



# 《机器学习公式详解》 (南瓜书)

## 第10章 降维和度量学习(下)

本节主讲：秦州

南瓜书对应章节：10.5 10.6

### 1.流形学习

- 等度量映射
- 局部线性嵌入

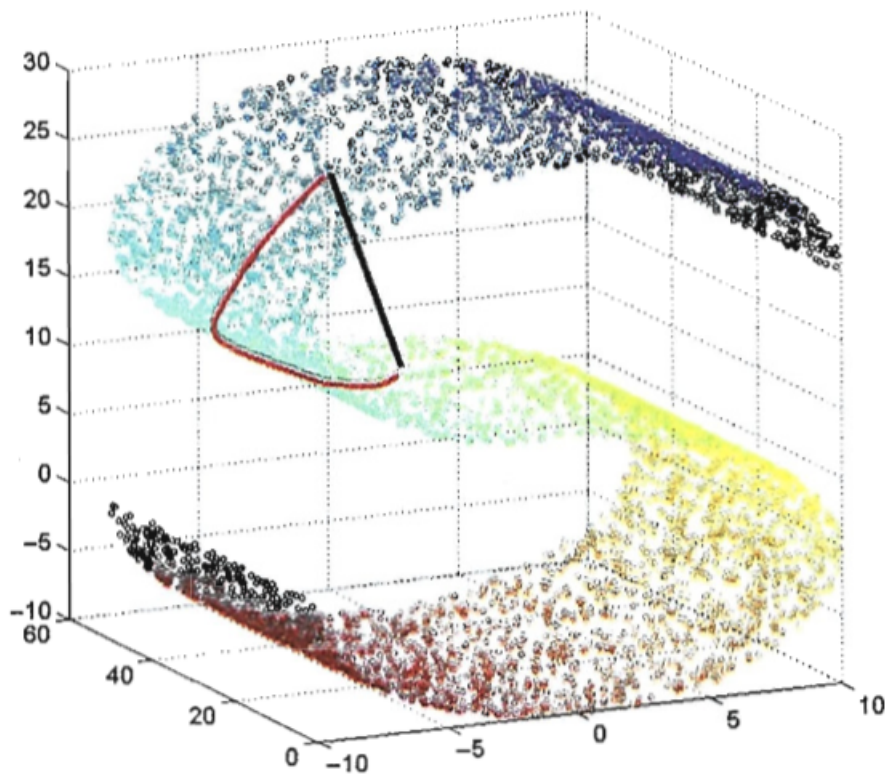
### 2.度量学习

“流形”：在局部具有欧式空间的性质。

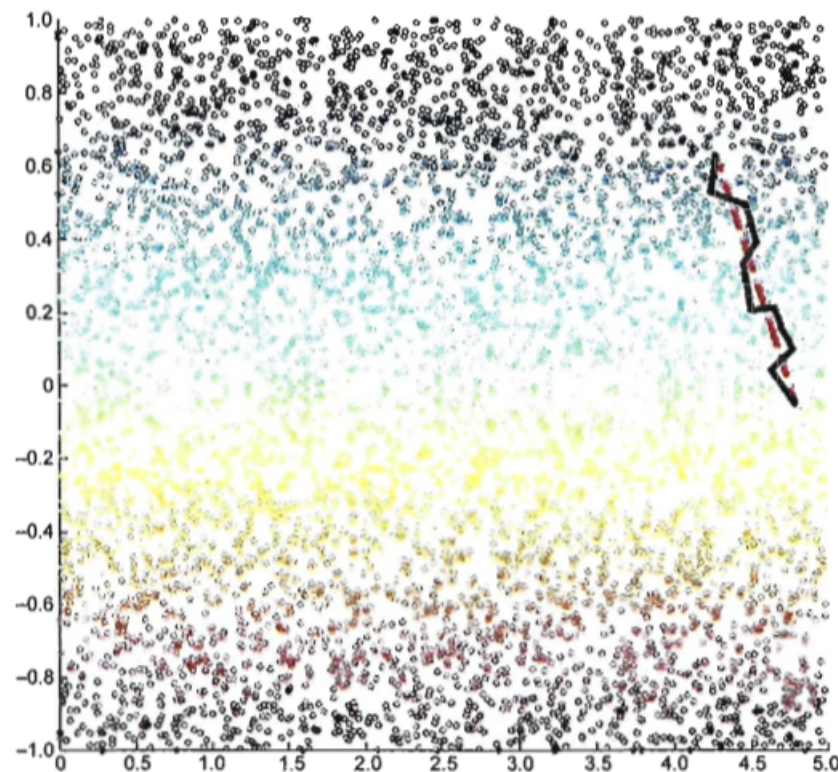
虽然样本在高维空间中的分布看起来非常复杂，但是只要它们在局部上仍然具有欧式空间的性质，因此可以在局部建立降维映射关系，然后再设法将局部映射关系推广到全局。

## 等度量映射(Isometric Mapping)

保持近邻样本之间的距离，因为测地线距离和高维空间的直线距离是不相等的



(a) 测地线距离与高维直线距离



(b) 测地线距离与近邻距离

## 等度量映射2

利用流形在局部上与欧式空间同胚的性质，找出每个点的欧式近邻，建立近邻的链接图，计算两个点之间的测地线距离转化为计算近邻链接图上两点之间的最短路径问题。

---

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

过程:

- 1: **for**  $i = 1, 2, \dots, m$  **do**
- 2:   确定  $\mathbf{x}_i$  的  $k$  近邻;
- 3:    $\mathbf{x}_i$  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  $\text{dist}(\mathbf{x}_i, \mathbf{x}_j)$ ;
- 6: 将  $\text{dist}(\mathbf{x}_i, \mathbf{x}_j)$  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  $Z = \{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_m\}$ .

---

图 10.8 Isomap 算法

### 1. 近邻图如何构建？

- $k$ 近邻图：指定 $k$ 个点为近邻点
- $\epsilon$ 近邻图：指定距离阈值 $\epsilon$

### 2. 对新样本如何变换

- 构建高维坐标和低维坐标的回归器

## 局部线性嵌入(Locally Linear Embedding)

保持邻域内样本之间的线性关系。即假设样本点 $\mathbf{x}_i$ 的坐标能通过它的邻域样本 $\mathbf{x}_j, \mathbf{x}_k, \mathbf{x}_l$ 的坐标通过线性组合而重构得到：

$$\mathbf{x}_i = w_{ij}\mathbf{x}_j + w_{ik}\mathbf{x}_k + w_{il}\mathbf{x}_l$$

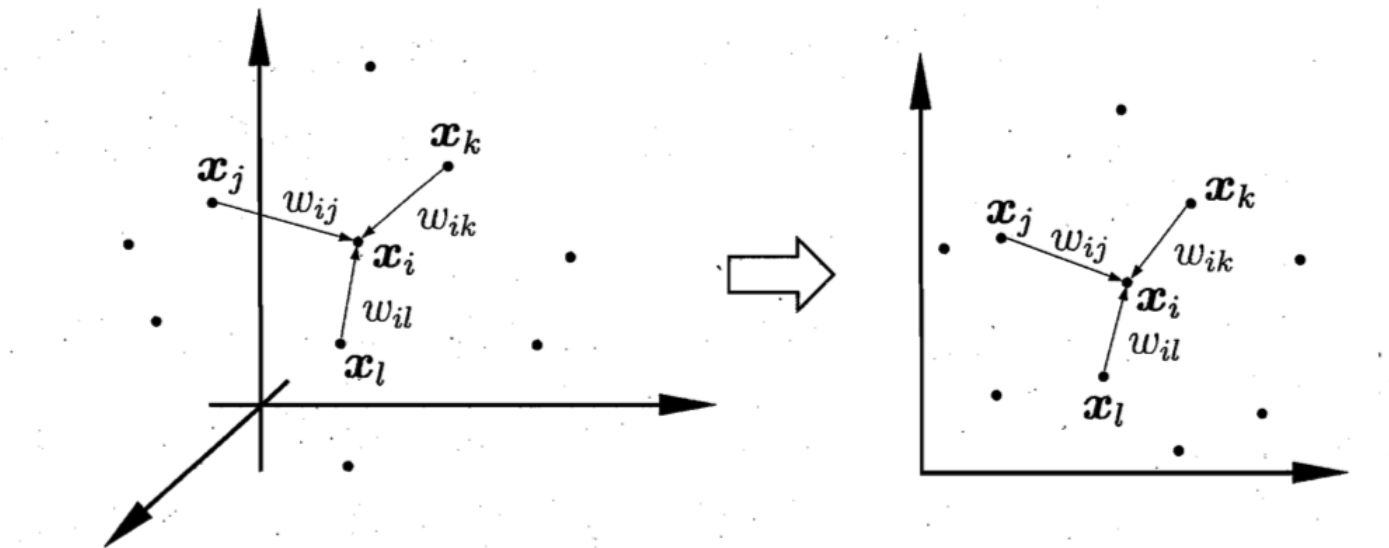


图 10.9 高维空间中的样本重构关系在低维空间中得以保持

每个空间其实对应了样本属性上定义的距离度量，寻找合适的空间的本质是寻找一个合适的距离度量。度量学习的基本动机是学习得到一个合理的“度量”。

之前我们介绍了很多计算距离的表达式（欧式距离等），但是“它们”都是固定的，没有可调节的参数，因此要先做一个推广。

欧式距离定义为

$$\text{dist}_{\text{ed}}^2(\mathbf{x}_i, \mathbf{x}_j) = \|\mathbf{x}_i - \mathbf{x}_j\|_2^2 = \text{dist}_{ij,1}^2 + \text{dist}_{ij,2}^2 + \dots + \text{dist}_{ij,d}^2$$

假设不同属性的重要性不同，则可引入属性权重 $w$

$$\begin{aligned}\text{dist}_{\text{wed}}^2(\mathbf{x}_i, \mathbf{x}_j) &= \|\mathbf{x}_i - \mathbf{x}_j\|_2^2 = w_1 \cdot \text{dist}_{ij,1}^2 + w_2 \cdot \text{dist}_{ij,2}^2 + \dots + w_d \cdot \text{dist}_{ij,d}^2 \\ &= (\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{W} (\mathbf{x}_i - \mathbf{x}_j)\end{aligned}$$

其中  $w_i \geq 0$ ,  $\mathbf{W} = \text{diag}(\mathbf{w})$  是一个对角矩阵,  $(\mathbf{W})_{ii} = w_i$



通过以上方式，我们引入了变量 $\mathbf{W}$ ，但是这样引入存在一个问题，因为 $\mathbf{W}$ 的非对角线元素为0，因此坐标轴是正交的，意味着属性之间无关。为了更加符合实际，引入属性之间的关联性，可将 $\mathbf{W}$ 替换成一个半正定对称矩阵 $\mathbf{M}$ ，且 $\mathbf{M} = \mathbf{P}^T \mathbf{P}$ ，于是就得到了马氏距离

$$\text{dist}_{\text{mah}}^2(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{M} (\mathbf{x}_i - \mathbf{x}_j) = (\mathbf{P} (\mathbf{x}_i - \mathbf{x}_j))^T (\mathbf{P} (\mathbf{x}_i - \mathbf{x}_j)) = \|\mathbf{x}_i - \mathbf{x}_j\|_{\mathbf{M}}^2$$

如何学习 $\mathbf{M}$ ？ $\mathbf{M}$ 的学习与下游任务有关，比如如果我们希望提高近邻分类器的性能，则可将 $\mathbf{M}$ 嵌入近邻分类器的评价指标里去，下面以近邻成分分析(Neighborhood Component Analysis) 为例解释这个过程。

前面介绍过得近邻算法（kNN）以邻居投票的形式决定自身的标签，假设将一邻一票替换成概率投票，即对于样本 $\mathbf{x}_i$ ，他的邻居 $\mathbf{x}_j$ 给他投票

$$p_{ij} = \frac{\exp \left( - \|\mathbf{x}_i - \mathbf{x}_j\|_{\mathbf{M}}^2 \right)}{\sum_l \exp \left( - \|\mathbf{x}_i - \mathbf{x}_l\|_{\mathbf{M}}^2 \right)}$$

那么样本 $\mathbf{x}_i$ 被预测正确的概率为

$$p_i = \sum_{j \in \Omega_i} p_{ij}$$

其中 $\Omega_i$ 表示和 $\mathbf{x}_i$ 标签一致的样本集合（不包含 $\mathbf{x}_i$ 自身）。

整个样本集上的错误率为

$$\epsilon = 1 - \sum_{i=1}^m p_i = 1 - \sum_{i=1}^m \sum_{j \in \Omega_i} p_{ij}$$

带入 $p_{ij}$ 的计算公式，那么为了最小化分类错误率，有

$$\min_{\mathbf{P}} 1 - \sum_{i=1}^m \sum_{j \in \Omega_i} \frac{\exp \left( - \left\| \mathbf{P}^T \mathbf{x}_i - \mathbf{P}^T \mathbf{x}_j \right\|_2^2 \right)}{\sum_l \exp \left( - \left\| \mathbf{P}^T \mathbf{x}_i - \mathbf{P}^T \mathbf{x}_l \right\|_2^2 \right)}$$

上式可以用随机梯度下降的方式求解。

下一节：特征选择和稀疏学习

特征选择

稀疏学习

西瓜书对应章节：第11章

欢迎加入【南瓜书读者交流群】，我们将在群里进行答疑、勘误、本次直播回放、本次直播PPT发放、下次直播通知等最新资源发放和活动通知。

加入步骤：

1. 关注公众号【Datawhale】，发送【南瓜书】三个字获取机器人“小豚”的微信二维码
2. 添加“小豚”为微信好友，然后对“小豚”发送【南瓜书】三个字即可自动邀请进群

