

## Dimensionality Reduction 降维

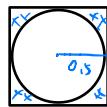
### 一、背景

降维是抑制过拟合的手段之一。

维度灾难：每增加一个特征，若加二值，它也会以指数的速度增长。

几何角度：

例一：



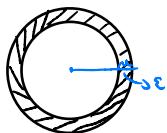
当维度增加时

$$V_{\text{超立方体}} = 1 \quad \text{当 } D \text{ 增大时, } V_{\text{超立方体}} \rightarrow 0.$$

$$V_{\text{超球体}} = \pi^{\frac{D}{2}}$$

故在高维空间中，立方体几乎是空的，所以在高维空间中，样本分布在角落中，造成样本的稀疏性，并且分布不均，这种分布很难做分类。

内部的圆表示  
示空间内部



$$V_{\text{球}} = \pi^{\frac{D}{2}} = K$$

$$V_{\text{环}} = V_{\text{球}} - V_{\text{内}} = K - K \cdot (1-\epsilon)^2$$

当  $D \rightarrow \infty$  时,  $V_{\text{环}} \rightarrow K$ .

同样，可以发现，在  $D$  很大时，其内部几乎不存在任何空间这导致其样本部分部在外球壳上。

降维方法  $\rightarrow$ 

直接降维(即特征选择): 如 L1 (Lasso)	
	线性降维: PCA, MDS (多维缩放)
	非线性降维: 流形

## 二. 样本均值及方差矩阵

$$X = (x_1, x_2, \dots, x_N)^T = \begin{pmatrix} x_{11}, x_{12}, \dots, x_{1p} \\ x_{21}, x_{22}, \dots, x_{2p} \\ \vdots \\ x_{N1}, x_{N2}, \dots, x_{Np} \end{pmatrix}_{N \times p}$$

$x_i \in \mathbb{R}^p, i=1, 2, \dots, N$

$$\text{样本均值: } \bar{X}_{px} = \frac{1}{N} \sum_{i=1}^N x_i$$

$$\text{样本方差: } S_{px} = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{X})(x_i - \bar{X})^T$$

$$\bar{X} = \frac{1}{N} \sum_{i=1}^N x_i = \frac{1}{N} (x_1, x_2, \dots, x_N) \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}_{N \times 1} = X^T \cdot \mathbf{1}_N \quad \mathbf{1}_N = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}_{N \times 1}$$

$$S = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{X})(x_i - \bar{X})^T$$

$$= \frac{1}{N} \underbrace{(x_1 - \bar{X}, x_2 - \bar{X}, \dots, x_N - \bar{X})}_{(x_1, x_2, \dots, x_N) - \bar{X}(1, 1, \dots, 1)} \underbrace{\begin{pmatrix} (x_1 - \bar{X})^T \\ (x_2 - \bar{X})^T \\ \vdots \\ (x_N - \bar{X})^T \end{pmatrix}}_{\mathbf{H}}$$

$$\frac{1}{N} \cdot \underbrace{x^T (I_N - \frac{1}{N} \mathbf{1}_N \mathbf{1}_N^T)}_{\text{令 } H_N \text{ 为 centering matrix (中心矩阵)}} \cdot (I_N - \frac{1}{N} \mathbf{1}_N \mathbf{1}_N^T)^T x = \frac{1}{N} x^T H \cdot H^T x$$

H 的性质:  $H = H^T, H^T = H$

$$\therefore S = \frac{1}{N} x^T H x$$

PCA:

一个中心：原始特征空间的重心。  
 两个基本点：  
 ① 最大投影方差  
 ② 最小重构距离

最大投影方差：

目的：找到  $u$ ，使得中心化后的样本在  $u$  上的投影的方差之和最大。

$$\text{归一化 } J = \frac{1}{N} \sum_{i=1}^N ((x_i - \bar{x})^T u)^2$$

$$= \frac{1}{N} \sum_{i=1}^N (u^T (x_i - \bar{x}) \cdot (x_i - \bar{x})^T u)$$

$$= u^T \underbrace{\left( \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(x_i - \bar{x})^T \right)}_{\text{协方差矩阵 } S} u$$

可以乘上一个  $\frac{1}{N}$ ，对结果无影响，即  $J = \frac{1}{N} \sum_{i=1}^N ((x_i - \bar{x})^T u)^2$

$$= u^T \underbrace{\left( \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(x_i - \bar{x})^T \right)}_{\text{协方差矩阵 } S} u$$

$$= u^T S u$$

即求  $\hat{u} = \arg \max_u u^T S u$   
 s.t.  $u^T u = 1$

拉格朗日乘子法： $L(u, \lambda) = u^T S u + \lambda(1 - u^T u)$

$$\text{令 } \frac{\partial L}{\partial u} = 2S u - 2\lambda u = 0,$$

$$\text{得 } S u = \lambda u$$

$\downarrow$   
特征向量      特征值

故  $u$  是协方差矩阵的特征向量。

令  $\|u\| = \sqrt{u^T u} = 1$   
 就相当于找主成分，要找样本  $x_i$  在  $u$  上的投影  
 想降到  $m$  维，那就找前  $m$  个  $u$  就行了。  
 $u$  是由  $x_i \cdot u$ ,  $P P x_i^T u$   
 均值  $\bar{x}$  的投影为  $\bar{x} \cdot u$   
 $P P \bar{x}^T u$ ，因此  $x_i$  在  $u$  上投影的方差为  
 $((x_i - \bar{x})^T u)^2$ .

最后解得符合条件的向量  $u$  是协方差矩阵  $S$  的特征向量，共  $p$  个，若想降到  $q$  维 ( $q < p$ )，可以选择对应特征值最大的前  $q$  个特征向量取出来作为投影方向，然后获得数据在这些方向上的投影即为重构坐标，即：

$$\begin{pmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_N^T \end{pmatrix}_{N \times p} (u_1 \ u_2 \ \dots \ u_q)_{p \times q} = \begin{pmatrix} x_1^T u_1 & x_1^T u_2 & \dots & x_1^T u_q \\ x_2^T u_1 & x_2^T u_2 & \dots & x_2^T u_q \\ \vdots & \vdots & \ddots & \vdots \\ x_N^T u_1 & x_N^T u_2 & \dots & x_N^T u_q \end{pmatrix}_{N \times q}$$

特征向量表示投影变换的方向，特征值表示投影变换的强度。

最小重构距离。

由于共  $p$  个特征向量符合条件，则原来的数据可表示为：

$$x_i - \bar{x} = \sum_{k=1}^p ((x_i - \bar{x})^T u_k) u_k$$

降维后的数据：

$$\hat{x}_i = \sum_{k=1}^q ((x_i - \bar{x})^T u_k) u_k$$

因此重构距离为：

$$(x_i - \bar{x}) - \hat{x}_i$$

故应最小化  $J$ ，其中：

$$\begin{aligned} J &= \frac{1}{N} \sum_{i=1}^N \| (x_i - \bar{x}) - \hat{x}_i \|^2 \\ &= \frac{1}{N} \sum_{i=1}^N \| \sum_{k=1}^p ((x_i - \bar{x})^T u_k) u_k \|^2 \\ &= \frac{1}{N} \sum_{i=1}^N \sum_{k=1}^p ((x_i - \bar{x})^T u_k)^2 \\ &= \frac{1}{N} \sum_{i=1}^N \sum_{k=1}^p ((x_i - \bar{x})^T u_k)^2 \\ &= \sum_{k=1}^p \underbrace{\frac{1}{N} \sum_{i=1}^N ((x_i - \bar{x})^T u_k)^2}_{u_k^T S \cdot u_k} \\ &= \boxed{\sum_{k=1}^p u_k^T S \cdot u_k}, \quad \text{s.t. } u_k^T u_k = 1 \end{aligned}$$

即求解下列最优化问题：

$$\left\{ \begin{array}{l} \hat{u} = \arg \min_u \sum_{k=1}^p u_k^T S \cdot u_k \\ \text{s.t. } u_k^T u_k = 1 \end{array} \right.$$

tips:

最大投影方差求得的  $p$  个特征向量，  
我们要的是特征值最大的前  $q$  个作为  
新的坐标方向向量，所以剩下的  
 $p-q$  个特征向量是对应的最小的几  
个特征值，所以最小重构距离所求  
得的  $u_k$  都是对应的最小的几个特征  
值。

概率角度的PCA (P-PCA)