WDS

DataScience@Web

机器学习讨论班

2018年暑期

WDS

DataScience@Web

度量学习
(DISTANCE) METRIC LEARNING

刘宾楚

09/12/2018

度量学习的定义：

- 降维的目的：在高维空间中，找到一个合适的低维空间，从而在低维空间中进行学习能够获得更好的性能。
- 降维的实质：每一个不同的空间都对应不同的距离度量，寻找合适的低维空间，实际就是寻找合适的度量。
- 度量学习的基本动机：通过学习，直接得出合适的距离度量。
- 度量学习 与 降维 : 密切相关

距离度量

- 四个条件：
  - $d(x; y) \geq 0$， 非负性
  - $d(x; y) = d(y; x)$， 对称性
  - $d(x; y) = 0 \leftrightarrow x = y$， 同一性/自反性
    - (不存在两个不同实体，其所有属性全部相同)
  - $d(x; z) \leq d(x; y) + d(y; z)$， 三角不等式

距离度量

- l_p 范数 与 l_p 距离：
- l_p 范数： $\left\|x\right\|_p = (\sum_{i=1}^{n}|x_i|^p)^{\frac{1}{p}}$, $p \geq 1$
- 当 $0 \leq p \leq 1$ 时， 不是范数， 不满足三角不等式
- l_0 范数： 通常用作 表示 x 中 非0元素的个数， 但不是有效的范数。
- 当 $p \geq 1, l_p$ 范数 可以用作距离： $d_p(x,y) = \left\|x-y\right\|_p$
- $p=1$ 街区距离 $p=2$ 欧式距离

距离度量

- 度量学习需要
  - 1. 非固定 2. 具有可调节参数 的距离度量
  - 非标准马氏距离
- 马氏距离由印度统计学家马哈拉诺比斯(P. C. Mahalanobis)提出
- 与欧氏距离不同的是它考虑到各种特性之间的联系。--- 协方差
- 表示数据的协方差距离：通过分解协方差矩阵 得到新的坐标基 （P9）
- 它是一种有效的计算两个未知样本集(分布之间)的相似度的方法

$$d(\vec{x},\vec{y}) = \sqrt{(\vec{x}-\vec{y})^T S^{-1} (\vec{x}-\vec{y})}.$$

距离度量

- 平方欧式距离： 单位阵

$$\mathrm{dist}^2_{\mathrm{ed}}(\boldsymbol{x}_i, \boldsymbol{x}_j) = ||\boldsymbol{x}_i - \boldsymbol{x}_j||_2^2 = dist^2_{ij,1} + dist^2_{ij,2} + \ldots + dist^2_{ij,d} \ ,$$

- 引入权重： 对角阵

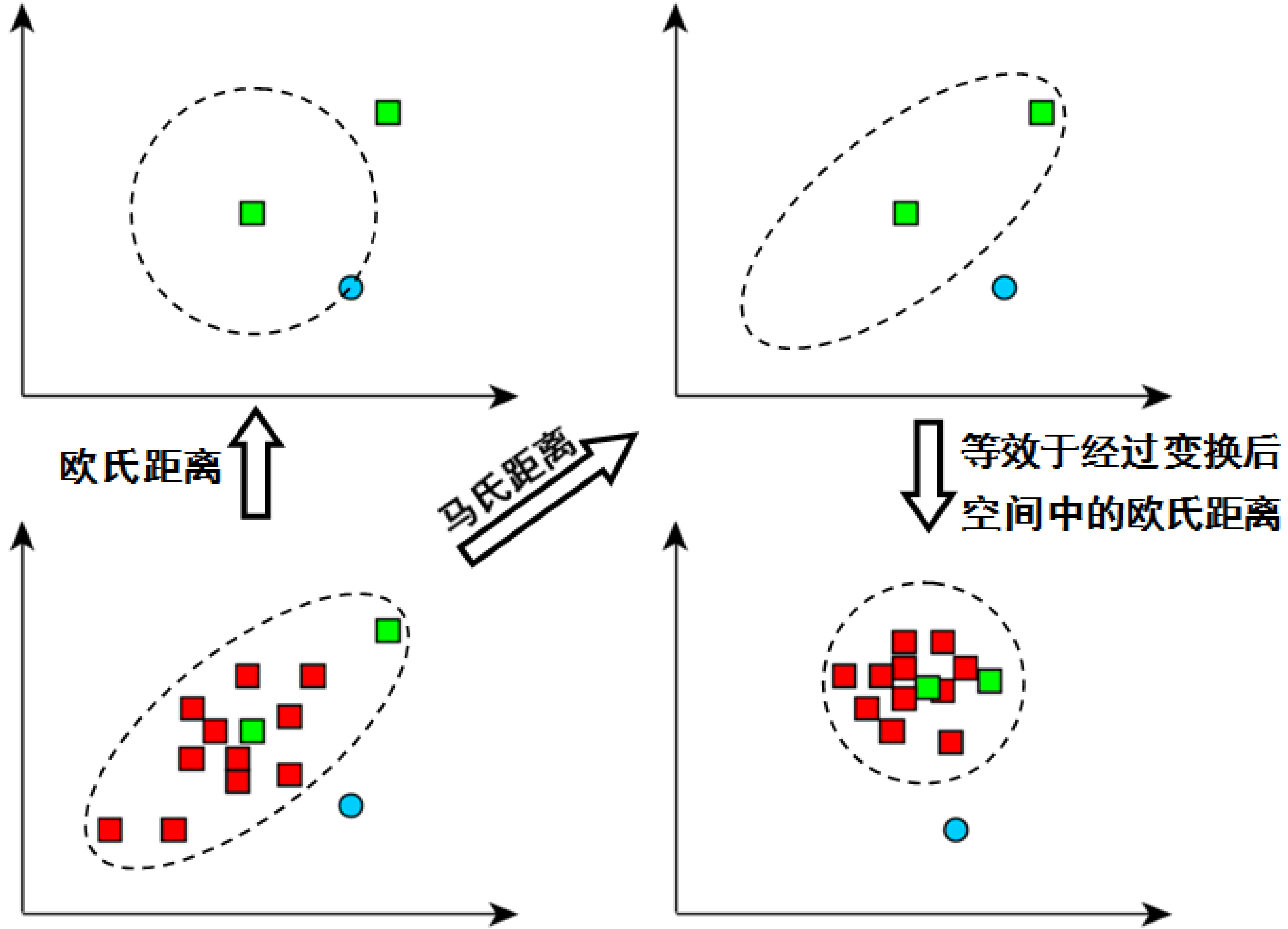
$$\begin{aligned} \mathrm{dist}^2_{\mathrm{wed}}(\boldsymbol{x}_i, \boldsymbol{x}_j) &= ||\boldsymbol{x}_i - \boldsymbol{x}_j||_2^2 = w_1 \cdot dist^2_{ij,1} + w_2 \cdot dist^2_{ij,2} + \ldots + w_d \cdot dist^2_{ij,d} \\ &= (\boldsymbol{x}_i - \boldsymbol{x}_j)^{\mathrm{T}} \mathbf{W} (\boldsymbol{x}_i - \boldsymbol{x}_j) \ , \end{aligned} \tag{10.33}$$

- 非标准马氏距离： 半正定对称阵

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

马氏距离

- 优点：
  - 不受量纲的影响
  - 尺度无关
  - 考虑到维度之间的相互关系
- 缺点：
  - 建立在总体样本的基础上
  - 要求总体样本数大于样本的维数
  - 马氏距离的计算 不稳定


欧氏距离
马氏距离
等效于经过变换后
空间中的欧氏距离

再看 主成分分析 PCA

算法描述：

输入：n维样本集 $D=(x_1, x_2, x_3, \ldots, x_m)$, 要降维到维数d'.

1) 对所有的样本进行中心化: $x_i = x_i - \frac{1}{m}\sum_{i=1}^{m} x_i$;

2) 计算样本的协方差矩阵$\boldsymbol{X}\boldsymbol{X}^{\boldsymbol{T}}$;

3) 对矩阵$\boldsymbol{X}\boldsymbol{X}^{\boldsymbol{T}}$进行特征值分解;

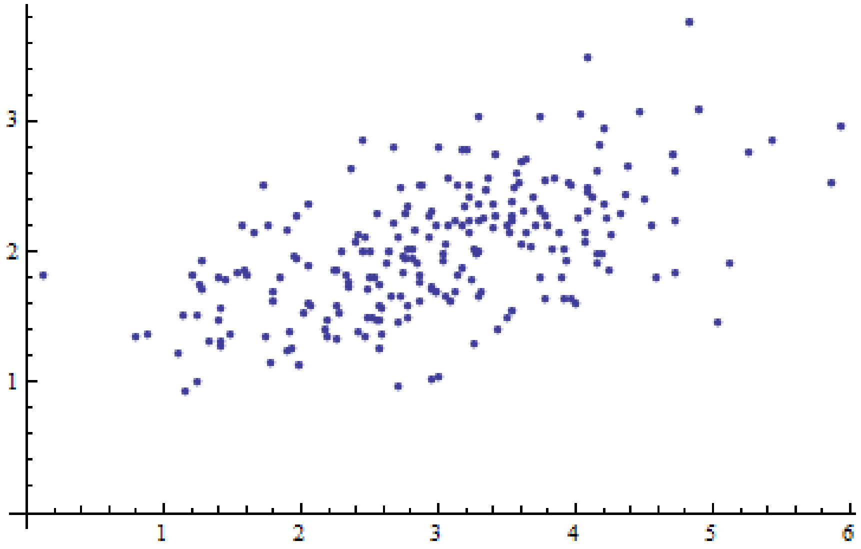
4) 取出最大的d'个特征值对应的特征向量$(w_1, w_2, w_3, \ldots, w_{d'})$

将所有的特征向量标准化后，组成特征向量矩阵$\boldsymbol{W}$;

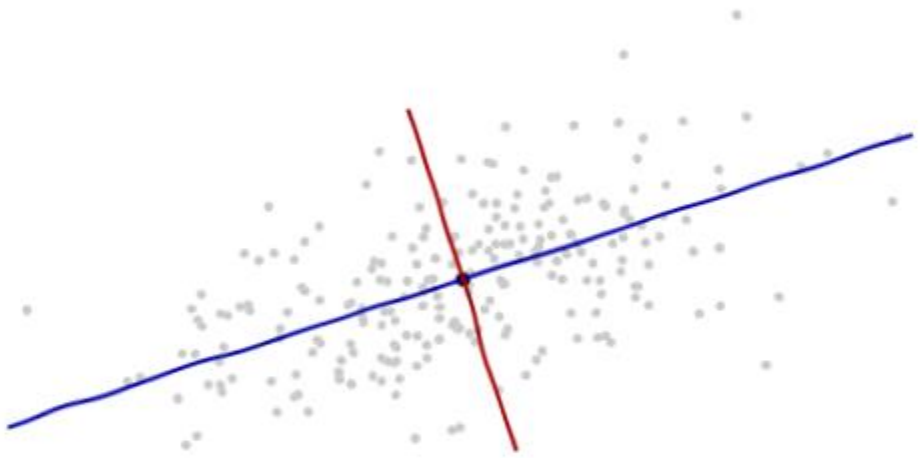
5）对样本集中的每一个样本,转化为新的样本$z_i = \boldsymbol{W}^{\boldsymbol{T}} x_i$

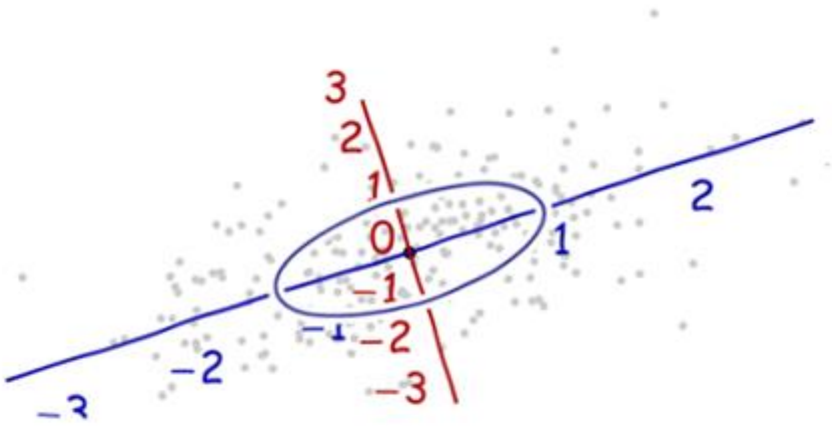
输出：得到输出样本集$D' = (z_1, z_2, \ldots, z_m)$.

马氏距离对应的空间变换

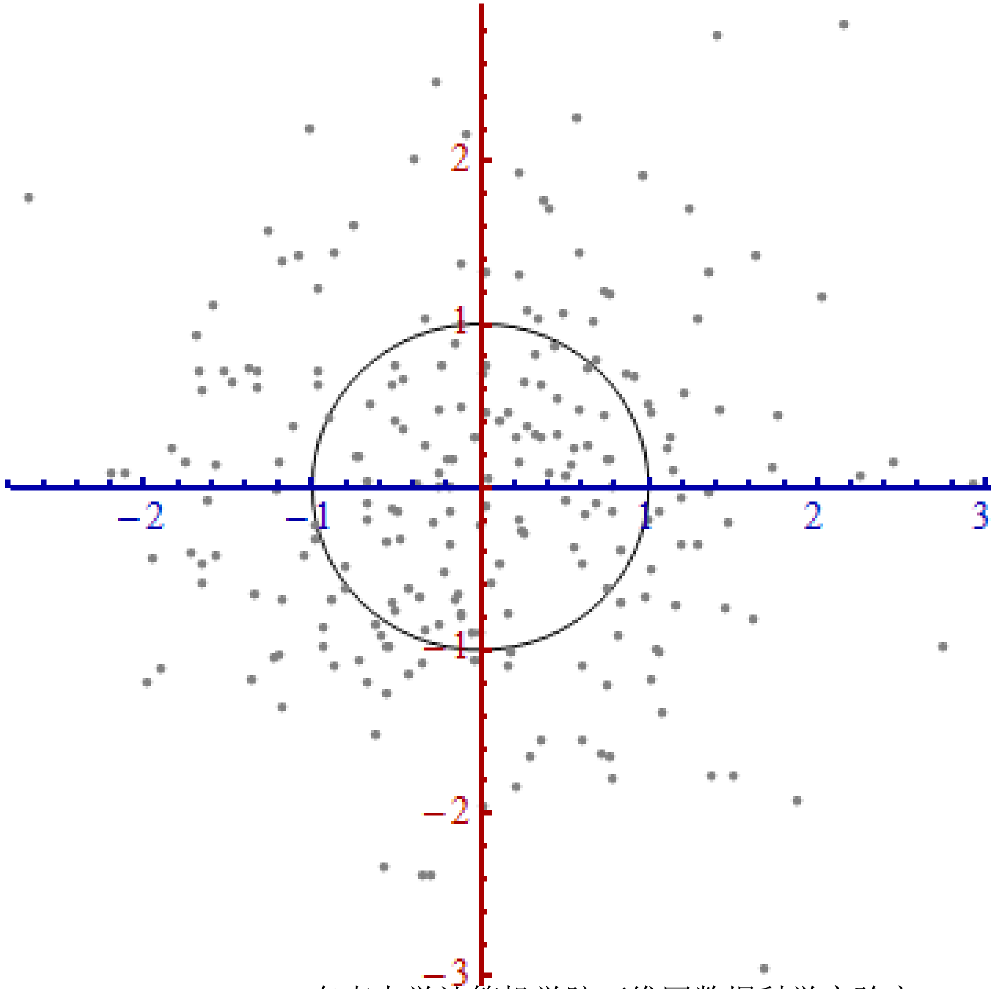


马氏距离对应的空间变换





马氏距离对应的空间变换



近邻成分分析 NCA

- 目标： 寻找合适的度量，提高近邻分类器的性能
- 方法： 将度量矩阵 M 嵌入近邻分类器的评价指标中，通过求解其最优值
- 对于任意样本 x_j，它对于x_i的分类结果的影响为：

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

- 分子: i 和 j 两个样本点的马氏距离
- 分母：所有点 距离 i点的距离之和
- x_j对x_i的影响随着它们之间距离的增大而减小

近邻成分分析 NCA

- x_i的留一法正确率：被自身之外的所有样本分类正确的概率
  - Ω_i 表示与x_i属于相同类别的样本的下标的集合

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

- 整个样本上的留一法正确率：

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

- 度量矩阵M 半正定对称 则必有 正交基 P 使得 $M = PP^T$

- 最优化目标：

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

引入领域知识

- 如果已知某些样本相似，某些样本不相似， 那么可以定义必连约束集合M和勿连约束集合C。
- 思想：希望相似的样本之间距离较小，不相似的样本之间距离较大。
- 最优化目标：

$$\min_{\mathbf{M}} \sum_{(\boldsymbol{x}_i,\boldsymbol{x}_j)\in\mathcal{M}} \|\boldsymbol{x}_i - \boldsymbol{x}_j\|_{\mathbf{M}}^2$$

$$\text{s.t.} \sum_{(\boldsymbol{x}_i,\boldsymbol{x}_k)\in\mathcal{C}} \|\boldsymbol{x}_i - \boldsymbol{x}_k\|_{\mathbf{M}}^2 \geqslant 1\ ,$$

- $M \geq 0$ 表示 M必须是半正定 $\mathbf{M} \succeq 0$,
  - E. P. Xing, A. Y. Ng, M. I. Jordan, and S. Russell. Distance metric learning with application to clustering with sideinformation. In: NIPS, pages 505-512, 2003

度量学习的分类

- http://www.cs.cmu.edu/~liuy/distlearn.htm
- A comprehensive survey on distance metric learning
- An overview of distance metric learning
- 有监督的度量学习：
  - 全局：同时满足所有的必连和勿连约束
    - Relevant Components Analysis (RCA) 线性
    - Discriminative Component Analysis (DCA) 线性
  - 局部：不必满足所有的约束 仅满足附近的约束
    - Local Fisher Discriminant Analysis (LFDA) 线性
    - Neighborhood Component Analysis (NCA) 线性

无监督的度量学习

- 保留全局信息
  - Principal Component Analysis(PCA) 线性
  - Multidimensional Scaling(MDS) 线性
  - ISOMAP 非线性
- 保留局部信息
  - Laplacian Eigenamp (LE) 非线性
  - Locality Preserving Projections (LPP) 线性
  - Locally Linear Embedding (LLE) 非线性
  - Neighborhood Preserving Embedding (NPE) 线性

度量学习 与 降维

- 不同的度量学习方法 针对 不同的目标 获得 “好” 半正定 距离度量矩阵M，
- 通过对 度量矩阵 进行 特征值分解
- 总能够找到一组 正交基 作为新空间中的坐标基
  - 若M是一个低秩矩阵，那么
  - 正交基的数目小于矩阵维数 --- 降维
- 一个问题 从不同角度考虑，一体两面

度量学习

- l_p 距离 与 马氏距离
- 马氏距离 对应的 投影变换 – PCA
- 近邻成分分析 NCA
- 引入领域知识的度量学习
- 度量学习 同 降维 的关系
- 度量学习 的拓扑分类

谢 谢