大学 朱致鹏

之前学习过 MDS 和 PCA，能将高维输数据降维到低维。其中 PCA 基本思想是通过线性变换尽可能地保留方差大的数据量，而 MDS 基本思想是在低维嵌入空间中尽量保持原始数据任意两点之间的欧式距离。但是，对于包含非线性结构的数据集而言，这两种方法往往是无效的。

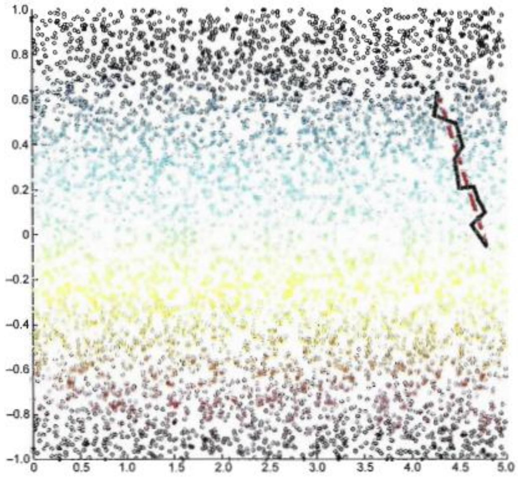



在高维空间中，如上图所示。若采用 MDS 降维，则两点的欧式距离为黑色直线，但实际上红色曲线才是距离最短的路径，即 S 曲面上的测地线，测地线距离是两点之间的本真距离。显然，在高维空间中计算直线距离是不恰当的。

因此，两点之间的距离应该为测地距离。我们可以在每个点基于欧式距离找出其近邻点，建立一个近邻连接图，图中近邻点之间存在连接，而非近邻点之间不存在连接，如下图所示。



因此计算高维空间中两个点的测地距离就近似转变成计算近邻连接图上两点之间的最短路径问题。




如上图所示，红色线段为实际的测地距离，而黑色线段为近邻连接图两点之间的欧式距离，可以看出近邻距离能获得测地线距离很好的近似。

而计算近邻连接图上两点的最短路径可采用 Dijkstra 算法，在得到两点之间的距离后，再通过 MDS 降维获得低维空间中的坐标。

以下是算法流程：

**输入:** 样本集 $D=\{\boldsymbol{x}_1,\boldsymbol{x}_2,\ldots,\boldsymbol{x}_m\}$;
近邻参数 $k$;
低维空间维数 $d'$.

**过程:**

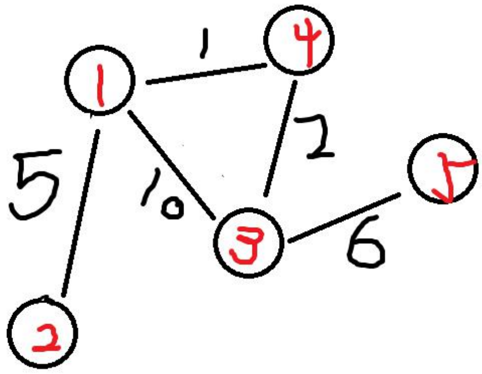
1: **for** $i=1,2,\ldots,m$ **do**
2: 确定 $\boldsymbol{x}_i$ 的 $k$ 近邻;
3: $\boldsymbol{x}_i$ 与 $k$ 近邻点之间的距离设置为欧氏距离, 与其他点的距离设置为无穷大;
4: **end for**
5: 调用最短路径算法计算任意两样本点之间的距离 $\mathrm{dist}(\boldsymbol{x}_i,\boldsymbol{x}_j)$;
6: 将 $\mathrm{dist}(\boldsymbol{x}_i,\boldsymbol{x}_j)$ 作为 MDS 算法的输入;
7: **return** MDS 算法的输出

**输出:** 样本集 $D$ 在低维空间的投影 $Z=\{\boldsymbol{z}_1,\boldsymbol{z}_2,\ldots,\boldsymbol{z}_m\}$.

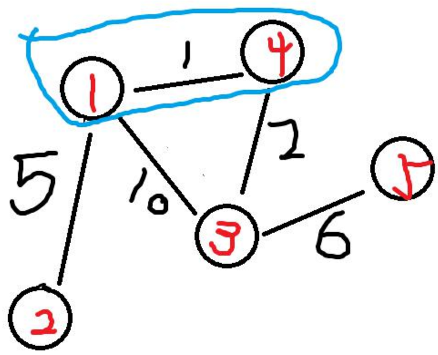
附录：

Dijkstra 算法：

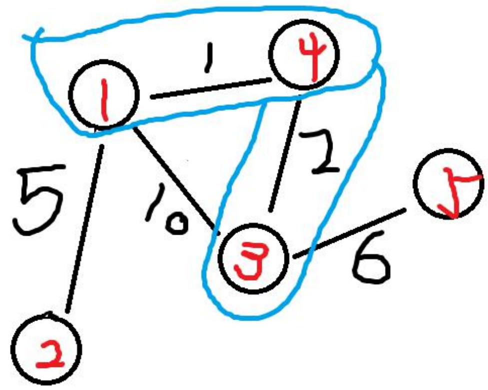
假设要求从①到其他点的最短路径



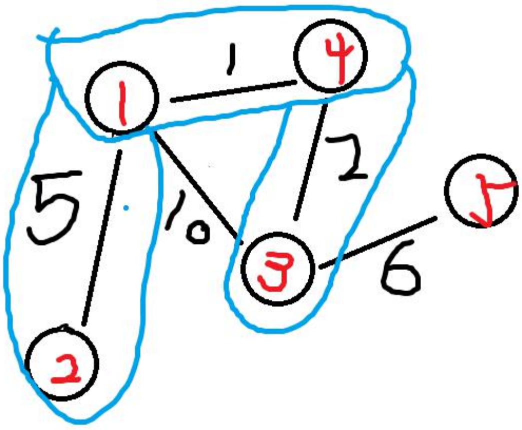

1. 从节点 1 出发，找到与 1 相邻的所有节点距离最近的一个，把这个节点标记为访问过。

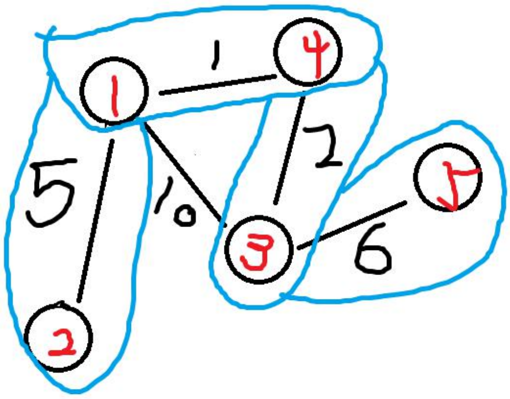


2. 找到与最短路相邻的所有点距离最近的一个（与节点 1 和 4 相邻的点）



3. 重复步骤 2






4. 最后包含所有点，上图为最短路径图。