降维与度量学习

1. KNN 算法

给定测试样本，基于某种距离度量找出最近的 K 个样本

根据投票或平均输出结果

KNN 属于懒惰学习　　训练：保存样本，收到样本后再处理.

相对的就有急切学习.

$$p(err) = 1 - \sum_{c \in y} P(c|x)\,P(c|z)$$

其中 z 即最邻近样本.

满足条件：x 附近任意小 δ 距离都有样本

$$\simeq 1 - \sum_{c \in y} P^2(c|x)$$

$$\le 1 - P^2(c^*|x)$$

c^*：贝叶斯最优分类器结果

$$= [1 + \underline{P(c^*|x)}]\,[1 - P(c^*|x)]$$

$0 \le P \le 1$

$$\le 2\,[1 - P(c^*|x)]$$

KNN 分类器泛化误差不超过贝叶斯最优分类器两倍错误率

⇓

要求足够的样本密度

1维：1000

2维：1000^2

与学习任务相关的只是低维分布.

没有足够的样本 ⟶ 降维.：通过某种数学变换将原始高维属性转化为低维子空间

2. MDS 算法

D：m 个样本的距离矩阵 $\begin{bmatrix} 11 & 12 \\ 12 & 22 \end{bmatrix}$

Z：转化后的空间向量 $\|z_i - z_j\| = D_{ij}$

B：降维后的内积矩阵 $D_{ij}^2 = B_{ii}^2 - 2B_{ij} + B_{jj}^2$

已知 $\sum_{i=1}^{m} z_i = 0 \Rightarrow \sum_{i=1}^{m} b_{ij} = \sum_{j=1}^{m} b_{ij} = 0$，$\sum_{i=1}^{m} d_{ij}^2 = tr(B) + m b_{jj}$，$\sum_{j=1}^{m} d_{ij}^2 = tr(B) + m b_{ii}$，$\sum_{i=1}^{m}\sum_{j=1}^{m} d_{ij}^2 = 2m\,tr(B)$

$d_{i\cdot}^2 = \frac{1}{m}\sum_{j=1}^{m} d_{ij}^2$，$d_{\cdot j}^2 = \frac{1}{m}\sum_{i=1}^{m} d_{ij}^2$，$d_{\cdot\cdot}^2 = \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} d_{ij}^2$

$$b_{ij} = -\frac{1}{2}\left(d_{ij}^2 - d_{i\cdot}^2 - d_{\cdot j}^2 + d_{\cdot\cdot}^2\right)$$

即从距离矩阵到内积矩阵的计算，$B = V\Lambda V^T$，$Z = \Lambda_*^{1/2} V_*^T \in \mathbb{R}^{d^* \times m}$.

小结：1. 计算 $d_{i\cdot}$，$d_{\cdot j}$，$d_{\cdot\cdot}$

2. 计算内积矩阵 B

3. 对 B 作特征值分解

4. 取最大的 d' 个特征值 $\tilde{\Lambda}$，对应特征向量 $\tilde{V}$

样本坐标：$\tilde{V}\tilde{\Lambda}^{1/2}$

3. 线性变换

$Z = W^T x$，W ⇒ 变换矩阵

4. PCA

合适的超平面：
- 重构：样本点到超平面距离足够近
- 可分：投影尽可能分开.

重构 $\hat{x}_i = \sum_{j=1}^{d'} z_{ij} w_j$

$$\Rightarrow \begin{cases} \min\limits_{W} \; -\mathrm{tr}(W^T X X^T W) \\ \text{s.t.} \; W^T W = I. \end{cases}$$

$$XX^T w_i = \lambda_i w_i$$

1. 对样本中心化：$x_i \leftarrow x_i - \frac{1}{m}\sum_{i=1}^{m} x_i$
2. 计算协方差矩阵 XX^T （最大化方差）
3. 对 XX^T 作特征值分解
4. 取前 d' 个特征值对应特征向量 $w_1 \dots w_{d'}$

$$z = W^T x$$

注：d' 如何选择
- 交叉检验
- $\frac{\sum_{i=1}^{d'} \lambda_i}{\sum_{i=1}^{d} \lambda_i} \geq t$，$t$ 可以是 95%

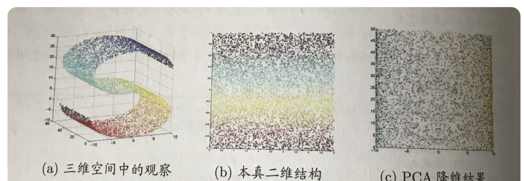

图 10.6 三维空间中观察到的 3000 个样本点，是从本真二维空间中矩形区域采样后以 S 形曲面嵌入，此情形下线性降维会丢失低维结构. 图中数据点的染色显示出低维空间的结构.

5. 核方法的引入

$$\left(\sum_{i=1}^{m} z_i z_i^T\right) w_j = \lambda_j w_j$$

$$w_j = \sum_{i=1}^{m} z_i \frac{z_i^T w_j}{\lambda_j}, \quad \text{令 } \alpha_i^j = \frac{z_i^T w_j}{\lambda_j}$$

$$= \sum_{i=1}^{m} z_i \alpha_i^j, \quad \text{另有 } z_i = \phi(x_i)$$

$$= \sum_{i=1}^{m} \phi(x_i) \alpha_i^j$$

$$\left(\sum_{i=1}^{m} \phi(x_i)\phi(x_i)^T\right) w_j = \lambda_j w_j$$

又有 $\kappa(x_i, x_j) = \phi(x_i)^T \phi(x_j)$

$$\Rightarrow K \sum_{i=1}^{m} \phi(x_i)\alpha_i^j = \lambda_j \sum_{i=1}^{m} \phi(x_i)\alpha_i^j$$

$$K\alpha^j = \lambda_j \alpha^j$$

得出 λ, α

$$z_j = w_j^T \phi(x) = \sum_{i=1}^{m} \alpha_i^j \phi(x_i)^T \phi(x)$$

$$= \sum_{i=1}^{m} \alpha_i^j \kappa(x_i, x)$$

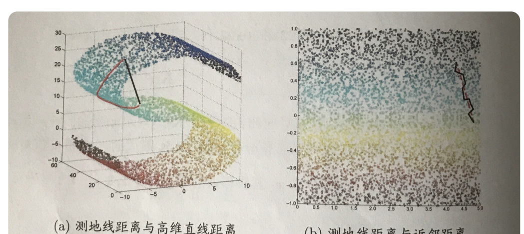

图 10.7 低维嵌入流形上的测地线距离(红色)不能用高维空间的直线距离计算，但能用近邻距离来近似

6. 流形学习

在局部与欧氏空间同胚

测地线：计算本真距离，不能脱离曲面行走.

(1) ISOMAP：等度量映射

solution：对于每个点基于欧氏距离寻找邻近点，邻近点相连到终点

指定邻近点个数，距离阈值

利用 Dijkstra 或 Floyd 计算两点最短距离，输入 MDS 降维.

学习器
- 输入：高维
- 输出：低维

不大理解这里的短路，断路.

(2) LLE：局部线性嵌入，保留样本间的线性关系

① $$\min_{w_1, w_2, \dots, w_m} \sum_{i=1}^{m} \| x_i - \sum_{j \in Q_i} w_{ij} x_j \|_2^2$$

$$\sum_{j \in Q_i} w_{ij} = 1 \quad \Rightarrow \quad w_{ij} = \frac{\sum_{k \in Q_i} C_{jk}^{-1}}{\sum_{l,s \in Q_i} C_{ls}^{-1}}$$

优化w

$$C_{jk} = (x_i - x_j)^T (x_i - x_k)$$

② $$\min_{z_1, z_2, \dots, z_m} \sum_{i=1}^{m} \| z_i - \sum_{j \in Q_i} w_{ij} z_j \|_2^2$$

优化 z

$$\Rightarrow \min_{Z} tr(Z M Z^T) \qquad M = (I - W)^T (I - W)$$

$$s.t. \quad Z Z^T = I$$

③ 对 M 作特征值分解，返回最小的 d' 个特征值

$$Z = \{ z_1, z_2, \dots, z_m \}$$

7. 度量学习

目标：找到合适的低维空间性能优于原始空间

对应样本属性的距离度量

加属性权重的 d：$d^2(x_i, x_j) = w_1 d_{ij,1}^2 + \dots + w_d d_{ij,d}^2$

$$= (x_i - x_j)^T W (x_i - x_j)$$

W通过学习确定，W为对角阵，$(W)_{ii} = w_i$

扩展 ⇒ 属定相交

$$= (x_i - x_j)^T M (x_i - x_j)$$

M为半正定对称矩阵，$M = PP^T$

NCA：近邻成分分析

样本 x_j 对 x_i 分类结果的影响随距离增大而减小.

$$p_{ij} = \frac{\exp(-\|x_i - x_j\|_M^2)}{\sum_l \exp(-\|x_i - x_l\|_M^2)}$$

$$\sum_{i=1}^{m} p_i = \sum_{i=1}^{m} \sum_{j \in \Omega_i} p_{ij} \quad \Rightarrow \quad \max_P \sum_{i=1}^{m} \sum_{j \in \Omega_i} \frac{\exp(-\|P^T x_i - P^T x_j\|_2^2)}{\sum_l \exp(-\|P^T x_i - P^T x_l\|_2^2)}$$

Ω_i：与 x_i 同类的下标集

也可引入必连，勿连

$$\min_M \sum_{(x_i, x_j) \in \mathcal{M}} \|x_i - x_j\|_M^2$$

$$s.t. \sum_{(x_i, x_k) \in C} \|x_i - x_k\|_M \geq 1$$

$$M \succeq 0.$$