线性代数5

矩阵分解

从整数分解谈矩阵特征分解

- 分解：分解为多个组成部分，找到表达方式不变的通用属性，帮助更好理解。

从整数分解谈矩阵特征分解

- 分解：分解为**多个组成部分**，找到**表达方式不变**的**通用属性**，帮助**更好理解**。

$$12 = 2 \times 2 \times 3$$

从整数分解谈矩阵分解

- 分解：分解为**多个组成部分**，找到**表达方式不变**的**通用属性**，帮助**更好理解**。

$$12 = 2 \times 2 \times 3$$

- 12不能被5整除
- 12的倍数可以被3整除

- 特征值分解
- 奇异值分解

特征值分解

- 如果说一个非零向量 v 是方阵 A 的特征向量，则一定可以表示成下面的形式：

$$Av = \lambda v$$

- λ 被称为特征向量 v 对应的特征值

特征值分解

- 如果说一个非零向量 v 是方阵 A 的特征向量，则一定可以表示成下面的形式：

$$Av = \lambda v$$

- λ 被称为特征向量 v 对应的特征值
- 性质：
  - 非零性
  - v是特征向量，则 sv 也是特征向量

特征值分解

- 特征值分解是将一个矩阵分解成下面的形式：

$$A = V diag(\lambda) V^{-1}$$

- V是矩阵A的线性无关的特征向量组成的矩阵

$$\mathrm{V} = \{v^{(1)}, v^{(2)}, \ldots, v^{(n)}\}$$

- Σ是一个对角阵，对角线上元素就是一个特征值 $\{v^{(1)}, v^{(2)}, \ldots, v^{(n)}\}$ 对应的特征向量

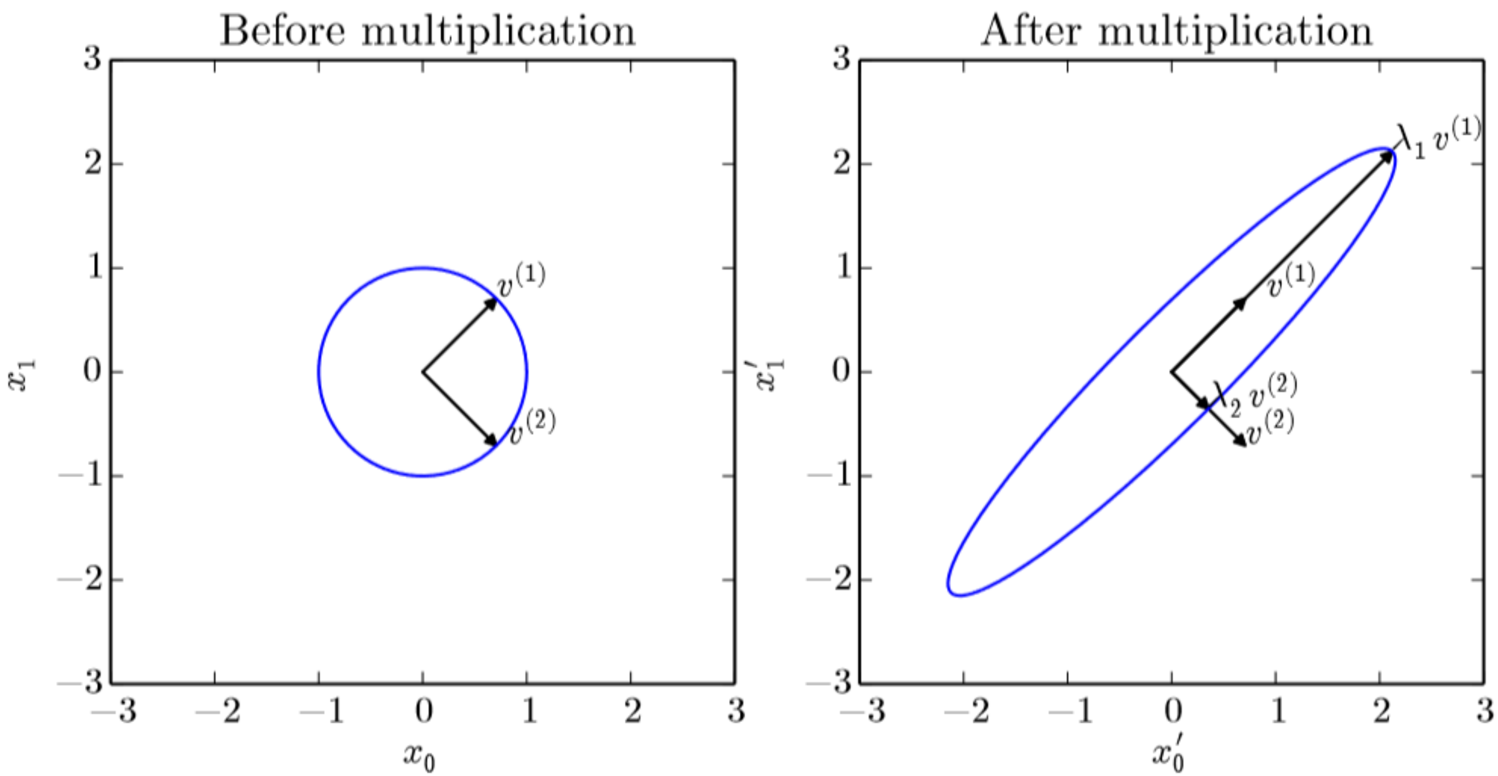
$$\{\lambda^{(1)}, \lambda^{(2)}, \ldots, \lambda^{(n)}\}$$

特征值分解

特征值分解

- 特征值分解的存在性
  - 不是每一个矩阵都可以特征分解
  - 实对称矩阵都可以特征分解
  - $A = Q\Lambda Q^T$
  - Q是A的特征向量组成的正交矩阵，Λ是对角矩阵
- 特征值分解的唯一性
  - 可能并不唯一

特征值分解



Before multiplication
After multiplication
$v^{(1)}$
$v^{(2)}$
$\lambda_1 v^{(1)}$
$\lambda_2 v^{(2)}$
x_0
x_1
x'_0
x'_1

特征值分解

- 正定矩阵
- 半正定矩阵
- 负定矩阵
- 半负定矩阵

$$\forall \boldsymbol{x}, \boldsymbol{x}^\top \boldsymbol{A} \boldsymbol{x} \geq 0$$

$$\boldsymbol{x}^\top \boldsymbol{A} \boldsymbol{x} = 0 \Rightarrow \boldsymbol{x} = \boldsymbol{0}$$

特征值分解

- 特征值分解是一个提取矩阵特征很不错的方法
- 但是它只是对方阵而言的，在现实的世界中，我们看到的大部分矩阵都不是方阵
- 比如说有N个学生，每个学生有M科成绩，这样形成的一个N * M的矩阵就不可能是方阵

我们怎样才能描述这样普通的矩阵呢的重要特征呢？

奇异值分解可以用来干这个事情，奇异值分解是一个能适用于任意矩阵的一种分解的方法

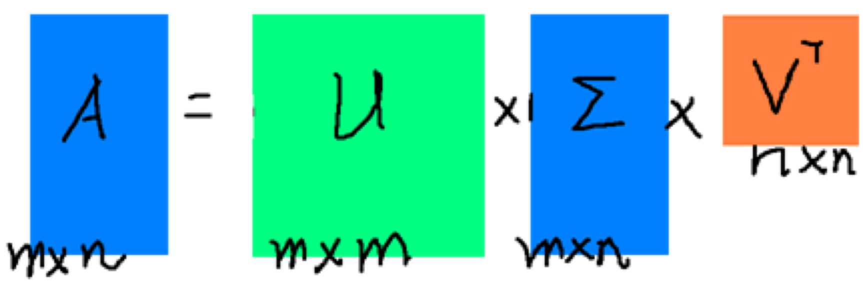
特征值分解

- 应用广泛
  - 每个实数矩阵都有一个奇异值分解，但不一定都有特征分解
  - 非方阵的矩阵没有特征分解，这时只能使用奇异值分解

奇异值分解

分解形式： $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^{\top}.$

- A 是一个N × M的矩阵
- U 是一个M × M的方阵
  - U的列向量称为左奇异向量
- Σ 是一个N × M的矩阵
  - 除了对角线的元素都是0，对角线上的元素称为奇异值
- $V^{\top}$ 是一个N × N的矩阵
  - $V^{\top}$列向量称为右奇异向量）



奇异值分解

- 那么奇异值和特征值是怎么对应起来的呢？
- 将一个矩阵A的转置 乘以 A，并将会得到一个方程：

$$(A^T A)v_i = \lambda_i v_i$$

- 我们利用这个方阵求特征值 λ_i 以及特征向量组 V，就是我们上面的右奇异向量。此外我们还可以得到：

$$\sigma_i = \sqrt{\lambda_i}$$

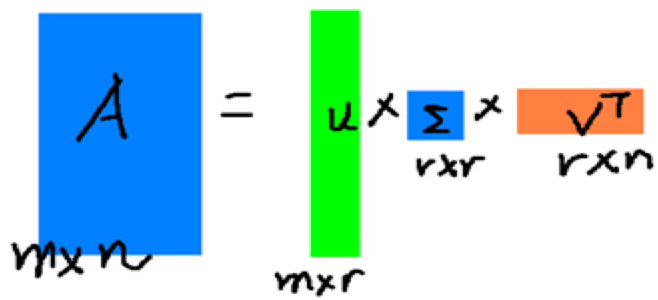
$$u_i = \frac{1}{\sigma_i} A v_i$$

- 这里的σ就是奇异值，u就是上面说的左奇异向量。

奇异值分解

奇异值σ跟特征值类似，在矩阵Σ中也是从大到小排列，而且σ的减少特别的快，在很多情况下，前10%甚至1%的奇异值的和就占了全部的奇异值之和的99%以上了。也就是说，我们也可以用前r（ r远小于m、n ）个的奇异值来近似描述矩阵，即部分奇异值分解：

$$A_{m\times n}=U_{m\times m}\Sigma_{m\times n}V_{n\times n}^{T}$$



右边的三个矩阵相乘的结果将会是一个接近于A的矩阵，在这儿，r越接近于n，则相乘的结果越接近于A。

奇异值分解

$$A_{m\times n}=U_{m\times m}\Sigma_{m\times n}V_{n\times n}^{T}$$

其实整个求解 SVD 的过程就是求解这 3 个矩阵的过程，而求解这 3 个矩阵的过程就是求解特征值和特征向量的过程，问题就在于 **求谁的特征值和特征向量。**

- U 的列由 AA^T 的单位化过的特征向量构成
- V 的列由 A^TA 的单位化过的特征向量构成
- Σ 的对角元素来源于 AA^T 或 A^TA 的特征值的平方根，并且是按从大到小的顺序排列的

1. 求 AA^T 的特征值和特征向量，用单位化的特征向量构成 U
2. 求 A^TA 的特征值和特征向量，用单位化的特征向量构成 V
3. 将 AA^T 或者 A^TA 的特征值求平方根，然后构成 Σ

奇异值分解

$$A=\begin{pmatrix}2&4\\1&3\\0&0\\0&0\end{pmatrix}$$

$$AA^T=\begin{pmatrix}20&14&0&0\\14&10&0&0\\0&0&0&0\\0&0&0&0\end{pmatrix}$$

$$AA^Tx=\lambda x$$

$$(AA^T-\lambda E)x=0$$

$$\begin{vmatrix}20-\lambda&14&0&0\\14&10-\lambda&0&0\\0&0&-\lambda&0\\0&0&0&-\lambda\end{vmatrix}=0$$

$$\lambda_1\approx 29.86606875,\ \lambda_2\approx 0.13393125,\ \lambda_3=\lambda_4=0$$

因为有4个特征，特征向量为4*4的矩阵

$$x=\begin{pmatrix}0.8174&-0.5766&0&0\\0.576&0.8174&0&0\\0&0&1&0\\0&0&0&1\end{pmatrix}$$

$$A_{m\times n}=U_{m\times m}\Sigma_{m\times n}V_{n\times n}^T$$

这个就是U

奇异值分解

同样的过程求解 A^TA 的特征值和特征向量，求得 $\lambda_1 \approx 0.13393125$，$\lambda_2 \approx 29.86606875$，将**特征值降序排列后**对应的单位化过的特征向量为

$$\begin{pmatrix} 0.40455358 & -0.9145143 \\ 0.9145143 & 0.40455358 \end{pmatrix}$$

$$A_{m\times n} = U_{m\times m}\Sigma_{m\times n}V_{n\times n}^T$$

这个就是V

奇异值分解

而矩阵 Σ 根据上面说的为特征值的平方根构成的对角矩阵

$$\begin{pmatrix} 5.4649857 & 0 \\ 0 & 0.36596619 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$

奇异值分解

到此，SVD 分解就结束了，原来的矩阵 A 就被分解成了 3 个矩阵的乘积。

$$A_{4\times 2} = U_{4\times 4}\Sigma_{4\times 2}V_{2\times 2}^T$$

$$\begin{pmatrix} 2 & 4 \\ 1 & 3 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$

$$= \begin{pmatrix} 0.81741556 & -0.57604844 & 0 & 0 \\ 0.57604844 & 0.81741556 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 5.4649857 & 0 \\ 0 & 0.36596619 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0.40455358 & -0.9145143 \\ 0.9145143 & 0.40455358 \end{pmatrix}^T$$