

专题 3: PCA&SVD

教授: 陈宝权

作者: 樊泽嘉、翁伊嘉、龚晨、史梦芝、秘玮晨、邓茵琳

免责声明: 该笔记尚未进行正式出版物的通常审查, 未经教师允许不能向课程外分发.

3.1 主成分分析 (PCA)

3.1.1 介绍

PCA(Principal Component Analysis) 是一种有用的线性分解, 主要用于对给定数据找最优坐标系、数据降维。主成分分析可以找到一组能最好表达给定数据的正交基, 例如, 在平面直角坐标系中找到新的坐标, 使点到 x 轴的距离的平方和最小 Fig.3.7。

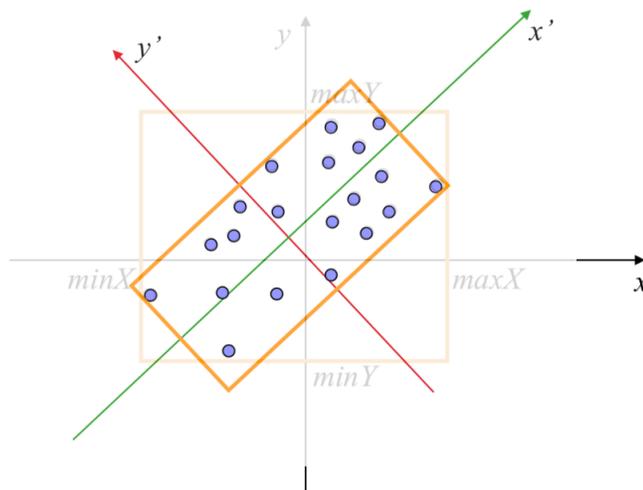


图 3.1: 找到了更好的坐标系

3.1.2 概念

将每一个点定义为向量 \mathbf{x} , $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n \in \mathbf{R}^d, \mathbf{x}_i = (x_i^1, x_i^2, \dots, x_i^d)^T$, 则这些点的质心是

$$\mathbf{m} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i$$

, 由求导可知

$$\mathbf{m} = \arg \min_x \sum_{i=1}^n \|\mathbf{x}_i - \mathbf{x}\|^2$$

记 $\mathbf{y}_i = \mathbf{x}_i - \mathbf{m}, i = 1, 2, \dots, n$

$$\mathbf{S} = \mathbf{Y}\mathbf{Y}^T$$

$\mathbf{y}_k \mathbf{Y}^T$ ($k = 1, 2, \dots, n$), \mathbf{Y} 为 \mathbf{x}_i 的协方差矩阵。S 衡量了数据在不同方向上的分散程度。

对任一过质心的直线, 定义点集在该直线上的投影的方差是

$$\text{var}(L) = \frac{1}{n} \sum_{i=1}^n \|\mathbf{x}'_i - \mathbf{m}\|^2$$

给定单位向量 \mathbf{v} , 直线 $L = \mathbf{m} + \mathbf{v}t$, x_i 在直线 L 上的投影为 x'_i ,

$$\|\mathbf{x}'_i - \mathbf{m}\| = \left| \frac{\langle \mathbf{v}, \mathbf{x}_i - \mathbf{m} \rangle}{\|\mathbf{v}\|} \right| = |\langle \mathbf{v}, \mathbf{y}_i \rangle| = |\mathbf{v}^T \mathbf{y}_i|$$

计算得,

$$\text{var}(l) \frac{1}{n} \sum_{i=1}^n \|\mathbf{x}'_i - \mathbf{m}\|^2 = \frac{1}{n} \sum_{i=1}^n (\mathbf{v}^T \mathbf{y}_i)^2 = \frac{1}{n} \|\mathbf{v}^T \mathbf{Y}\|^2 = \frac{1}{n} (\mathbf{v}^T \mathbf{Y})(\mathbf{v}^T \mathbf{Y})^T = \frac{1}{n} \mathbf{v}^T \mathbf{Y} \mathbf{Y}^T \mathbf{v} = \frac{1}{n} \langle \mathbf{S} \mathbf{v}, \mathbf{v} \rangle$$

定理 3.1 函数 $f: \{v \in \mathbb{R}^d, \text{s.t. } \|v\| = 1\} \rightarrow \mathbb{R}, f(v) = \langle \mathbf{S} \mathbf{v}, \mathbf{v} \rangle, \mathbf{S}$ 是一个实对称矩阵, 则使 f 取到极值的单位向量一定在 \mathbf{S} 的正交特征向量中取到

$$\mathbf{S} = \mathbf{V} \begin{vmatrix} \lambda_1 & & & & \\ & \lambda_2 & & & \\ & & \lambda_3 & & \\ & & & \dots & \\ & & & & \lambda_n \end{vmatrix} \mathbf{V}^T = \mathbf{V} \mathbf{V}^T$$

$$(\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \dots \geq \lambda_n)$$

$$f(\mathbf{y}) = \langle \mathbf{S} \mathbf{y}, \mathbf{y} \rangle = \mathbf{y}^T \mathbf{V} \mathbf{V}^T \mathbf{y}, \quad \mathbf{y} = \mathbf{V}(y_1, y_2, \dots, y_n)^T$$

$$\text{let } \mathbf{p} = \mathbf{V}^T \mathbf{y}, \text{ then } f(\mathbf{y}) = \mathbf{p}^T \mathbf{p} = \sum_{i=1}^n y_i^2 \lambda_i$$

as $\|\mathbf{y}\| = 1$, the maximum of $f(\mathbf{y})$ is λ_1 , the minimum of $f(\mathbf{y})$ is λ_n .

类似地, 可以证明该定理在取到最值的特征向量垂直的子空间也是成立的, 于是通过 PCA 找到的正交基所代表的坐标系上的 bounding box 是一个足够紧的 bounding box Fig.3.2。

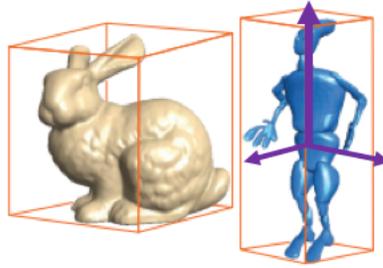


图 3.2: 以长方体代表三维物体

3.2 奇异值分解

3.2.1 引入: 线性变换的几何意义

对于用矩阵 A 定义的 \mathbb{R}^d 中线性映射 $\mathcal{A}: v \rightarrow Av$, 我们希望找到一种表示 A 的方式, 能够给我们这个线性变换的几何直观。

- A 为实对称矩阵时, A 可被正交对角化, 即

$$\exists V, D, VV^T = I, D = \text{diag}\{\lambda_1, \dots, \lambda_d\}$$

$$\text{s.t. } AV = VD \Rightarrow A = VDV^T$$

这一分解形式称为 A 的谱分解 (Spectral Decomposition). 由它我们得到

$$Av_i = \lambda_i v_i$$

其几何直观为, 对于 V 指定的标准正交基的任意一个坐标轴 v_i , v_i 在 \mathcal{A} 作用下被放缩到原本的 λ_i 倍。此时的 \mathcal{A} 是一个放缩变换。

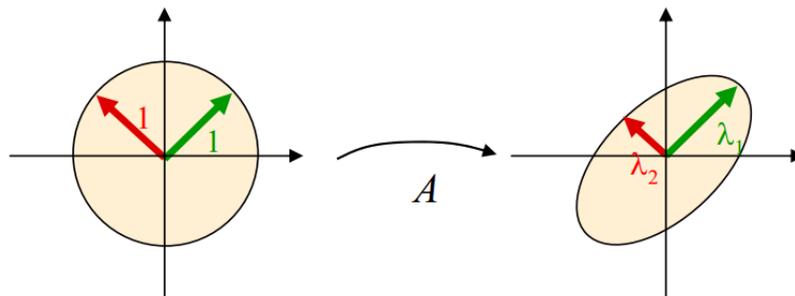


图 3.3: 对称矩阵 A 对应的线性变换是一个放缩变换

- 对于一般的 A , 我们可以将谱分解扩展为奇异值分解, 我们将证明, A 总能分解为

$$A = U\Sigma V^T$$

其中 U, V 为正交矩阵, $\Sigma = \text{diag}\{\sigma_1, \dots, \sigma_d\}, \sigma_i \geq 0$.

由这一分解我们得到

$$A = U\Sigma V^T \Rightarrow AV = U\Sigma \Rightarrow Av_i = \sigma_i u_i$$

其几何直观为, 对于 V 指定的标准正交基的任意一个坐标轴 v_i , v_i 在 A 作用下被放缩到原本的 σ_i 倍, 并被旋转 (可能带反射) 到另一标准正交基的轴 u_i 上。

A 的奇异值分解提供了线性变换的几何直观!

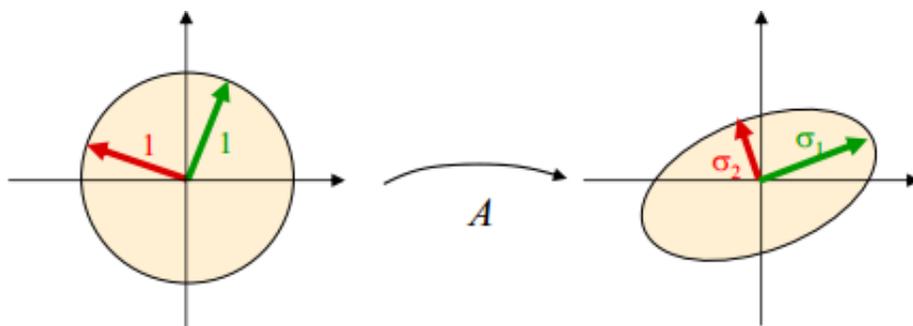


图 3.4: 一般矩阵 A 对应的线性变换是一个放缩 + 旋转变换

3.2.2 奇异值分解的定义

3.2.2.1 一般形式 (Full SVD)

对一个任意的矩阵 $A \in \mathbb{R}^{d \times n}$, \exists 正交矩阵 $U \in \mathbb{R}^{d \times d}$, $V \in \mathbb{R}^{n \times n}$, 对角矩阵 $\Sigma \in \mathbb{R}^{d \times n}$ 其对角线元素依次为 $\sigma_1, \sigma_2, \dots, \sigma_{\min\{d,n\}}$, 且有 $\sigma_i \geq 0, \forall i$,

$$s.t. A = U\Sigma V^T$$

这一分解称为 A 的**奇异值分解 (Singular Value Decomposition)**.

$\Sigma(\sigma_1, \dots, \sigma_{\min\{d,n\}})$ 的对角元素 σ_i 称为**奇异值 (Singular Values)**. 一般地, 我们规定 $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_{\min\{d,n\}}$

$U(u_1, \dots, u_d)$ 的列向量 u_i 称为**左奇异向量 (Left Singular Values)**

$V(v_1, \dots, v_d)$ 的列向量 v_i 称为**右奇异向量 (Right Singular Values)**

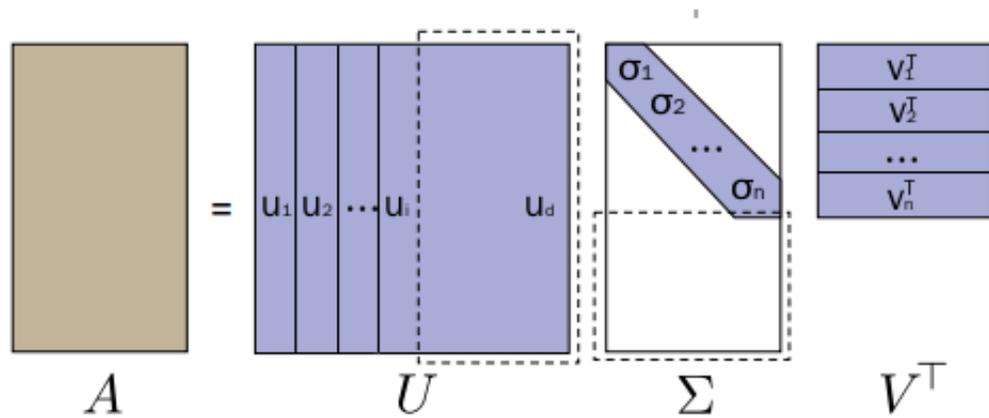


图 3.5: 矩阵 A 的 SVD 分解

3.2.2.2 简化形式 (Reduced SVD)

SVD 的一般形式在存储与运算上代价较高，下面介绍两种开销较小的 SVD 形式。

将 $A = U\Sigma V^T$ 按照矩阵乘法规律展开

$$A = [u_1, u_2, \dots, u_d] \begin{bmatrix} \sigma_1 & & & & \\ & \sigma_2 & & & \\ & & \dots & & \\ & & & \sigma_{\min\{n,d\}} & \\ 0 & 0 & \dots & 0 & \\ \vdots & \vdots & & \vdots & \\ 0 & 0 & \dots & 0 & \end{bmatrix} [v_1, v_2, \dots, v_n]^T$$

我们可以得到 A 的另一个形式，

$$A = \sum_{i=1}^{\min\{n,d\}} \sigma_i u_i v_i^T$$

这一形式启发我们， U, V 中下标 $i \leq \min\{n, d\}$ 的列向量由于均不出现在 A 的表达式中，可以在 SVD 分解中省去。由此，我们得到了 SVD 的简化形式 (Reduced Form)，不妨设 $A \in \mathbb{R}^{d \times n}, d > n$ ，我们有

$$A = U\Sigma V^T, U \in \mathbb{R}^{d \times n}, \Sigma \in \mathbb{R}^{n \times n}, V \in \mathbb{R}^{n \times n}$$

这里 U, V 为列向量单位正交的矩阵。

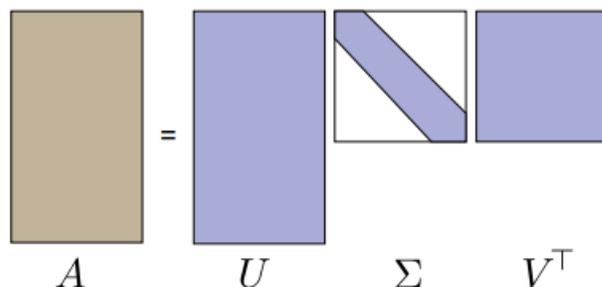


图 3.6: Reduced SVD

3.2.2.3 截断形式 (Truncated SVD)

在 $A = \sum_{i=1}^{\min\{n,d\}} \sigma_i u_i v_i^T$ 中, 我们容易发现与 $\sigma_i = 0$ 对应的 u_i, v_i 也可以直接丢弃。由于 σ_i 已经降序排列, 我们只需舍弃 U, V, Σ 的若干末尾行列。

进一步地, 我们还可以设置阈值 ϵ , 将 $\sigma_i \leq \epsilon$ 的部分丢弃, 此时有

$$\hat{A} = \sum_{i=1}^t \sigma_i u_i v_i^T$$

我们将这种形式称为 SVD 的截断形式 (Truncated Form)。此时的 \hat{A} 仅是 A 的一个近似, 我们将看到, SVD 的这种形式在数据降维、降噪方面有着广泛的应用。

3.2.3 线性代数的巨大助力

当我们有了计算奇异值分解的算法后, 我们可以做:

- 矩阵求逆: 解线性系统
- 定义矩阵的“数值逆”
- 主成分分析计算
- 解最小二乘问题

矩阵求逆与解线性问题

矩阵的逆:

$$A = U \Sigma V^T$$

$$A^{-1} = (U\Sigma V^T)^{-1} = (V^T)^{-1}\Sigma^{-1}U^{-1}$$

$$= V \begin{bmatrix} \frac{1}{\sigma_1} & & & \\ & \ddots & & \\ & & \ddots & \\ & & & \frac{1}{\sigma_n} \end{bmatrix} U^T$$

所以，当解方程 $Ax = b$ 时，我们可以用 $x = V\Sigma^{-1}U^Tb$

矩阵的秩

矩阵 A 的秩为非零的奇异值数。如果有一些非常小的奇异值，那么矩阵 A 很接近奇异矩阵。

我们可以设定一个临界值 $t: \text{numeric_rank}(A) = \#\{\sigma_i | \sigma_i > t\}$

另外，当 $\text{rank}(A) < n$ 时， A 是奇异矩阵，将整个空间 R^n 映射到某个子空间，比如说一个平面。（所以 A 可以看成某种投影）

主成分分析

现在我们有矩阵 Y 的奇异值分解： $Y = U\Sigma V^T$ 。那么“分散矩阵” $S = YY^T$ 可以简化成：

$$YY^T = U\Sigma V^T(U\Sigma V^T)^T$$

$$= U\Sigma V^T V \Sigma^T U^T$$

$$= U(\Sigma\Sigma^T)U^T$$

3.2.4 奇异值分解的另一角度

将 $d \times n$ 矩阵 A 的每一列看成 d 维空间的一个点，通过奇异值分解，我们可以找到对这 n 个点的最适合的 k 维子空间。

这里的“最适合”指的是：在取相同维数的子空间时

- 最小化点到子空间的距离的平方和：类似最小二乘拟合的概念
- 最大化点在子空间上的投影的平方和：这表明子空间包含了数据点的最多内容

矩阵 A 的第 i 列 a_i ，在 u 上的投影的长度为 $|u^T \cdot a_i|$ 。整个矩阵的投影长度的平方和为 $|u^T A|^2$ 。

定义 A 的第一个左奇异向量 u_1 为 $u_1 = \arg \max_{|u|=1} |u^T A|$ 。

第二左奇异向量 u_2 定义为垂直于 u_1 的最好的拟合方向，即 $u_2 = \arg \max_{u \perp u_1, |u|=1} |u^T A|$

$$\sigma_i(A) = |u_i^T A|$$

$$v_i = \frac{1}{\sigma_i(A)} u_i^T A$$

$$A = \sum_{i=1}^r \sigma_i u_i v_i^T$$

我们可以用上面的等式取前 k 项，对矩阵 A 做 k 阶估计。

贪心算法的正确性 记 A 为 $d \times n$ 的矩阵, 左奇异向量为 u_1, u_2, \dots, u_r . 对 $1 \leq k \leq r$, 记 U_k 由 u_1, u_2, \dots, u_r 扩展的子空间. 对每个 k , U_k 是对矩阵 A 拟合得最好的 k 维空间。

证明当 $k=1$ 显然成立. 对 $k=2$, 记 W 为对 A 拟合得最好的 2 维子空间. 对 W 的任一组标准正交基 (w_1, w_2) , $|w_1^T A|^2 + |w_2^T A|^2$ 为 A 的列向量在 W 上投影的平方和. 取一组 (w_1, w_2) 使得 w_2 垂直于 w_1 . 因为 w_1 最大化 $|u^T A|^2$, 我们有 $|w_1^T A|^2 \leq |u_1^T A|^2$. 因为 w_2 在所有与 w_1 垂直的向量中最大化 $|u^T A|^2$, $|w_2^T A|^2 \leq |v_2^T A|^2$. 所以 $|w_1^T A|^2 + |w_2^T A|^2 \leq |u_1^T A|^2 + |u_2^T A|^2$. 因此 U_2 至少和 W 一样好, 即是最好的二维子空间。

对于一般的 k , 用归纳法证明。

3.2.5 数据的中心化

在主成分分解中, 我们将数据点的中心设为坐标原点。
在找矩阵的低秩近似时, 不需要将每个数据减去中心。

3.2.6 附录

3.2.6.1 奇异值分解的存在性

步骤一 在上面的贪心算法中, 我们构造了矩阵 U, V , 其中 U 根据构造方法是一个正交矩阵. 当我们找到左奇异向量 u_1, u_2, \dots, u_r , 奇异值 $\sigma_1, \sigma_2, \dots, \sigma_r$, 且 $\max_{u \perp u_1, u_2, \dots, u_r, |u|=1} |u^T A| = 0$ 时, 算法停止。

步骤二 现在我们需要证明 V 矩阵的正交性。

直观上看, 如果两个向量 $v_i, v_j, i \neq j$ 不相互垂直, 我们会怀疑左奇异向量 u_j 中有向量 u_i 的部分, 这与 u_i 和 u_j 的正交性矛盾. 下面证明的思路就是在假设存在这样的向量 $v_i, v_j, i \neq j$ 时, 构造 u'_i , 使得 u'_i 垂直于 u_1, \dots, u_{i-1} , 但 $|u'_i{}^T A| > \sigma_i$.

证明假设 i 是使得 v_i 不垂直于其他某个向量的最小整数 i . 不失一般性, 我们可以假设 $v_i^T v_j = \delta > 0$, 否则就用 $-u_j$ 代替 u_j . 显然 $j > i$, 因为 i 是满足上述条件的最小整数。

取 $\epsilon > 0$, 令 $u'_i = \frac{u_i + \epsilon u_j}{|u_i + \epsilon u_j|}$. 注意 u'_i 是一个单位向量. 且 $u'_i{}^T A = \frac{\sigma_i v_i + \epsilon \sigma_j v_j}{\sqrt{1 + \epsilon^2}}$ 的模长不小于它在向量 v_i 上投影的长度. 投影长度为:

$$\begin{aligned} v_i^T \left(\frac{\sigma_i v_i + \epsilon \sigma_j v_j}{\sqrt{1 + \epsilon^2}} \right) &> (\sigma_i + \epsilon \sigma_j \delta) \left(1 - \frac{\epsilon^2}{2} \right) \\ &> \sigma_i - \frac{\epsilon^2}{2} \sigma_i + \epsilon \sigma_j \delta - \frac{\epsilon^2}{2} \sigma_j \delta > \sigma_i, \end{aligned}$$

对充分小的 ϵ .

产生矛盾, 说明不存在这样的 i 。

3.2.6.2 一些关于“数据中心化”的证明

定理 3.2 拟合数据点集最好的直线（点到直线的距离的平方之和最小）一定经过点集的中心。

证明 从数据点中减去点集的中心，此时中心为坐标原点。设此时拟合最好的直线为 l 。用反证法，假设 l 不经过坐标原点，则直线 l 可以写成 $\{a + \lambda v | \lambda \in R\}$ ，其中 a 是直线上离坐标原点最近的点， v 是 l 的单位方向向量， v 与 a 垂直。对数据点 a_i ，记点到直线 l 的距离为 $dist(a_i, l)$ 。由勾股定理有 $dist(a_i, l)^2 = |a_i - a|^2 - (v \cdot a_i)^2$ 。

对 n 个数据点进行求和有：

$$\begin{aligned} \sum_{i=1}^n dist(a_i, l) &= \sum (|a_i|^2 + |a|^2 - 2a_i \cdot a - (v \cdot a_i)^2) \\ &= \sum_{i=1}^n |a_i|^2 + n|a|^2 - 2a \cdot (\sum_i a_i) - \sum_{i=1}^n (v \cdot a_i)^2 = \sum_i |a_i|^2 + n|a|^2 - \sum_i (v \cdot a_i)^2 \end{aligned}$$

$a = 0$ 时，上述表达式取最小值。

定义仿射空间为一个子空间经平移产生： $\{v_0 + \sum_{i=1}^k c_i v_i | c_1, c_2, \dots, c_k \in R\}$ 。

其中 v_0 代表平移操作 v_1, \dots, v_k 是原来子空间的一组正交基。

定理 3.3 在所有的 k 维仿射空间中，最小化点到空间的距离的平方和的空间一定经过数据点的中心。

证明 和前一个定理相似，现在我们需要计算 $\sum_{j=1}^k (v_j \cdot a_i)$ ，其中 $v_j, j = 1, 2, \dots, k$ 是一个过原点的平行于仿射空间的子空间的一组正交基。

3.3 直接应用 PCA 和 SVD 的例子

3.3.1 求解点云的近似法向量

对三维空间中的点云做 PCA，那么前两个特征值对应的特征向量构成了这些点的一个近似平面，最小的特征值对应的特征向量就是近似的法向量。

3.3.2 数据降维

3.3.2.1 图片压缩

将图片看作二维矩阵做 SVD 分解， $A = U \Sigma V^T = \sum_{i=1}^r \sigma_i u_i v_i^T$ ， $A_k = \sum_{i=1}^k \sigma_i u_i v_i^T$ 是 A 的最佳 k 阶近似。因此可以选择合适的 k ，例如选择 $\frac{\sum_{i=1}^k \sigma_i}{\sum_{i=1}^r \sigma_i} \geq 90\%$

3.3.2.2 特征脸

通过把一张图片投影到一个低维子空间中来得到它的低维表示

1. 获取一个训练集 T , 包含 n 张人脸的图片。把每张图片转成一个 d 维向量, 把 n 张图片聚合成一个 $n * d$ 维的矩阵 A ;
2. 对 A 进行 PCA 分解, 得到若干个特征向量, 取最大的 k 个特征值对应的特征向量张成了一个 k 维的特征空间;
3. 把人脸图片投影到这个特征空间中, 可以得到它的低维表示。

3.4 Orthonormal Procrustes Problem

3.4.1 问题描述

$P_1 \in R^{2*N}$ 和 $P_2 \in R^{2*N}$ 是在二维平面上的两个点集。求解一个旋转矩阵 R (即 R 是正交矩阵且 $\det(R) = 1$) 使 E 最小化。即得到下面的式子: $P_1 = RP_2 + E$ 。

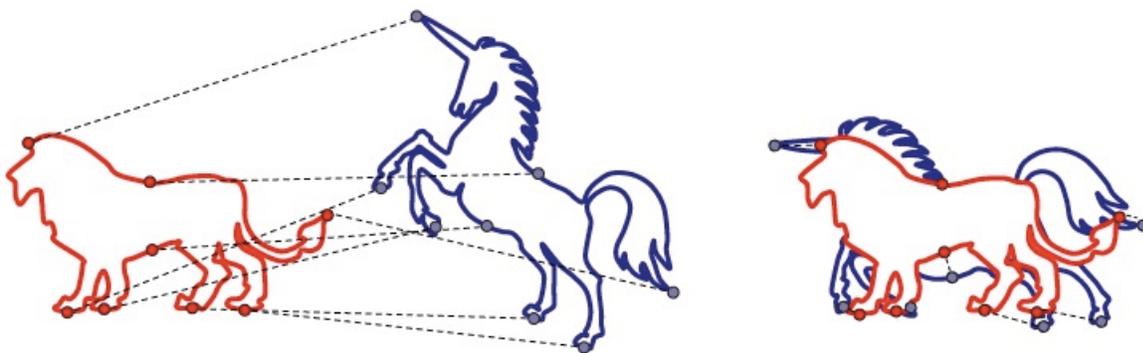


图 3.7: 刚性变换

3.4.2 数学推导

$$\begin{aligned}
 E(R) &= \|P_1 - RP_2\|_F^2 \\
 &= \text{trace}((P_1 - RP_2)^T(P_1 - RP_2)) \\
 &= \text{trace}(P_1^T P_1 + P_2^T R^T R P_2 - P_2^T R^T P_1 - P_1^T R P_2) \\
 &= \text{trace}(P_1^T P_1 + P_2^T P_2 - P_2^T R^T P_1 - P_1^T R P_2) // R^T R = I \\
 &= \text{trace}(P_1^T P_1 + P_2^T P_2) - 2\text{trace}(P_1^T R P_2)
 \end{aligned}$$

第一项是常数，第二项是我们要最大化的式子：

$$\begin{aligned} \text{trace}(P_1^T R P_2) &= \text{trace}(R P_2 P_1^T) \\ &= \text{trace}(R U' S' V'^T) \text{ (使用了 SVD 分解 } P_2 P_1^T = U' S' V'^T) \\ &= \text{trace}(S' V'^T R U') = \text{trace}(S' X) \text{ (X 是一个正交的矩阵)} \\ &= \sum_i S'_{ii} X_{ii} \end{aligned}$$

由于 S'_{ii} 都是奇异值 (S_{11}, S_{22})，均为非负的值，故想要最大化则必须 $X = I$ ，则我们有 $V'^T R U' = I$ ，给出了 $R = V' U'^T$ 。

这样我们得到了一个正交矩阵，但不一定是旋转矩阵，我们通过 X 矩阵的修改，得到 $X = \text{diag}(1, \det(V' U'^T))$ ，再求解 R 即可。我们可以通过一些数学的分析得出 $E(R)$ 的极值点在 $(X_{11}, X_{22}) : (-1, 1), (1, 1), (1, -1), (-1, -1)$ 点上取到，其中关于旋转矩阵的极值点在 $X = \text{diag}(1, \det(V' U'^T))$ 上取到，证明详见 [4]

3.5 As-Rigid-As-Possible Surface Modeling

3.5.1 问题描述

物体表面编辑重建，包括旋转、平移，不包括伸缩。考虑尽可能的保持局部信息，使物体表面的细节得以保留。

3.5.2 数学描述

用 p_i 表示原图形中的一点， p'_i 表示编辑后图形中 p_i 对应的点， $N(i)$ 表示点 p_i 的邻域。

定义 $E(C_i, C'_i) = \sum_{j \in N(i)} w_{ij} \|(p'_i - p'_j) - R_i(p_i - p_j)\|^2$ 其中 R_i 为正交矩阵。 $E(C_i, C'_i)$ 表示了向量 $\overrightarrow{p_j p'_i}$ 经过变换 R_i 后的向量与 $\overrightarrow{p'_j p'_i}$ 之间的差值之和，其中 p_j 为 p_i 邻域中的点， w_{ij} 为求和权重。

取 $C_1, C_2 \dots C_n$ 构成图形表面 S 的一个有限子覆盖，定义 $E(S) = \sum_{i=1}^n w_i E(C_i, C'_i)$ ，为了保持图形局部的稳定性，我们想要让 $E(S)$ 尽可能的小。

3.5.3 数学推导

$E(S) = \sum_{i=1}^n w_i \sum_{j \in N(i)} w_{ij} \|(p'_i - p'_j) - R_i(p_i - p_j)\|^2$ ，采用迭代算法逐步逼近最优解：固定 \mathbf{p}' 找到最优的 $R_i (i = 1, 2, \dots, n)$ ；固定最优的 R_i 寻找最优的 \mathbf{p}'

下面说明如何利用 SVD，固定 \mathbf{p}' ，找最优的 R_i

$$e_{ij} : p_i - p_j$$

$$\begin{aligned}
E(C_i, C'_i) &= \sum_{j \in N(i)} w_{ij} \|(p'_i - p'_j) - R_i(p_i - p_j)\|^2 \\
&= \sum_j w_{ij} (e'_{ij}{}^T e'_{ij} - 2e'_{ij}{}^T R_i e_{ij} + e_{ij}^T e_{ij}) \\
&\quad \arg \min_{R_i} \sum_j -2w_{ij} e'_{ij}{}^T R_i e_{ij} \\
&= \arg \min_{R_i} \sum_j w_{ij} e'_{ij}{}^T R_i e_{ij} \\
&= \arg \max_{R_i} \text{Tr} \left(\sum_j w_{ij} R_i e_{ij} e'_{ij}{}^T \right) \\
&= \arg \max_{R_i} \text{Tr} \left(R_i \sum_j w_{ij} e_{ij} e'_{ij}{}^T \right)
\end{aligned}$$

令 $S_i = \sum_j w_{ij} e_{ij} e'_{ij}{}^T$, 则原问题转化为求 $\arg \max_{R_i} (R_i S_i)$, 其中 S_i 已知, R_i 为正交矩阵, 可用 SVD 求解

参考文献

- [1] AVRIM BLUM, JOHN HOPCROFT and RAVINDRAN KANNAN, Foundations of Data Science
- [2] JOHN R. HURLEY, RAYMOND B. CATTELL, The Procrustes Program: Producing Direct Rotation to Test a Hypothesized Factor Structure
- [3] SCHÖNEMANN, P.H., A generalized solution of the orthogonal Procrustes problem
- [4] SUWAJANAKORN S, SNAVELY N, TOMPSON J Discovery of Latent 3D Keypoints via End-to-end Geometric Reasoning[J]. 2018.
- [5] SHINJI UMEYAMA Least squares estimation of transformation parameters between two point patterns, IEEE Transactions on Pattern Analysis and Machine Intelligence, 1991.