线性代数基础

机器学习版

VCG

数据的表示：向量与空间

向量

- 在机器学习中，向量 $\mathbf{v} = [\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n]^T$
  - 作为数据点 (Data Point)
    - n 维特征空间中的一个位置
    - 静态的视角
  - 作为空间矢量 (Geometric Vector)
    - 从原点出发，具有方向和长度的箭头
    - 定义了空间中的一个“运动”或“位移”
    - 动态的视角

向量的线性无关

【线性组合】给定$\mathbb{R}^n$中向量$\mathbf{v}_1$，⋯，$\mathbf{v}_p$和标量c_1，⋯，c_p，向量

$$\boldsymbol{y}=c_1\mathbf{v}_1+\cdots+c_p\mathbf{v}_p$$

称为向量$\mathbf{v}_1$，$\mathbf{v}_2$，⋯，$\mathbf{v}_p$以c_1，c_2，⋯，c_p为权的线性组合【数乘和加法】

【**Span**】若$\mathbf{v}_1$，$\mathbf{v}_2$，⋯，$\mathbf{v}_p$是$\mathbb{R}^n$中的向量,$\mathrm{Span}\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_p\}$称为由$\mathbf{v}_1$，$\mathbf{v}_2$，⋯，$\mathbf{v}_p$所生成(张成)的$\mathbb{R}^n$的子集,即，

$$\mathrm{Span}\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_p\}=\{c_1\mathbf{v}_1+\cdots+c_p\mathbf{v}_p \mid c_1, c_2, \cdots, c_p\text{为标量}\}$$

向量空间

【向量空间】

一个向量空间是由向量构成的非空集合V。且定义了运算：加法和标量乘法。

对V中所有向量**u**，**v**，**w**及所有标量c和d，满足

- **u**，**v**之和(表示为$\mathbf{u}+\mathbf{v}$)属于**V**
- $\boldsymbol{u}+\mathbf{v}=\mathbf{v}+\mathbf{u}$
- $(\mathbf{u}+\mathbf{v})+\mathbf{w}=\mathbf{u}+(\mathbf{v}+\mathbf{w})$
- **V**中存在一个零向量0，使得$\mathbf{u}+0=\mathbf{u}$
- 对**V**中每个向量$\boldsymbol{u}$，存在**V**中一个向量-$\boldsymbol{u}$，使得$\boldsymbol{u}+(-\boldsymbol{u})=\mathbf{0}$
- $\boldsymbol{u}$与标量c的标量乘法(记为$c\boldsymbol{u}$)属于**V**
- $c(\boldsymbol{u}+\mathbf{v})=c\boldsymbol{u}+\mathrm{c}\mathbf{v}$
- $(c+d)\mathbf{u}=c\mathbf{u}+d\boldsymbol{u}$
- $c(d\boldsymbol{u})=(cd)\boldsymbol{u}$
- $1\boldsymbol{u}=\boldsymbol{u}$

子空间

【子空间】向量空间**V**的子集**H**，如果**H**满足以下三个性质，则成**H**为**V**的子空间:

V中的零向量在**H**中

H对向量加法封闭，即对**H**中任意向量**u**，**v**，和$\mathbf{u}+\mathbf{v}$仍在**H**中

H对标量乘法封闭，即对**H**中任意向量**u**和任意标量c，向量$c\mathbf{u}$仍在**H**中

子空间的基：线性无关集；标准基

【基】$\mathbb{R}^n$中子空间**H**的一组基是**H**中一个线性无关集，它生成**H**

可逆$n\times n$矩阵的各列构成$\mathbb{R}^n$的一组基，因为它们线性无关，而且生成$\mathbb{R}^n$

【标准基$\mathcal{E}$】$n\times n$单位矩阵，它的各列用$\mathbf{e}_1$，$\cdots$，$\mathbf{e}_n$:

$$\mathbf{e}_1=\begin{bmatrix}1\\0\\\vdots\\0\end{bmatrix},\ \mathbf{e}_2=\begin{bmatrix}0\\1\\\vdots\\0\end{bmatrix},\ \cdots,\ \mathbf{e}_n=\begin{bmatrix}0\\\vdots\\0\\1\end{bmatrix}$$

$\{\mathbf{e}_1，\ \cdots，\ \mathbf{e}_n\}$称为$\mathbb{R}^n$的标准基$\mathcal{E}$

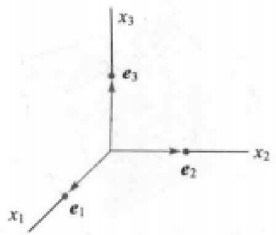


图 2-27 $\mathbb{R}^3$ 的标准基

数据的操作：矩阵与线性变换

矩阵：数据集合 vs. 空间变换

- 一个矩阵 $\mathbf{A}$ 有两种核心视角
  - 数据集合 (Container of Data)
    - 每一行或每一列是一个数据样本或特征
      - 例如：一个用户-物品评分矩阵
    - 静态的视角
  - 线性变换 (Linear Transformation)
    - 一个函数，将一个输入向量 $\mathbf{x}$ 映射到一个输出向量 $\mathbf{y} = \mathbf{A}\mathbf{x}$
      - 描述空间运动：将整个输入空间进行旋转、缩放、切变等操作，变成一个新的输出空间
    - 动态的，也是线性代数最核心的视角

矩阵向量乘法的本质：基向量的“着陆点

【矩阵与向量乘法】$m \times n$矩阵$\mathbf{A} = [\mathbf{a}_1 \quad \mathbf{a}_2 \quad \cdots \quad \mathbf{a}_n]$。$\mathbf{x} \in \mathbb{R}^n$，则$\mathbf{A}$与$\mathbf{x}$的积(记为$Ax$)

$$Ax = [\mathbf{a}_1 \quad \mathbf{a}_2 \quad \cdots \quad \mathbf{a}_n]\begin{bmatrix}\mathbf{x}_1\\ \mathbf{x}_2\\ \vdots\\ \mathbf{x}_n\end{bmatrix} = \mathbf{x}_1\mathbf{a}_1 + \mathbf{x}_2\mathbf{a}_2 + \cdots + \mathbf{x}_n\mathbf{a}_n$$

通常把矩阵$\mathbf{A}$当作一种对象，它通过乘法“作用”于向量$\mathbf{x}$，产生的新向量称为$A\mathbf{x}$。

解方程$\mathbf{Ax} = \mathbf{b}$，就是要求出所有经过乘以$\mathbf{A}$的“作用”后变为$\mathbf{b}$的向量$\mathbf{x}$

线性变换T的标准矩阵

$\mathbf{Ax}$可以看作是$\mathbf{A}$代表的span{$\mathbf{e}_1$， ⋯， $\mathbf{e}_n$}空间中，$\mathbf{x}$作为权重生成的变量

【矩阵变换定理】设$\mathbf{T}: \mathbb{R}^n \to \mathbb{R}^m$为线性变换，则存在唯一的矩阵$\mathbf{A}$，使得对$\mathbb{R}^n$中一切$\mathbf{x}$，

$$\mathrm{T}(\mathbf{x}) = \mathbf{Ax}$$

事实上，$\mathbf{A}$是$\boldsymbol{m} \times \boldsymbol{n}$矩阵，它的第$\boldsymbol{j}$列是向量$\mathbf{T}(\mathbf{e}_j)$，其中$\boldsymbol{e}_j$是$\mathbb{R}^n$中单位矩阵$\boldsymbol{I}_n$的第$j$列:

$$\mathbf{A} = [\mathrm{T}(\mathbf{e}_1) \cdots \mathrm{T}(\mathbf{e}_n)]$$

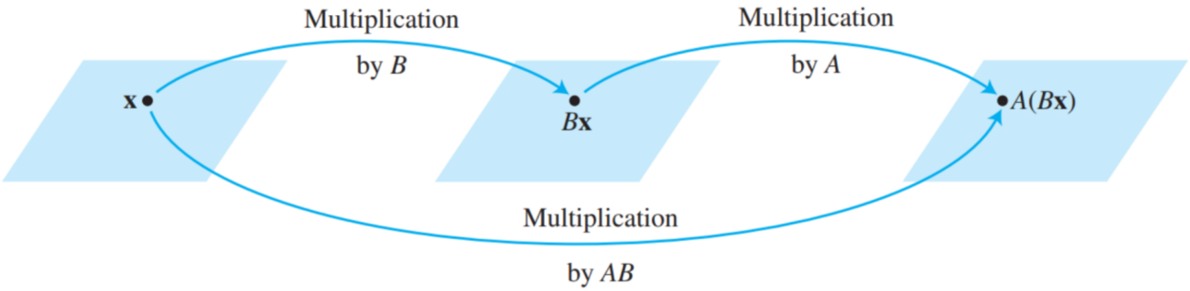
矩阵$\mathbf{A}$称为线性变换T的标准矩阵。

矩阵乘法

【矩阵乘法】矩阵B乘以向量x（将x变换Bx），向量Bx再乘以矩阵A，得A(Bx)。于是，x经复合映射（对应矩阵AB）变换得到A(Bx)：

$$\mathbf{A}(\mathbf{B}\mathbf{x}) = (\mathbf{A}\mathbf{B})\mathbf{x}$$

【矩阵的组合预算】



矩阵的逆，可逆矩阵，奇异矩阵

【可逆矩阵】一个$n \times n$矩阵$\mathbf{A}$是可逆的，若存在一个$n \times n$矩阵$\mathbf{C}$使

$$\mathbf{CA} = \mathbf{I}，\ \mathbf{AC} = \mathbf{I}$$

其中$\mathbf{I} = \mathbf{I}_n$是$n \times n$单位矩阵。这时称$\mathbf{C}$是$\mathbf{A}$的逆

若$\mathbf{A}$可逆，它的逆是唯一的，记为$\mathbf{A}^{-1}$，于是

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}，且\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$$

不可逆矩阵有时称为奇异矩阵，而可逆矩阵也称为非奇异矩阵

坐标系，坐标，线性变换

线性变换

【线性变换】线性变换T将向量空间V中的$\mathbf{x}$映射成向量空间$\mathbf{W}$中唯一向量$\mathrm{T}(\mathbf{x})$，且:

$\mathrm{T}(\mathbf{u}+\mathbf{v})=\mathrm{T}(\mathbf{u})+\mathrm{T}(\mathbf{v})$，对V中所有$\mathbf{u}$，$\mathbf{v}$均成立

$\mathrm{T}(c\mathbf{u})=c\mathrm{T}(\boldsymbol{u})$，对V中所有$\mathbf{u}$及所有数$c$均成立

线性无关集

【线性无关集】V中向量的一个指标集$\{\mathbf{v}_1, \cdots, \mathbf{v}_p\}$称为线性无关，如果向量方程

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p = 0$$

只有平凡解（即$c_1 = 0, \cdots, c_p = 0$）

集合$\{\mathbf{v}_1, \cdots, \mathbf{v}_p\}$称为线性相关，如果方程有一个非平凡的解

基和标准基

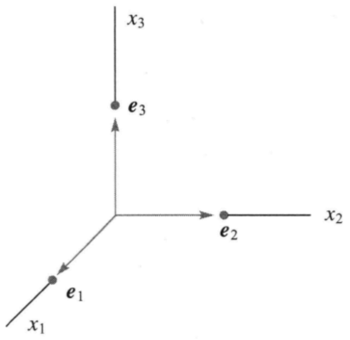
【基】H是向量空间V的子空间。V中一组向量$\mathcal{B}=\{\mathbf{b}_1,\cdots,\mathbf{b}_p\}$称为H的一个基，如果

- $\mathcal{B}$是一线性无关集。
- 由$\mathcal{B}$生成的子空间与H相同，即$\mathbf{H}=\mathrm{Span}\{\mathbf{b}_1,\cdots,\mathbf{b}_p\}$

基是：最小的生成集(再小生成不了H)； 最大线性无关集(再多一个就不是线性无关)

【标准基】单位矩阵 $\mathbf{I}_n=[\mathbf{e}_1\cdots\mathbf{e}_n]$， $\mathcal{E}=\{\mathbf{e}_1, \cdots, \mathbf{e}_n\}$是$\mathbb{R}^n$中的标准基

基$\mathcal{B}$引导了一个新的坐标系，$\{\mathbf{b}_1,\cdots,\mathbf{b}_p\}$为基（坐标轴）为在标准坐标系下的表达



坐标系与坐标

对一个向量空间V，指定一个基$\mathcal{B}$：在V上加上一个“坐标系”，给V以一个新的“视角”。

【坐标】$\mathcal{B}=\{\boldsymbol{b}_1, \ \cdots, \ \boldsymbol{b}_n\}$是V的一个基，记$\mathbf{P}_{\mathcal{B}}=[\mathbf{b}_1 \cdots \mathbf{b}_n]$， $[\mathbf{x}]_{\mathcal{B}}=\begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$。如

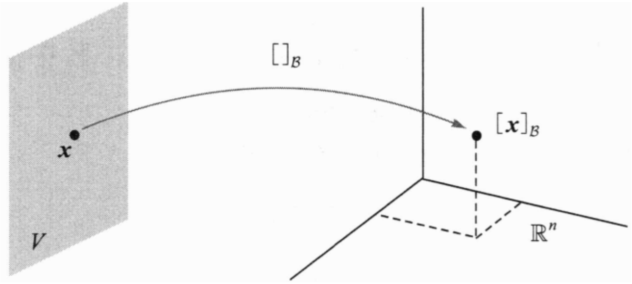
$\mathbf{x}=c_1\mathbf{b}_1+\cdots+c_n\mathbf{b}_n$，即$\mathbf{x}=\mathbf{P}_{\mathcal{B}}[x]_{\mathcal{B}}$

则，称向量$[\mathbf{x}]_{\mathcal{B}}$为$\mathbf{x}$(相对于基$\mathcal{B}$)的坐标向量（或，$\mathbf{x}$的$\mathcal{B}$ –坐标向量）

称$\mathbf{P}_{\mathcal{B}}$为从$\mathcal{B}$到标准基$\mathcal{E}$的坐标变换矩阵$\underset{\mathcal{E}\leftarrow\mathcal{B}}{\mathbf{P}}$。【$\mathcal{B}$中的向量，乘以$\mathbf{P}_{\mathcal{B}}$，变换到$\mathcal{E}$】

【坐标映射】(由$\mathcal{B}$确定的)坐标映射$\mathbf{x}\mapsto[\mathbf{x}]_{\mathcal{B}}$:

$$[\mathbf{x}]_{\mathcal{B}}=\mathbf{P}_{\mathcal{B}}^{-1}\mathbf{x}$$



$\mathbf{b}_1=[\mathbf{b}_1]_{\mathcal{E}}$

【坐标变换矩阵和坐标映射矩阵是逆变换】

基的变换（坐标系变换）

【定理】向量空间V的基$\mathcal{B}=\{\mathbf{b}_1,\ \cdots,\ \mathbf{b}_n\}$和$\mathcal{C}=\{\mathbf{c}_1,\ \cdots,\ \mathbf{c}_n\}$，

称为由$\mathcal{B}$到$\mathcal{C}$的坐标变换矩阵$\underset{\mathcal{C}\leftarrow\mathcal{B}}{\mathbf{P}}$，将$\mathcal{B}$-坐标向量 转成$\mathcal{C}$-坐标向量，即

$$[\mathbf{x}]_{\mathcal{C}}=\underset{\mathcal{C}\leftarrow\mathcal{B}}{\mathbf{P}}[\mathbf{x}]_{\mathcal{B}}$$

那么

$$\underset{\mathcal{C}\leftarrow\mathcal{B}}{\mathbf{P}}=[[\mathbf{b}_1]_{\mathcal{C}}\quad[\mathbf{b}_2]_{\mathcal{C}}\quad\cdots\quad[\mathbf{b}_n]_{\mathcal{C}}]=\mathbf{P}_{\mathcal{C}}{}^{-1}\mathbf{P}_{\mathcal{B}}$$

【计算过程】

$[[\mathbf{x}]_{\mathcal{C}}=\underset{\mathcal{C}\leftarrow\mathcal{B}}{\mathbf{P}}[\mathbf{x}]_{\mathcal{B}}$，两边同时乘以$P_{\mathcal{C}}$

$\mathbf{P}_{\mathcal{C}}[\mathbf{x}]_{\mathcal{C}}=\mathbf{P}_{\mathcal{C}}\underset{\mathcal{C}\leftarrow\mathcal{B}}{\mathbf{P}}[\mathbf{x}]_{\mathcal{B}}$，又$\mathbf{P}_{\mathcal{C}}[\mathbf{x}]_{\mathcal{C}}=\mathbf{P}_{\mathcal{B}}[\mathbf{x}]_{\mathcal{B}}$，所以

$\mathbf{P}_{\mathcal{B}}[[\mathbf{x}]_{\mathcal{B}}=\mathbf{P}_{\mathcal{C}}\underset{\mathcal{C}\leftarrow\mathcal{B}}{\mathbf{P}}[\mathbf{x}]_{\mathcal{B}}$，于是$\underset{\mathcal{C}\leftarrow\mathcal{B}}{\mathbf{P}}=\mathbf{P}_{\mathcal{C}}^{-1}\mathbf{P}_{\mathcal{B}}$

相当于，$\mathcal{B}$-坐标到标准坐标，再映射到$\mathcal{C}$-坐标

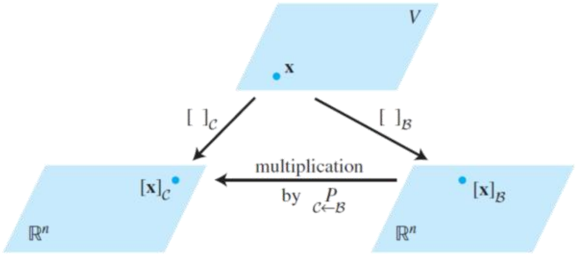


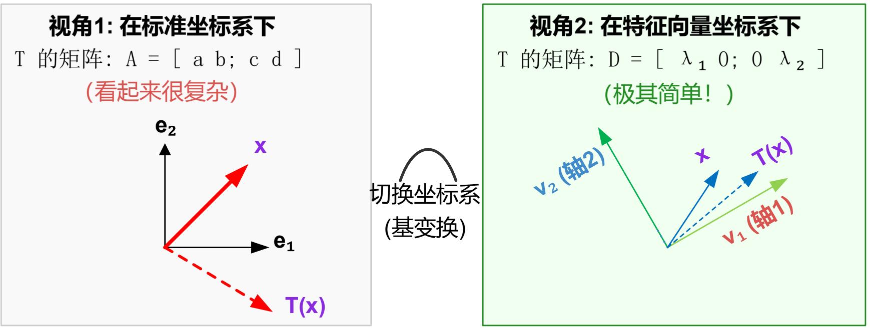
FIGURE 2 Two coordinate systems for V.

特征值与特征向量

矩阵特征向量和特征值：线性变换内在结构的描述

- 特征向量和特征值：描述线性变换核心的，不随观察坐标系改变的性质
  - 变换的主轴
    - 特征向量定义了线性变换的“主运动方向”。 $\mathbf{Ax} = \lambda\mathbf{x}$
  - 空间形变的骨架与自然坐标系 (Natural Coordinate System)
    - 标准形状（单位圆）线性变换后变成椭圆，特征向量（不一定正交）为变换骨架
    - 变换矩阵**A**在特征向量组成的“自然”坐标系（变换**P**）下，变成沿坐标轴缩放的对角矩阵**D**

同一个变换 T，在不同坐标系下的观察



变换 T 表现为复杂的剪切和旋转

变换 T 简化为纯粹的沿轴缩放

$$\mathbf{Ax} = (\mathbf{PDP}^{-1})\mathbf{x}$$

- $\mathbf{P}^{-1}\mathbf{x}$:
  - 把头歪过来，让视线和物体的主轴对齐，变换到自然坐标系
- $\mathbf{DP}^{-1}\mathbf{x}$:
  - 在然坐标系下，沿着坐标轴缩放
- $(\mathbf{PDP}^{-1})\mathbf{x}$
  - 把头摆正，回到原来的观察角度

矩阵特征值与特征向量

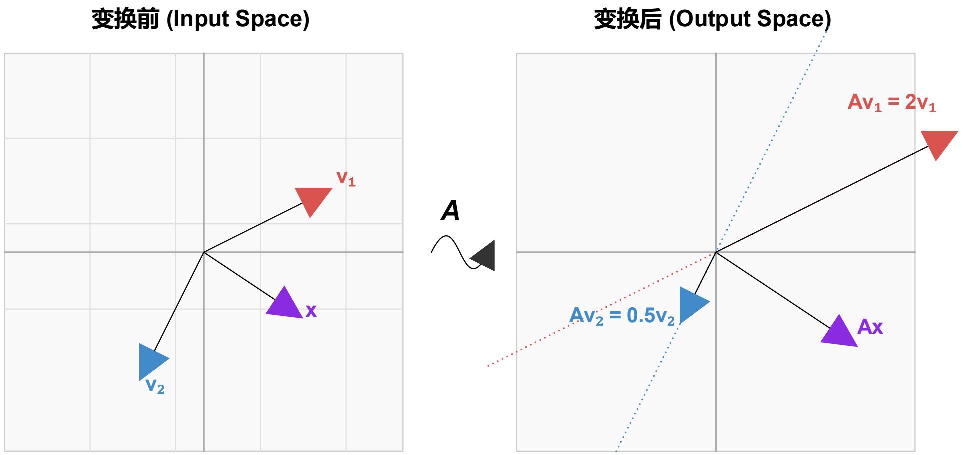
【特征值与特征向量】**A**为$n \times n$矩阵，**x**为非零向量，若存在数λ使

$$\mathbf{Ax} = \lambda \mathbf{x}$$

有非平凡解**x**，则称λ为**A**的特征值，**x**称为对应于λ的特征向量

1.矩阵与向量相乘是一种线性变换

2. 特征向量经过矩阵变换后，方向与原方向保持一致（缩放系数为特征值）



特征向量 v_1 和 v_2 的方向保持不变，仅长度被缩放 (λ_1=2, λ_2=0.5)。
普通向量 x 的方向和长度都发生了改变。

可逆矩阵，正交矩阵

可逆矩阵（Invertible Matrix）

- 方阵$\mathbf{A}_{n\times n}$被称为可逆的（Invertible）或非奇异的（Non-singular）
  - 存在方阵$\mathbf{B}_{n\times n}$，使得$\mathbf{AB}=\mathbf{BA}=\mathbf{I}$。其中$\mathbf{I}_{n\times n}$是单位矩阵。矩阵$\mathbf{B}=\mathbf{A}^{-1}$是A的逆矩阵（Inverse Matrix）

- 【可逆矩阵代表了一种“良好”的、不丢失信息的、可以撤销的线性变换】
- 变换 T是双射
  - 单射：不同的输入向量**x**必然映射到不同的输出向量**y**
  - 满射：整个目标空间中的每一个点**y**都能在输入空间中找到一个对应的点**x**
  - 可逆：存在一个逆变换 $\mathrm{T}^{-1}(\mathbf{x})=\mathbf{A}^{-1}\mathbf{x}$，可以将变换后的结果完美地还原回去
- 变换 T不会“压扁”空间（ $\det(\mathbf{A})\neq 0$ ）
  - 将 n 维空间映射到 n 维空间，而不是降维到一条线或一个平面。变换后图形的体积不为零

正交矩阵

- 正交矩阵 $\mathbf{P}$ 满足$\mathbf{P}^{-1} = \mathbf{P}^{\mathrm{T}}$。显然，正交矩阵是可逆矩阵
- 【正交矩阵所代表的线性变换只有旋转或者反射，没有缩放，剪切或形变】
- 行向量和列向量是标准正交基
  - 任意不同的列向量（或行向量）的点积为 0（互相垂直）；列向量（或行向量）模长为1
- 行列式的值为 +1 或 -1
  - 1，变换保向，代表旋转；-1，变换反向，代表反射或旋转加反射；体积没有变化
- 特征值的模长为 1
  - 变换不会在任何特征向量方向上进行缩放，与 " 保持长度 " 的几何特性一致
- 保持内积 / 点积
  - $\mathbf{Qx} \cdot \mathbf{Qy} = \mathbf{x} \cdot \mathbf{y}$
  - 保持向量间的夹角：向量之间的夹角，变换前后不变，没有发生剪切或形变
  - 保持向量长度：向量在经过正交变换后不会被拉伸或压缩

相似矩阵，对角化

相似性

【矩阵相似】假如$\mathbf{A}$和$\mathbf{B}$是$n \times n$矩阵，如果存在可逆矩阵$\mathbf{P}$，使得

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

则称$\mathbf{A}$相似于$\mathbf{B}$。(那么$\mathbf{B}$也相似于$\mathbf{A}$，$\mathbf{A} = \mathbf{PBP}^{-1}$)

称$\mathbf{A}$和B是相似的，把$\mathbf{A}$变成$\mathbf{P}^{-1}\mathbf{AP}$的变换称为相似变换

【定理】若$n \times n$矩阵$\mathbf{A}$和$\mathbf{B}$是相似的，那么它们有相同的特征多项式，从而有相同的特征值(和相同的重数)

【注】

即使有相同的特征值也可能不相似

相似性与行等价不一样。(假如$\mathbf{B}$行等价于$\mathbf{B}$，则存在可逆矩阵$\mathbf{E}$，使得$\mathbf{B} = \mathbf{EB}$。)对矩阵作行变换通常会改变矩阵的特征值

相似矩阵的几何意义

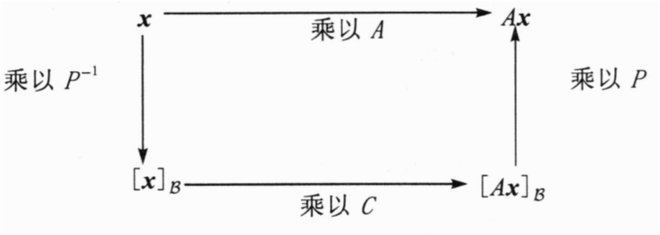
若A相似于C，即$\mathbf{A}=\mathbf{PCP}^{-1}$

记$\mathbf{P}=[\boldsymbol{b}_1\cdots\boldsymbol{b}_n]$，$\boldsymbol{\mathcal{B}}=\{\boldsymbol{b}_1, \cdots, \boldsymbol{b}_n\}$，则P是坐标变换矩阵$\mathbf{P}_{\boldsymbol{\mathcal{B}}}$，其中

$$\boldsymbol{x}=\mathbf{P}[\boldsymbol{x}]_{\boldsymbol{\mathcal{B}}}, \qquad [\boldsymbol{x}]_{\boldsymbol{\mathcal{B}}}=\mathbf{P}^{-1}\boldsymbol{x}$$

则C是变换$\mathbf{x}\mapsto\mathbf{Ax}$的$\boldsymbol{\mathcal{B}}$-矩阵【C代表变换在$\boldsymbol{\mathcal{B}}$坐标系下的矩阵】

即，$\mathbf{A}=\mathbf{PCP}^{-1}$



【几何意义】

相似矩阵为线性变换在不同坐标系下的，不同的矩阵表达形式

对向量的变换，可以看作：转换到$\boldsymbol{\mathcal{B}}$坐标系（$\mathbf{P}^{-1}$），变换（C），再回原来坐标系（P）

相似矩阵的不变量：特征值

- 所有相似的矩阵都具有相同的
  - 特征值（特征多项式）
    - 最重要的不变量！
    - 特征值代表了变换在特征向量上的缩放因子。这些“固有”的缩放方向和比例是变换本身的“特征”
  - 行列式
    - 变换对空间体积的缩放比例

- 矩阵相似性的最终应用和目标，就是对角化
  - 希望找到一个“最理想的视角”（由特征向量组成的基），在这个视角下，线性变换的表达形式变得极其简单(一个对角矩阵)

对角矩阵与对角化

【对角矩阵】对角矩阵(diagonal matrix)是一个主对角线之外的元素皆为0的矩阵。

$$\text{记作}\mathbf{D} = \operatorname{diag}(\lambda_1,\ \lambda_2,\ \cdots,\ \lambda_n) = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}$$

在很多情况下，从分解式$\mathbf{A} = \mathbf{PDP}^{-1}$(其中，$\mathbf{D}$是对角矩阵)，能够了解到有关矩阵$A$的特征值和特征向量的信息

【对角化】如果方阵A相似于对角矩阵，即存在可逆矩阵$\mathbf{P}$和对角矩阵$\mathbf{D}$，使得

$$\mathbf{A} = \mathbf{PDP}^{-1}$$

则称$\mathbf{A}$可对角化

对角化矩阵

【对角化】$\mathbf{A}_{n\times n}$有n个线性无关的特征向量，则$\mathbf{A}$可对角化
假设$\mathbf{A}$的n个线性无关的特征向量为$\mathbf{v}_1,\mathbf{v}_2\cdots\mathbf{v}_n$ （对应特征值$\lambda_1,\cdots,\lambda_n$）

令$\mathbf{P}=[\mathbf{v}_1\quad\mathbf{v}_2\quad\cdots\quad\mathbf{v}_n]$为特征向量基变换矩阵，对角阵$\mathbf{D}=\begin{bmatrix}\lambda_1 & 0 & \cdots & 0\\ 0 & \lambda_2 & \cdots & 0\\ \vdots & \vdots & & \vdots\\ 0 & 0 & \cdots & \lambda_n\end{bmatrix}$

那么，$\mathbf{AP}=[\mathbf{Av}_1\quad\mathbf{Av}_2\quad\cdots\quad\mathbf{Av}_n]$，$\mathbf{PD}=[\lambda_1\mathbf{v}_1\quad\lambda_2\mathbf{v}_2\quad\cdots\quad\lambda_n\mathbf{v}_n]$
λ_i和$\mathbf{v}_i$分别为特征值和特征向量，$\mathbf{Av}_\mathrm{i}=\lambda_\mathrm{i}\mathbf{v}_i$，则$\mathbf{AP}=\mathbf{PD}$

从而$\mathbf{A}=\mathbf{PDP}^{-1}$，于是$\mathbf{A}$对角化矩阵为$\mathbf{D}=\begin{bmatrix}\lambda_1 & 0 & \cdots & 0\\ 0 & \lambda_2 & \cdots & 0\\ \vdots & \vdots & & \vdots\\ 0 & 0 & \cdots & \lambda_n\end{bmatrix}$

【重要结论】
如果一个 n 阶方阵$\mathbf{A}$有 n 个互不相同的特征值，那么它一定可以被对角化
任何一个实对称矩阵 ($\mathbf{A}=\mathbf{A}^\mathrm{T}$) 都一定可以被对角化