SVD 习题

文档作者：李小山

说明：

- 最近完成了**Datawhale** 的项目“《统计学习方法》（第2版）习题解答” 中第15章奇异值分解的课后习题解答；

  项目地址：https://github.com/datawhalechina/statistical-learning-method-solutions-manual

- 这是我最初写的答案，不是很规范(尤其是代码部分，写得很不**pythonic**，自己需要再多学习学习漂亮的代码)，经过项目组织者**锋哥**的整理和修改，更加完善的版本可以去项目在线阅读（https://datawhalechina.github.io/statistical-learning-method-solutions-manual/）查看，这一不完善版本仅作为记录放在我的博客（https://github.com/L3Y1Q2/MyBrain）中。

课后习题

问1：试求矩阵 A 的奇异值分解

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

答：

- 方法1，对应函数 `svd1()`

  步骤：

  1. 求出 A^TA 的特征值和特征向量，并按特征值从大到小排列；
  2. 求 n 阶正交矩阵 V ；
  3. 求 $m\times n$ 对角矩阵 Σ ；
  4. 求出 AA^T 的特征值和特征向量，并按特征值从大到小排列；
  5. 求 m 阶正交矩阵 U 。

- 方法2，对应函数 `svd2()`

  步骤：

  1. 求出 A^TA 的特征值和特征向量，并按特征值从大到小排列；
  2. 求 n 阶正交矩阵 V ；
  3. 求 $m\times n$ 对角矩阵 Σ ；
  4. 通过公式 $Av_i=\sigma_iu_i$ 来求正交矩阵 U ，如果矩阵 A 的秩 r 小于 m ，矩阵 U 剩余部分通过求 AA^T 的特征向量来补齐。

代码：

```
import numpy as np

def diag(m,n,value):
    """
    因为不一定是方阵，所以单独写一个函数对角化。
    """
    matrixA = np.zeros((m,n))
    r = np.min([m,n])
    for i in range(r):
```

```
        matrixA[i,i] = value[i]
    return matrixA

def svd1(A):
    #第一步：计算对称矩阵 A^TA 的特征值与特征向量,，并按特征值从大到小排列
    W = np.dot(A.T,A)
    eigen_value1, eigen_vector1 = np.linalg.eigh(W)
    index1 = eigen_value1.argsort()[::-1]
    eigen_value1 = eigen_value1[index1]
    eigen_vector1 = eigen_vector1[:,index1]
    #第二步：构造正交矩阵 V
    V = eigen_vector1
    #第三步：求对角矩阵 Sigma
    m = A.shape[0]
    n = A.shape[1]
    sigma = np.sqrt(eigen_value1)
    Sigma = diag(m,n,sigma)
    #第四步：计算对称矩阵 AA^T 的特征值与特征向量,，并按特征值从大到小排列
    eigen_value2, eigen_vector2 = np.linalg.eigh(np.dot(A,A.T))
    index2 = eigen_value2.argsort()[::-1]
    eigen_value2 = eigen_value2[index2]
    eigen_vector2 = eigen_vector2[:,index2]
    #第五步：构造正交矩阵 U
    U = eigen_vector2

    return U, Sigma, V


def svd2(A):
    #第一步：计算对称矩阵 A^TA 的特征值与特征向量,，并按特征值从大到小排列
    eigen_value1, eigen_vector1 = np.linalg.eigh(np.dot(A.T,A))
    index1 = eigen_value1.argsort()[::-1]
    eigen_value1 = eigen_value1[index1]
    eigen_vector1 = eigen_vector1[:,index1]
    #第二步：构造正交矩阵 V
    V = eigen_vector1
    #第三步：求对角矩阵 Sigma
    m = A.shape[0]
    n = A.shape[1]
    sigma = np.sqrt(eigen_value1)
    Sigma = diag(m,n,sigma)
    #第四步：求正交矩阵 U
    r = np.linalg.matrix_rank(A)
    U = np.hstack([(np.dot(A, V[:, i]) / sigma[i])[:, np.newaxis] for i in
range(r)])
    if r<m:
        eigen_value2, eigen_vector2 = np.linalg.eigh(np.dot(A,A.T))
        index2 = eigen_value2.argsort()[::-1]
        eigen_value2 = eigen_value2[index2]
        eigen_vector2 = eigen_vector2[:,index2]
        U = np.hstack((U,eigen_vector2[:,r:]))

    return U, Sigma, V
if __name__ == "__main__":
    print("*****************第一题**********************")
```

```
    A1_real = np.array([[1, 2, 0],
                        [2, 0, 2]])
    U1, S1, V1 = svd1(A1_real)
    print("U=",U1)
    print("S=",S1)
    print("V=",V1)
    A1_calc1 = np.dot(np.dot(U1,S1),V1.T)
    print("A=",A1_calc1)
    print("--------------------------------------")
    U2, S2, V2 = svd2(A1_real)
    print("U=",U2)
    print("S=",S2)
    print("V=",V2)
    A1_calc2 = np.dot(np.dot(U2,S2),V2.T)
    print("A=",A1_calc2)
    print("--------------------------------------")
    U, S, V = np.linalg.svd(A1_real)
    print("U=",U)
    print("S=",S)
    print("V=",V.T)
    print("A=",np.dot(np.dot(U,S1),V))
```

问2：试求矩阵 A 的奇异值分解并写出其外积展开式

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

答：

- 写出矩阵 A 的外积展开式；

  摘录自《统计学习方法（第2版）》P291、294

  > 任意一个实矩阵 A 可以由其外积展开式表示
  >
  > $$A = \sum_{k=1}^{n} \sigma_k u_k v_k^T = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_n u_n v_n^T$$
  >
  > 其中 $u_k v_k^T$ 为 $m \times n$ 矩阵，是列向量 u_k 和行向量 v_k^T 的外积，σ_k 为奇异值，u_k, v_k^T, σ_k 通过矩阵 A 的奇异值分解得到。

- 按照习题15.1中的步骤进行奇异值分解计算；

代码：

```
if __name__ == "__main__":
    print("*****************第二题**********************")
    A2_real = np.array([[2, 4],
                        [1, 3],
                        [0, 0],
```

```
                    [0,0]])
U1, S1, V1 = svd1(A2_real)
print("U=",U1)
print("S=",S1)
print("V=",V1)
A2_calc1 = np.dot(np.dot(U1,S1),V1.T)
print("A=",A2_calc1)
print("--------------------------------------")
U2, S2, V2 = svd2(A2_real)
print("U=",U2)
print("S=",S2)
print("V=",V2)
A2_calc2 = np.dot(np.dot(U2,S2),V2.T)
print("A=",A2_calc2)
print("--------------------------------------")
U, S, V = np.linalg.svd(A2_real)
print("U=",U)
print("S=",S)
print("V=",V.T)
print("A=",np.dot(np.dot(U,S1),V))
```

- 将奇异值分解的求解得到的奇异值、行向量与列向量带入外积展开式即可。

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

  其中

  $$\sigma_1 = 5.4649857, \sigma_2 = 0.36596619$$

  $$u_1 = \begin{bmatrix} 0.81741556 \\ 0.57604844 \\ 0 \\ 0 \end{bmatrix}, u_2 = \begin{bmatrix} -0.57604844 \\ 0.81741556 \\ 0 \\ 0 \end{bmatrix}$$

  $$v_1^T = [0.40455358, 0.9145143], v_2^T = [-0.9145143, 0.40455358]$$

  ```
  S2[0][0]*np.outer(U2[:,0],V2[:,0].T)+S2[1][1]*np.outer(U2[:,1],V2[:,1].T)
  ```

问3：比较矩阵的奇异值分解与对称矩阵的对角化的异同

解答：

注意： 与书中一致，不考虑复数矩阵。

解答思路：

1. 列出矩阵奇异值分解的定义；
2. 列出对称矩阵和方阵对角化定义；
3. 对比两者之间的异同。

解答步骤：

第1步：列出矩阵奇异值分解的定义；

矩阵的奇异值分解定义如下：

摘录自《统计学习方法（第2版）》P271

> 矩阵的奇异值分解是指，将一个非零的 $m \times n$ 实矩阵 $A, A \in \boldsymbol{R}^{m\times n}$，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：
>
> $$A = U\Sigma V^T$$
>
> 其中 U 是 m 阶正交矩阵（orthogonal matrix），V 是 n 阶正交矩阵，Σ 是由降序排列的非负的对角元素组成的 $m \times n$ 矩形对角矩阵（rectangular diagonal matrix）.
>
> - $UU^T = I$
> - $VV^T = I$
> - $\Sigma = diag(\sigma_1, \sigma_2, \cdots, \sigma_p)$
> - $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0$
> - $p = \min(m, n)$
>
> σ_i 称为矩阵 A 的奇异值（singular value），U 的列向量称为左奇异向量（left singular vector），V 的列向量称为右奇异向量（right singular vector）。

第2步：列出对称矩阵和方阵对角化定义；

对称矩阵的对角化定义如下：

摘录自 张丽静, 刘白羽, 申亚男. 实对称矩阵对角化教学的应用案例[J]. 大学数学, 2019, 35(2):6.

> 设 A 为 n 阶实对称矩阵，则存在一个正交矩阵 $P = (\boldsymbol{p_1}, \boldsymbol{p_2}, \cdots, \boldsymbol{p_n})$，使得
>
> $$A = P\Lambda P^T$$
>
> 其中 Λ 为对角元素是 A 的特征值 $\lambda_1, \lambda_2, \cdots, \lambda_n$ 所构成的对角阵；$\boldsymbol{p_1}, \boldsymbol{p_2}, \cdots, \boldsymbol{p_n}$ 为特征值 $\lambda_1, \lambda_2, \cdots, \lambda_n$ 对应的单位特征向量。

摘录自 https://www.matongxue.com/parts/4628

> 如果 n 阶方阵 A 有 n 个线性无关的特征向量 $\boldsymbol{p_1}, \boldsymbol{p_2}, \cdots, \boldsymbol{p_n}$，那么如下矩阵：
>
> $$P = (\boldsymbol{p_1}, \boldsymbol{p_2}, \cdots, \boldsymbol{p_n})$$
>
> 可以使得：$A = P\Lambda P^{-1}$，其中 Λ 为对角阵 $\Lambda = diag\{\lambda_1, \lambda_2, \cdots, \lambda_n\}$
>
> 特征值 $\lambda_1, \lambda_2, \cdots, \lambda_n$ 为特征向量 $\boldsymbol{p_1}, \boldsymbol{p_2}, \cdots, \boldsymbol{p_n}$ 对应的特征值，该过程为**对角化（Diagonalizable）**。

第3步：对比两者之间的异同。

相同：

- 两者都可以实现特征分解，都需要求解特征值与特征向量；
- 实对称矩阵的对角化是方阵对角化的特例，矩阵奇异值分解又是方阵对角化的推广；
- 实对称矩阵一定可以对角化，矩阵的奇异值分解也一定存在；
- 都可以被正交矩阵对角化；
- 本质上都是用不同的基来描述线性变化；

不同：

- 矩阵对角化要求矩阵 A是一个方阵，矩阵奇异值分解不需要对矩阵 A 有特殊要求；
- 奇异值分解中要求对角元素降序排列，对称矩阵的对角化没有这一要求；

问4：证明任何一个秩为1的矩阵可写成两个向量的外积形式，并给出实例。

解答：

解答思路：

1. 假设矩阵 A 的秩为1，写出其奇异值分解形式；
2. 利用矩阵A 的秩与对角矩阵 Σ 的秩之间的关系，以及奇异值分解的性质，将矩阵Σ 写成外积形式；
3. 将矩阵A 的奇异值分解写成外积形式；
4. 给出实例。

解答步骤：

第1步：假设矩阵 A 的秩为1，写出其奇异值分解形式；

设矩阵 $A_{m\times n}$ 的秩为1，因为矩阵的奇异值分解 $A=U\Sigma V^T$一定存在（奇异值分解基本定理）

第2步：利用矩阵A 的秩与对角矩阵 Σ 的秩之间的关系，以及奇异值分解的性质，将矩阵Σ 写成外积形式；

因为 $rank(\Sigma)=rank(A)=1$ 。

且对角矩阵 Σ 的奇异值是降序排列的，所以其中的非零元素一定位于第一行第一列，设为 $\begin{bmatrix} \sigma_1 & \\ & O_{(m-1)\times(n-1)} \end{bmatrix}_{m\times n}$。

设两向量 $a=[1,0,\cdots,0]^T_{m\times 1}, b=[\sigma_1,0,\cdots,0]^T_{n\times 1}$，可得：$\Sigma=ab^T$

第3步：将矩阵A 的奇异值分解写成外积形式；

将上式$\Sigma=ab^T$代入分解矩阵可得：$A=Uab^TV^T=(Ua)(Vb)^T$

其中 Ua 是 $m\times 1$ 列向量，Vb 是$n\times 1$ 列向量，矩阵 A 可以写成向量 Ua 和向量 Vb 的外积形式。

第4步：给出实例。

书上例子 15.5 中矩阵 $A=\begin{bmatrix}1 & 1\\ 2 & 2\\ 0 & 0\end{bmatrix}$ 的秩为1，其奇异值分解为

$$A=U\Sigma V^T=\begin{bmatrix}\frac{1}{\sqrt5} & -\frac{2}{\sqrt5} & 0\\ \frac{2}{\sqrt5} & \frac{1}{\sqrt5} & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}\sqrt{10} & 0\\ 0 & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt2} & \frac{1}{\sqrt2}\\ \frac{1}{\sqrt2} & -\frac{1}{\sqrt2}\end{bmatrix}$$

$$\Sigma=\begin{bmatrix}\sqrt{10} & 0\\ 0 & 0\\ 0 & 0\end{bmatrix}=\begin{bmatrix}1\\0\\0\end{bmatrix}[\sqrt{10},0]=ab^T\ ,\ Ua=\begin{bmatrix}\frac{1}{\sqrt5}\\ \frac{2}{\sqrt5}\\ 0\end{bmatrix}\ ,\ Vb=\begin{bmatrix}\sqrt5\\ \sqrt5\end{bmatrix}$$

$$(Ua)(Vb)^T=\begin{bmatrix}1 & 1\\ 2 & 2\\ 0 & 0\end{bmatrix}=A$$

问5：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，构成了一个二部图，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。

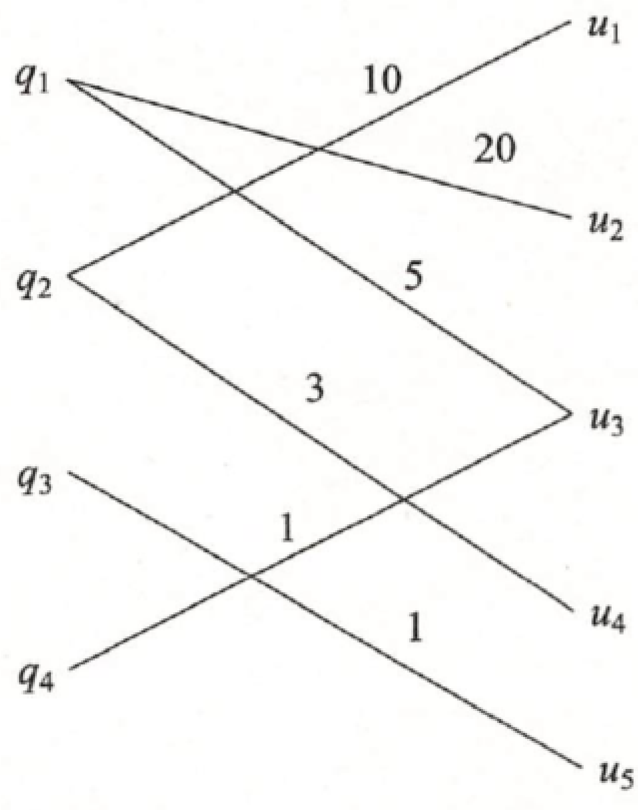

图 15.2　搜索点击数据例

答:

- 第一步：根据二部图写出对应矩阵

$$A=\begin{bmatrix}0 & 20 & 5 & 0 & 0\\ 10 & 0 & 0 & 3 & 0\\ 0 & 0 & 0 & 0 & 1\\ 0 & 0 & 1 & 0 & 0\end{bmatrix}$$

  其中，行向量分别对应 (q_1,q_2,q_3,q_4) ，列向量分别对应(u_1,u_2,u_3,u_4,u_5) 。

- 第二步：对矩阵 A 进行奇异值分解

$$A=\begin{bmatrix}0&20&5&0&0\\10&0&0&3&0\\0&0&0&0&1\\0&0&1&0&0\end{bmatrix}=U\Sigma V^T$$

$$U=\begin{bmatrix}-0.99993.5&0&0&0.01178999\\0&-1&0&0\\0&0&1&0\\-0.01178999&0&0&-0.99993.5\end{bmatrix}$$

$$\Sigma=\begin{bmatrix}20.61695792&0&0&0&0\\0&10.44030651&0&0&0\\0&0&1&0&0\\0&0&0&0.97007522&0\end{bmatrix}$$

$$V^T=\begin{bmatrix}0&-0.970008&-0.243074&0&0\\-0.957826&0&0&-0.287348&0\\0&0&0&0&1\\-0&0.243074&-0.970008&0&0\\-0.287348&0&0&0.957826&0\end{bmatrix}$$

- 第三步：解释奇异值分解得到的矩阵 U、Σ、V 代表的含义

  简单理解：

  U 表示用户输入的查询语句与网页特征之间的关系矩阵

  V 表示网页与网页特征之间的相似性

  Σ 表示用户输入的查询语句映射到网页的权重

  举个例子: 从矩阵 U 的第一列可以看成查询语句 q_1 与网页特征1的关系密切，通过矩阵 V 可以看成网页 u_2 与特征1的相似度最高，所以查询 q_1 的时候点击网页 u_2 的次数是最多的。

  奇异值分解要做的就是提取网页的特征来将用户输入与网页映射到一个低纬度空间中。

  通过第二步的计算，我们发现σ_1, σ_2 的值比较大，计算 $A'=\sigma_1 u_1 v_1^T+\sigma_2 u_2 v_2^T$ 可以发现其与矩阵A 基本相等。

参考资料：

1. http://charleshm.github.io/2016/03/Singularly-Valuable-Decomposition/# 漫谈奇异值分解
2. https://blog.sciencenet.cn/blog-696950-699432.html 奇异值分解(SVD) --- 几何意义
3. http://lazynight.me/3355.html 一个关于SVD的例子
4. https://zhuanlan.zhihu.com/p/392588258 不错的文章
5. https://www.cnblogs.com/bigmonkey/p/11913079.html 对称矩阵的一些性质
6. https://zhuanlan.zhihu.com/p/353637184?utm_medium=social&utm_oi=643096634247090176 奇异值与特征值辨析
7. https://zhuanlan.zhihu.com/p/43578482
8. https://math.stackexchange.com/questions/2359992/how-to-resolve-the-sign-issue-in-a-svd-problem
9. http://web.stanford.edu/class/cs246/slides/06-dim_red.pdf
10.