第二章 线性代数

目录

线性代数的表示

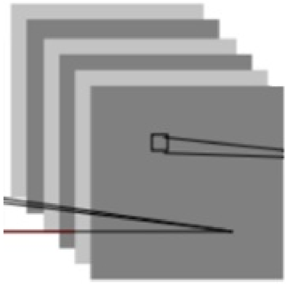
- 标量(Scalars)
  单个数值：整数5、实数0.5、有理数1/3等
  用小写字母表示，如 a, n, x

- 向量(Vectors)
  一维数组，无特别说明，即指列向量
  加粗小写，如

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix} = \left(x_1, x_2, \ldots, x_n\right)^T$$

- 矩阵(Matrix)
  二维数组
  大写字母表示 $A_{m\times n}$ 或 $A \in R^{m\times n}$
  向量可视为 $m \times 1$ 的矩阵

- 张量(Tensor)
  数组的扩展



矩阵的转置

- 定义：行、列坐标互换，行变列，列变行

$(A^T)_{i,j} = A_{j,i}$

如:

$$A_{3\times 2} = \begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \\ A_{3,1} & A_{3,2} \end{bmatrix} \text{则 } A^T_{2\times 3} = \begin{bmatrix} A_{1,1} & A_{2,1} & A_{3,1} \\ A_{1,2} & A_{2,2} & A_{3,2} \end{bmatrix}$$

矩阵的转置有性质:

$$(AB)^T = B^T A^T$$

矩阵相乘投影矩阵

矩阵相乘

- 计算方法1：元素为列乘以行

有矩阵$A_{m\times n}$ $B_{l\times k}$ 若 $n = l$，AB 存在

矩阵A B 相乘，其元素由下式计算

$$(AB)_{i,j} = \sum_k A_{i,k} B_{k,j}$$

如

$$AB = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} = \begin{bmatrix} (1,2,3)(1,2,3)^T & (1,2,3)(4,5,6)^T \\ (4,5,6)(1,2,3)^T & (4,5,6)(4,5,6)^T \end{bmatrix}$$

- 计算方法2：行乘以列的各矩阵相加

$$AB = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} = (1,4)(1,4)^T + (2,5)(2,5)^T + (3,6)(3,6)^T = \begin{bmatrix} 1 & 4 \\ 4 & 16 \end{bmatrix} + \begin{bmatrix} 4 & 10 \\ 10 & 25 \end{bmatrix} + \begin{bmatrix} 9 & 18 \\ 18 & 36 \end{bmatrix}$$

- 怎么应用

行的线性组合可通过左乘行向量取得

$$\begin{bmatrix} c_1 & c_2 \end{bmatrix} \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} = \begin{bmatrix} c_1 a_1 + c_2 a_3 & c_1 a_2 + c_2 a_4 \end{bmatrix}$$

同理，列的线性组合可通过右乘列向量

$$\begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} c_1 a_1 & c_2 a_2 \\ c_1 a_3 & c_2 a_4 \end{bmatrix}$$

- 如何理解

考虑以r为半径，自x轴正方向逆时针旋转角度α所得的向量$(x, y)^T$，其坐可表示为

$$\begin{cases} x = rcos(\alpha) \\ y = rsin(\alpha) \end{cases}$$

对其进行旋转变换，逆时针旋转角度 θ，旋转后坐标为$(x^{'}, y^{'})^T$

则有

$$\begin{cases} x^{'} = rcos(\alpha + \theta) \\ y^{'} = rsin(\alpha + \theta) \end{cases}$$

应用三角公式可得

$$\begin{cases} x^{'} = xcos\theta - ysin\theta \\ y^{'} = xsin + ycos\theta \end{cases}$$

系数可表示为矩阵

$$A = \begin{bmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{bmatrix}$$

考虑将x经过两次旋转，角度分别为ψ和θ，则系数可表示为

$$C = \begin{bmatrix} cos(\psi + \theta) & -sin(\psi + \theta) \\ sin(\psi + \theta) & cos(\psi + \theta) \end{bmatrix} = \begin{bmatrix} cos\theta cos\psi - sin\theta sin\psi & -sin\theta cos\psi - cos\theta sin\psi \\ sin\theta cos\psi + cos\theta sin\psi & cos\theta cos\psi - sin\theta sin\psi \end{bmatrix}$$

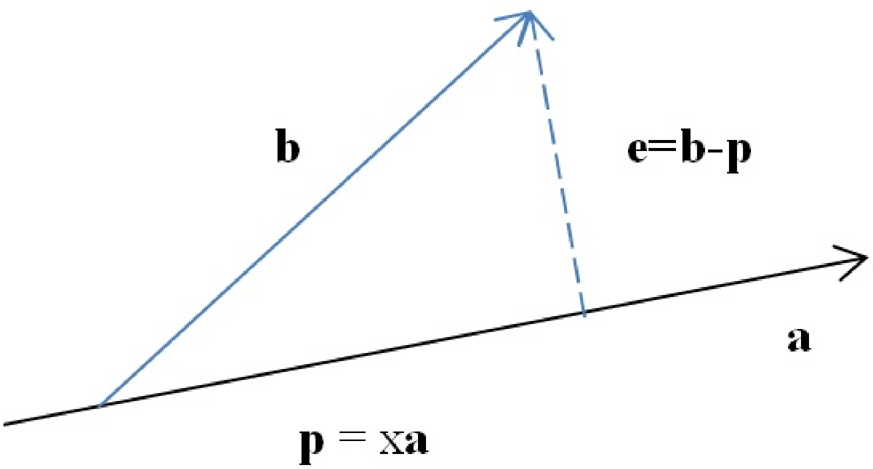
再观察

$$A = \begin{bmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{bmatrix} B = \begin{bmatrix} cos\psi & -sin\psi \\ sin\psi & cos\psi \end{bmatrix}$$

C各元素可由A的对应行乘B的对应列所得，于是由此规定矩阵乘法

投影矩阵

若 $Ax = b$ 无解，则可转而求 $A^TA\hat{x} = A^Tb$



$a \perp e$

从而 $a^T(b-p) = 0$

即 $a^T(b-xa) = 0$

故 $x = \frac{a^Tb}{a^Ta}\ p = \frac{aa^T}{a^Ta}b = Pb$

$P = \frac{aa^T}{a^Ta}_{n\times n}$ 被称为投影矩阵

易证 $A = \begin{bmatrix} a_1 & a_2 \end{bmatrix}$ 时 $P = A(A^TA)^{-1}A^T$

若A 可逆，则$P = I$，即无需投影

对任意投影矩阵，有性质 $P^T = P$ 及 $P^2 = P$，即朝一个方向反复投影，结果不变

向量的线性相关性与矩阵的秩

- 定义
  对向量组$\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_s (s \geq 1)$ 若存在一组不全为0的数

  $$k_1, ..., k_s$$

  使得

  $$k_1\mathbf{x}_1 + ... + k_s\mathbf{x}_s = 0$$

  则称该向量组线性相关，反之，则线性无关
  线性无关的充要条件：其中任何一个向量不能由其余向量线性表出

- 极大线性无关组
  对于一个向量组$\mathbf{x}_1, ..., \mathbf{x}_s$，若存在一个部分组，满足：
  这个部分组线性无关

如果从向量组的其余部分中任取一个添加进去，则到的新部分组都线性相关，则这个部分组称为向量组$\mathbf{x}_1, \ldots, \mathbf{x}_s$的极大线性无关组

向量组与它的任意一个极大线性无关组等价（等价——互可线性表出）
不含零向量的向量组如果线性相关，则它的极大线性无关组肯定不止一个

- 正交向量组
若向量α和β的内积为零，即$\alpha^T\beta = 0$，则称α与β正交，记为$\alpha \perp \beta$
若不含零的向量组α_i内的向量两两相交，则α_i为正交向量组
若一个正交向量组内的任一个向量是单位向量，则说向量组是**标准正交向量组**

- 向量组的秩
向量组的一个极大线性无关组所含的向量的个数，记为 $rank\{\mathbf{x}_1, \ldots, \mathbf{x}_s\}$

- 矩阵的秩
矩阵的列微量组的秩称为 A 的列秩，A 的行向量组的秩称为 A 的行秩
任一矩阵的行秩等于其列秩，统称为矩阵 A 的秩，记作 $rank(A)$

单位矩阵与逆矩阵

- 方阵：行数等于列数 $A_{n\times n}$
- 对角矩阵：对于一个方阵，对角线外的元素均为零，即为对角矩阵
- 单位矩阵：对于一个对角矩阵，对角线上的元素均为1，即为单位矩阵，用I_n表示
- 逆矩阵：对方阵A_n，若存在方阵B_n使得 $AB = BA = I$，则A为可逆矩阵，B与A互为逆矩阵
- 矩阵可逆的等价条件
- A为满秩矩阵(满秩矩阵——矩阵的秩等于方阵的列数/行数)
- A的各列线性无关
- 非满秩方阵称为**奇异矩阵**

相似矩阵

空间的基

线性空间V中的n个线性无关向量$\alpha_1, \ldots, \alpha_n$，且V中任何一个向量都可由其线性表出，

$\alpha = k_1\alpha_1 + \ldots + k_n\alpha_n$

则称其为V的一个基，$(k_1, \ldots, k_n)^T$为α在该基下的坐标

由n个标准正交向量组成的基称为标准正交基，标准正交基不唯一

求矩阵标准正交基的方法——Gram-Schmidt正交化：

i. 对第一个基直接取用$\beta_1 = \alpha_1$
ii. 第2个基减去在第一个基上的投影$\beta_2 = \alpha_2 - \frac{\alpha_2^T\beta_1}{\beta_1^T\beta_1}\beta_1$
iii. 第n个基减去在第1到第n-1个基上的投影，以此类推

以上方法得到的正交基，可以有多个，并不唯一

基与基之间的关系及其求解方法

设 $\alpha_1, ..., \alpha_n$ 与 $\beta_1, ..., \beta_n$ 为 V 中的两组基，它们的关系可表示为

$$\beta_i = (\alpha_1, ..., \alpha_n)\begin{bmatrix} a_{1i} \\ \cdots \\ a_{ni} \end{bmatrix}$$

用矩阵可表示为

$$(\beta_1, ..., \beta_n) = (\alpha_1, ..., \alpha_n)\begin{bmatrix} a_{11} & \ldots & a_{1n} \\ \cdots & & \\ a_{n1} & \ldots & a_{nn} \end{bmatrix}$$

右侧称为过渡矩阵，每一列对应右侧一个基向量过渡到左侧一个基向量的系数

过渡矩阵的求解方法

i. 直观求解，直接取得二者向量的对应关系，如$\beta_i = A\mathbf{a}$，直接将$\mathbf{a}$写作过渡矩阵的各列即可

ii. 若A, B均为方阵，则可对$[A|B]$辅助矩阵进行初等变换，令左侧为单位矩阵，则右侧即为过渡矩阵，方法同求矩阵的逆
    证明：$B = AP \Rightarrow P = A^{-1}B$ 对辅助矩阵 $[A|B] \Rightarrow [A^{-1}A|A^{-1}B] \Rightarrow [I|P]$
    注记：可见，凡是涉及单个矩阵求逆的运算，均可用辅助矩阵求解
    待证明，当AB不为方阵时，该方法同样有效，逆矩阵替换为左伪逆

相似矩阵

有矩阵在两组基$\alpha_1, ..., \alpha_n$， $\beta_1, ..., \beta_n$的表示下分别为A与B，两组基的过渡矩阵为P，则有

$$B = P^{-1}AP$$

将以上式推广，则得出相似矩阵的定义:若存在$B = P^{-1}AP$，则称A与B相似

相似矩阵，事实上是将同一座标系相形状相同的矩阵归为一类，尽管从同一个角度看去，二者并不相同，但本质上，二者的形状完全一样，那么，是否存在一个座标系，使得相同形状的矩阵获得相同的表示呢？这就是特征值和特征矩阵，尽管每个矩阵各有其特征向量，这代表了他们映射到目标坐标的不同姿势，但特征值暴露了他们形状相同的本质

特征值分解与奇异值分解

特征值分解

- 对矩阵A，有与之平行的向量v，二者关系表示为

  $Av = \lambda v$

要求得上式的v与λ，或求解方程$|\lambda E - A| = 0$

如：对矩阵

$$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

解方程 $\lambda E - A = (\lambda - 2)^2(\lambda - 6)$

取得特征值 $\lambda_1 = \lambda_2 = 2, \lambda_3 = 6$

将λ_i代入方程$(\lambda I - A) = 0$ 可求得与对角矩阵

$$\Lambda = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

对应的特征矩阵

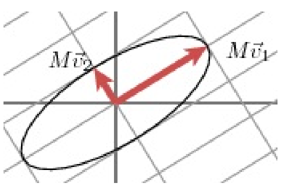
$$V = \begin{bmatrix} -1 & 1 & 1 \\ 1 & 0 & -2 \\ 0 & 1 & 3 \end{bmatrix}$$

当n阶矩阵A有n个线性无关特征向量(P可逆)时，矩阵可对角化为

$$P^{-1}AP = diag(\lambda_1, \ldots, \lambda_n)$$

即

$$A = P\Lambda P^{-1}$$



正定矩阵的特征向量相互正交，实对称矩阵为正定矩阵，对正定矩阵有

$$A = Q\lambda Q^T$$

例如，有一二次曲线，方程为

$$5x^2 + 4xy + 2y^2 - 2 = 0$$

很难从方程上直观得到曲线的形状，将方程列改写为矩阵相乘的形式，且使得系数矩阵为实对称矩阵

$$(x, y)\begin{bmatrix} 5 & 2 \\ 2 & 2 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}$$

对系数矩阵进行特征值分解，可得到其特征值和正交特征向量组，将特征向量组单位化，有

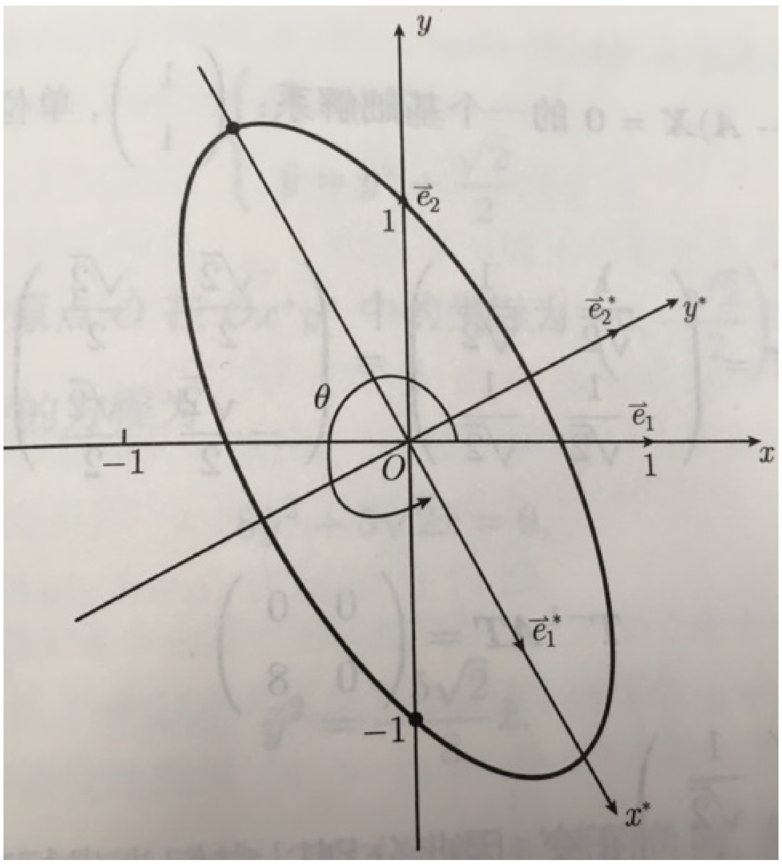
$$\Lambda = \begin{pmatrix} 1 \\ 6 \end{pmatrix} T = \begin{bmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ -2/\sqrt{5} & 1/\sqrt{5} \end{bmatrix} = \begin{bmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{bmatrix}$$

令新坐标为$(x^{'}, y^{'})$，则有

$$(x^{'}, y^{'})T^{'}AT(x^{'}, y^{'})^T = (x^{'}, y^{'})\begin{bmatrix} 1 & 0 \\ 0 & 6 \end{bmatrix}\begin{bmatrix} x^{'} \\ y^{'} \end{bmatrix} = x^{'2} + 6y^{'2}$$

故原方程变为$x^{'2} + y^{'2} - 2 = 0$，即$x^{'2} + y^{'2} = 2$

很明显是一个椭圆



奇异值分解与伪逆

问：特征值分解仅对方阵有效，非方阵如何取得近似？
答：构造实对称矩阵——A^TA与AA^T(二者均为对称矩阵)，对二者进行特征值分解，再开根号

设$A = U\Sigma V^{-1} = U\Sigma V^T$则

$$A^TA = V\Sigma U^TU\Sigma V^T = V\Sigma^2V^T$$

$$AA^T = U\Sigma V^TV\Sigma U^T = U\Sigma^2V^T$$

V为A^TA的特征向量单位化矩阵
U为AA^T的特征向量单位化矩阵

Σ为A^TA的特征值开方对角矩阵

与之思路类似，非方阵与不可逆矩阵可求左右伪逆
A不可逆，但A^TA可逆，对列满秩矩阵，有

$$(A^TA)^{-1}(A^TA) = I$$

即

$$[(A^TA)^{-1}A^T]A = I$$

括号中即为左伪逆，同理可得行满秩矩阵的右伪逆

$$A[A^T(AA^T)^{-1}] = I$$

向量的范数

- 向量的大小用**范数**来衡量，形式上，

$$L^P$$

  范数定义

$$||x||_P = (\sum_i |x_i|^p)^{\frac{1}{p}}$$

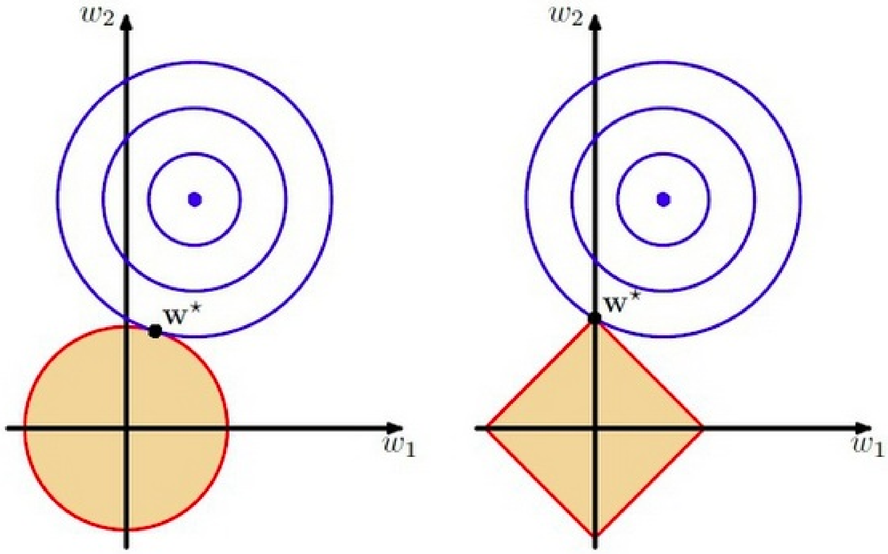
$$||\begin{bmatrix}1\\2\\3\end{bmatrix}||_2 = \sqrt{1^2+2^2+3^2}$$

  范数满足下列性质
- 非负性：若 $\mathbf{x} \neq \mathbf{0}$，则 $f(\mathbf{x}) > 0$；$f(\mathbf{x}) = 0 \Rightarrow \mathbf{x} = \mathbf{0}$
- 三角不等式（不要与凸函数性质混淆）：$f(\mathbf{x}+\mathbf{y}) \leq f(\mathbf{x}) + f(\mathbf{y})$
- 齐次性：$\forall \alpha \in \mathrm{R}, f(\alpha\mathbf{x}) = |\alpha| f(\mathbf{x})$
- 几个常见范数
- L2范数

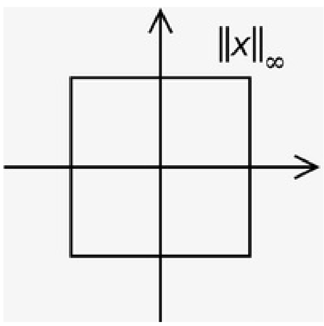
$$p = 2$$

  ，欧里几得范数 $L^2 = x^Tx$，机器学习中最常用
  推广至矩阵Frobenius norm $||A||_F = \sqrt{\sum_{i,j}A_{i,j}^2}$
- L1范数 $p = 1$，$L^1 = ||\mathbf{x}||_1 = \sum_i|x_i|$，也较常用
- 最大范数 $||\mathbf{x}||_\infty = max_i|x_i|$

正则化效果(结合带有不等式约束的极值求解)



w_2
$\mathbf{w}^\star$
w_1
w_2
$\mathbf{w}^\star$
w_1



$\|x\|_\infty$