大数据机器学习第十次理论作业

常智星 2024214438

2026 年 1 月 22 日

一、 试求矩阵 $A=\begin{bmatrix}2&4\\1&3\\0&0\\0&0\end{bmatrix}$ 的奇异值分解并写出其外积展开式。

1. 问题分析与相关知识点

1). 奇异值分解

矩阵的奇异值分解是指，将一个非零的 $m\times n$ 实矩阵 A，$A\in R^{m\times n}$，表示为以下三个实矩阵乘积形式的运算，即进行矩阵的因子分解：

$$A=U\Sigma V^{\intercal} \tag{1}$$

其中，U 是 m 阶正交矩阵，V 是 n 阶正交矩阵，Σ 是由降序排列的非负对焦线元素组成的 $m\times n$ 矩形对角矩阵，满足：

$$UU^{\intercal}=I$$

$$VV^{\intercal}=I$$

$$\Sigma=(\sigma_1,\sigma_2,\cdots,\sigma_p)$$

$$\sigma_1\geq\sigma_2\geq\cdots\geq\sigma_p\geq 0$$

$$p=\min(m,n)$$

其中，$U\Sigma V^{\intercal}$ 称为矩阵的奇异值分解 (singular value decomposition, SVD)；σ_i 称为矩阵的奇异值；U 的列向量称为左奇异向量，V 的列向量称为右奇异向量。

2). 矩阵的基本子空间

I. 向量子空间

若 S 是向量空间 V 的非空子集，且 S 满足以下条件：

i. 对任意实数 a，若 $x\in S$，则 $ax\in S$；

ii. 若 $x\in S$ 且 $y\in S$，则 $x+y=S$；

则 S 称为 V 的子空间。

设 $v_1,v_2,\cdots,v_n$ 为向量空间 V 中的向量，则其线性组合

$$a_1v_1+a_2v_2+\cdots+a_nv_n$$

构成 V 的子空间，称 $v_1, v_2, \cdots, v_n$ 张成的子空间，或 $v_1, v_2, \cdots, v_n$ 的张成，记为：

$$\text{span}(v_1, v_2, \cdots, v_n)$$

如果 $\text{span}\{v_1, v_2, \cdots, v_n\} = V$，则称称 $v_1, v_2, \cdots, v_n$ 张成 V。

II. 向量空间的基和维数

向量空间 V 中向量 $v_1, v_2, \cdots, v_n$ 称为空间 V 的基，如果满足条件：

i. $v_1, v_2, \cdots, v_n$ 线性无关；

ii. $v_1, v_2, \cdots, v_n$ 张成 V；

向量空间基的个数即向量空间的维数。

III. 矩阵的行空间和列空间

设 A 为一个 m 行 n 列 $m \times n$ 的矩阵，其每一行可以看作 R^n 中的一个向量，称为 A 的行向量；类似地，其每一列可以看作 R^m 中的一个向量，称为 A 的列向量。

设 A 为一个 m 行 n 列 $m \times n$ 的矩阵，则由 A 的行向量张成的 R^n 的子空间，称为 A 的行空间；由 A 的列向量张成的 R^m 的子空间，称为 A 的列空间。

矩阵 A 的行空间的维数等于列空间的维数，称为矩阵的秩。

IV. 矩阵的零空间

设 A 为一个 m 行 n 列 $m \times n$ 的矩阵，令 $N(A)$ 为齐次方程组 $Ax = 0$ 的所有解的集合，则 $N(A)$ 为的 R^n 的一个子空间，称为 A 的零空间 (null space)：

$$N(A) = \{x \in R^n | Ax = 0\} \tag{2}$$

一个矩阵零空间的维数称为矩阵的零度。

【秩-零度定理】设 A 为一 $m \times n$ 的矩阵，则 A 的秩和 A 的零度之和为 n。若 A 的秩为 r，则方程组 $Ax = 0$ 的独立变量有 r 个，自由变量有 $n - r$ 个。

V. 子空间的正交补

设 X 和 Y 为 R^n 的子空间。若对每一个 $x \in X$ 和 $y \in Y$ 都满足 $x^\intercal y = 0$，则称 X 和 Y 是正交的，记作 $X \perp Y$。

令 Y 为 R^n 的子空间，R^n 中与 Y 中的每一个向量都正交的向量集合记为 $Y^\perp$，即

$$Y^\perp = \{x \in R | x^\intercal y = 0, \forall y \in Y\}$$

集合 $Y^\perp$ 称为 Y 的正交补。

VI. 矩阵的基本子空间

设 A 为一 $m \times n$ 的矩阵，可以把 A 看成是将 R^n 映射到 R^m 的线性变换。一个向量 $z \in R^m$ 在 A 的列空间的充要条件是存在 $x \in R^n$，使得 $z = Ax$。这样 A 的列空间和 A 的值域是相同的。记 A 的值域为 $R(A)$，则：

$$\begin{aligned} R(A) &= \{x \in R^m | \exists x \in R^n, z = Ax\} \\ &= A\text{的列空间} \end{aligned}$$

类似地，一个向量 $y \in R^n$，$y^\intercal$ 在 A 的行空间的充要条件是 $\exists x \in R^m$，使得 $y = A^\intercal x$。这样 A 的行空间和 $A^\intercal$ 的值域 $R(A^\intercal)$ 是相同的。

$$\begin{aligned} R(A^\intercal) &= \{y \in R^n | \exists x \in R^m, y = A^\intercal x\} \\ &= A\text{的行空间} \end{aligned}$$

VII. 矩阵的基本子空间

矩阵有 4 个基本子空间：列空间、行空间、零空间、A 的转置的零空间（左零空间），其关系如图 1 所示。

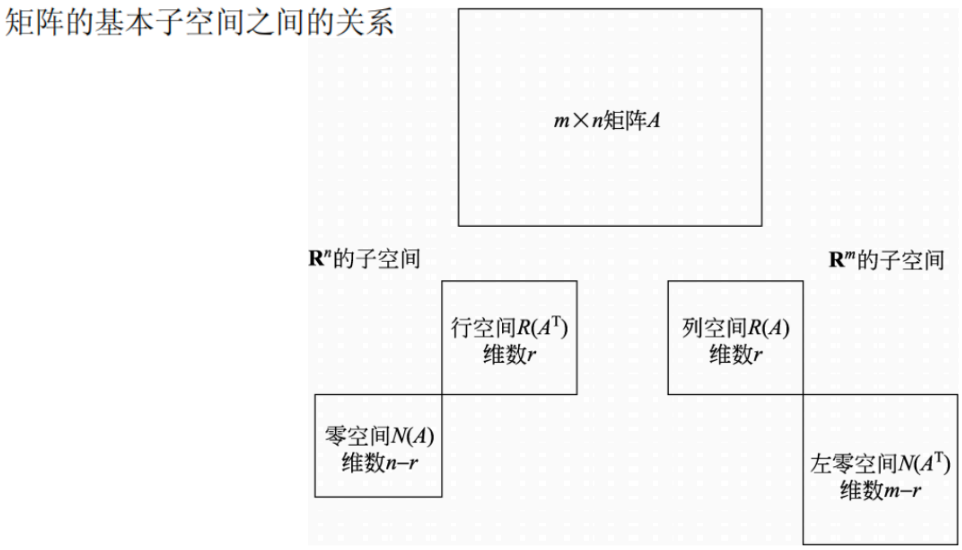

图 1: 矩阵的基本子空间的关系

有下面的定理成立：

【定理 1】若 A 为一 $m \times n$ 的矩阵，则 $N(A) = R(A^\intercal)^\perp$，且 $N(A^\intercal) = R(A)^\perp$。

3). 矩阵奇异值分解的计算

I. 求 $A^T A$ 的特征值和特征向量：

$$W = A^\intercal A$$

$$(W - \lambda I)x = 0$$

得到特征值 λ_i, 并将特征值从大到小排列：

$$\lambda_1 \geq \lambda_2 \geq \cdots \lambda_n \geq 0$$

并将特征值 $\lambda_i (i = 1, 2, \cdots, n)$ 代入特征方程求得对应的特征向量。

II. 求 n 阶正交矩阵 V：将特征向量单位化得到单位特征向量 $v_1, v_2, \cdots, v_n$，构成 n 阶正交矩阵 V:

$$V = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}$$

III. 求 $m \times n$ 对角矩阵 Σ: 计算 A 的奇异值：

$$\sigma_i = \sqrt{\lambda_i}, \quad i = 1, 2, \cdots, n$$

构造 $m \times n$ 对角矩，主对角线元素是奇异值，其余元素为 0:

$$\Sigma = \mathrm{diag}(\sigma_1, \sigma_2, \cdots, \sigma_n)$$

IV. 求 m 阶正交矩阵 U：对 A 的前 r 个正奇异值，令：

$$u_j = \frac{1}{\sigma_j} A v_j, \quad j = 1, 2, \cdots, r$$

得到

$$U_1 = \begin{bmatrix} u_1 & u_2 & \cdots & u_r \end{bmatrix}$$

求 A^T 的零空间的一组标准正交基 $\{u_{r+1}, u_{r+2}, \cdots, u_m\}$, 令

$$U_2 = \begin{bmatrix} u_{r+1} & u_{r+2} & \cdots & u_m \end{bmatrix}$$

并令

$$U = \begin{bmatrix} U_1 & U_2 \end{bmatrix}$$

V. 得到奇异值分解：

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

4). 矩阵的外积展开式

矩阵的奇异值分解 $U\Sigma V^\intercal$ 也可以由外积形式表示。如果把 A 的奇异值分解，看成矩阵 $U\Sigma$ 和 $V^\intercal$ 的乘积，将 $U\Sigma$ 按列向量分块，将 $V^\intercal$ 按行向量分块，即得到:

$$U\Sigma = \begin{bmatrix} \sigma_1 u_1 & \sigma_2 u_2 & \cdots & \sigma_n u_n \end{bmatrix}$$

$$V^\intercal = \begin{bmatrix} v_1^\intercal \\ v_2^\intercal \\ \vdots \\ v_n^\intercal \end{bmatrix}$$

则 A 的外积展开式：

$$A = \sigma_1 u_1 v_1^\intercal + \sigma_2 u_2 v_2^\intercal + \sigma_n u_n v_n^\intercal$$

即：

$$u_i v_j^\intercal = \begin{bmatrix} u_{1i} \\ u_{2i} \\ \vdots \\ u_{mi} \end{bmatrix} \begin{bmatrix} v_{1j} & v_{2j} & \cdots & v_{nj} \end{bmatrix} = \begin{bmatrix} u_{1i}v_{1j} & u_{1i}v_{2j} & \cdots & u_{1i}v_{nj} \\ u_{2i}v_{1j} & u_{2i}v_{2j} & \cdots & u_{2i}v_{nj} \\ \vdots & \vdots & \vdots & \vdots \\ u_{mi}v_{1j} & u_{mi}v_{2j} & \cdots & u_{mi}v_{nj} \end{bmatrix}$$

A 的外积展开式也可以写为：

$$A = \sum_{k=1}^{n} A_k = \sum_{k=1}^{n} \sigma_k u_k v_k^\intercal$$

由 A 的外积展开式可知，若 A 的秩为 n，则：

$$A = \sigma_1 u_1 v_1^\intercal + \sigma_2 u_2 v_2^\intercal + \sigma_n u_n v_n^\intercal$$

设矩阵

$$A_k = \sigma_1 u_1 v_1^\intercal + \sigma_2 u_2 v_2^\intercal + \sigma_k u_k v_k^\intercal$$

则 A_k 的秩为 k，并且 A_k 是秩为 k 的矩阵在弗罗贝尼乌斯范数意义 A 的最优近似矩阵，矩阵 A_k 就是 A 的截断奇异值分解。由于通常奇异值 σ_i 递减很快，所以 k 取很小时，A_k 也能对 A 有很好的近似。

2. 问题求解

1). 求矩阵的奇异值分解

I. 求 $A^{\intercal}A$ 的特征值和特征向量：

$$A^{\intercal}A=\begin{bmatrix}2&1&0\\4&3&0\end{bmatrix}\begin{bmatrix}2&4\\1&3\\0&0\\0&0\end{bmatrix}=\begin{bmatrix}5&11\\11&25\end{bmatrix}$$

特征值和特征向量满足方程:

$$(A^{\intercal}A-\lambda I)x=0$$

得到其次线性方程组：

$$\begin{cases}(5-\lambda)x_1+11\quad x_2=0\\11x_1+(25-\lambda)\quad x_2=0\end{cases}$$

方程组有解的充要条件是行列式 $\begin{vmatrix}5-\lambda&11\\11&25-\lambda\end{vmatrix}=0$

即

$$(5-\lambda)(25-\lambda)-121=\lambda^2-30\lambda+4=0$$

可求解得到特征值 λ_i, 并将特征值从大到小排列：

$$\lambda_1=15+\sqrt{221},\quad \lambda_2=15-\sqrt{221}$$

为方便计算，使用代码进行后续求解。

II. 代码实现

```
import numpy as np
from scipy.linalg import null_space

# 初始化A矩阵
A = np.array([[2,4],[1,3],[0,0],[0,0]],dtype= float)
m,n = A.shape

# 计算A的转置乘A，并求特征值、特征向量
ATA = np.array(A.T@A)
ATA_value,ATA_vector = np.linalg.eig(ATA)

# 对特征值按降序排序
argsort_value_index = np.argsort(-ATA_value)
ATA_sort_value = ATA_value[argsort_value_index]
# 按特征值降序排序特征向量作为V矩阵列向量并将其转置的VT矩阵
VT = ATA_vector[argsort_value_index]
print(f'V^T:\n{VT}')
# 对特征值取根号获取奇异值
sigma_list = []
for i in  range (0,ATA_sort_value.shape[0]):
    value = np.sqrt(ATA_sort_value[i])
    sigma_list.append(value)
print(f'奇异值:\n{sigma_list}')
```

```
# 构造m*n对角矩阵sigma_matrix
sigma_matrix = np.zeros((m,n))
sigma_matrix[: len(sigma_list),: len(sigma_list)] = np.diag(sigma_list)
print(f'对角矩阵:\n{sigma_matrix}')

# 计算U矩阵
U = np.zeros((m,m)) # 初始化U矩阵
for i in  range(m):
    # 计算U1
    if i <  len(sigma_list) and (sigma_list[i]>0):
        # i<=r时 计算u_j = 1/sigma_j*A*v_j
        uj = 1/sigma_list[i] * (A@ VT[i])
        U[i] = uj
    # 计算U2
    else:
         # i>r时 计算U2为A^T的零空间的标准正交基
        U2 = null_space(A.T).T
        for j in  range( len(U2)):
            U[i] = U2[j]
        break
# 转置得到U矩阵
U= U.T
print(f'U:\n{U}')

A_res = U@sigma_matrix@VT
# 验证结果
print(f'U * Sigma * V^T :\n{A_res}')
print(f'A:\n{A}')
```

结果

```
V^T:
[[ 0.40455358 -0.9145143 ]
 [-0.9145143  -0.40455358]]
奇异值:
[5.464985704219043, 0.3659661906262547]
对角矩阵:
[[5.4649857  0.        ]
 [0.         0.36596619]
 [0.         0.        ]
 [0.         0.        ]]
U:
[[-0.52130969 -9.41956667  0.          0.        ]
 [-0.4279955  -5.81522311  0.          0.        ]
 [ 0.          0.          0.          0.        ]
 [ 0.          0.          1.          0.        ]]
U * Sigma * V^T :
[[2. 4.]
 [1. 3.]
 [0. 0.]
 [0. 0.]]
```

```
A:
[[2. 4.]
 [1. 3.]
 [0. 0.]
 [0. 0.]]
```

2). 外积展开式表示

根据上面的计算结果可知：

$$\sigma_1 = 5.4650, \quad \sigma_2 = 0.3660$$

$$v_1^\intercal = \begin{bmatrix} 0.4046 & -0.9145 \end{bmatrix}, \quad v_2^T = \begin{bmatrix} -0.9145 & 0.4046 \end{bmatrix}$$

$$u_1 = \begin{bmatrix} -0.5213 \\ -0.4280 \end{bmatrix}, \quad u_2 = \begin{bmatrix} -9.4196 \\ -5.8152 \end{bmatrix}$$

A 的外积展开式：

$$A = \sigma_1 u_1 v_1^\intercal + \sigma_2 u_2 v_2^\intercal$$

代码实现

```python
# 外积展开式
# A = sigma_1u_1v_1^T+sigma_2u2v_2^T
A_outer = sigma_list[0]*np.outer(U[:,0],VT[0,:])+sigma_list[1]*np.outer(U[:,1],VT[1,:])
print(f'A外积展开式计算结果:\n{A_outer}')
```

结果

```
A外积展开式计算结果:
[[ 2.  4.]
 [ 1.  3.]
 [ 0. -0.]
 [ 0. -0.]]
```

二、 搜索中的点击数据记录用户搜索时提交的查询语句，点击的网页 URL 以及点击的次数构成一个二部图，其中一个结点集合 $\{q_i\}$ 表示查询，另一个结点集合 $\{u_j\}$ 表示 URL，边表示点击关系，边上的权重表示点击次数。图 15.2 是一个简化的点击数据例。点击数据可以由矩阵表示，试对该矩阵进行奇异值分解，并解释得到的三个矩阵所表示的内容

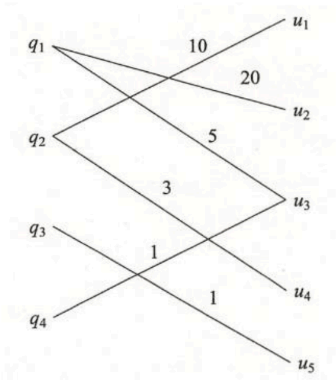

图15.2 搜索点击数据例

1. 问题求解

1). 写出二部图对应的矩阵 A

$$A = \begin{array}{c} \\ q_1 \\ q_2 \\ q_3 \\ q_4 \end{array} \begin{array}{c} \begin{array}{ccccc} u_1 & u_2 & u_3 & u_4 & u_5 \end{array} \\ \begin{bmatrix} 0 & 20 & 5 & 0 & 0 \\ 10 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \end{array}$$

2). 求矩阵 A 的奇异值分解：

```python
import numpy as np

A= [[0,20,5,0,0],[10,0,0,3,0],[0,0,0,0,1],[0,0,1,0,0]]
np.set_printoptions(precision=6,suppress=True)
U,sigma_matrix,V = np.linalg.svd(A)
print(f'U:\n{U}')
print(f'Sigma:\n{sigma_matrix}')
print(f'V:\n{V.T}')
```

结果

```
U:
[[ 0.99993 -0.        0.       -0.01179]
 [ 0.        1.        0.       -0.     ]
 [ 0.        0.        1.        0.     ]
 [ 0.01179  0.        0.        0.99993]]
Sigma:
[20.616958 10.440307  1.        0.970075]
```

```
V:
[[ 0.        0.957826 -0.       -0.        0.287348]
 [ 0.970008 -0.        0.       -0.243074 -0.      ]
 [ 0.243074  0.        0.        0.970008  0.      ]
 [ 0.        0.287348  0.        0.       -0.957826]
 [ 0.        0.        1.        0.        0.      ]]
```

3). 分析 3 个矩阵代表的含义

I. 矩阵 Σ：表示每个网页特征的重要程度。

II. 矩阵 V：V 的列向量表示每个 URL 与每个网页特征之间的关系。

矩阵 V 的第 1 列表示 URL1 的第 2 个特征较显著；第 2 列表示 URL2 的第 1 个特征较显著。

III. 矩阵 U：U 的列向量表示每个查询和每个网页特征之间的对应关系。

矩阵 U 的第 1 列表示第 1 个查询倾向于第 1 个特征较显著的 URL，即 URL2；矩阵 U 的第 2 列表示第 2 个查询倾向于第 2 个特征较显著的 URL，即 URL1。