第十二章重点（特征值与奇异值）

特征值与特征向量

- **Power Iteration计算主特征值与特征向量**
  - 主特征值的定义：
    The **dominant eigenvalue** $\lambda_{\max}$ of A: an eigenvalue λ_i s.t. $|\lambda_i| > |\lambda_j|$ for $i \neq j$.（最大的特征值）
  - 主特征向量的定义：
    If dominant eigenvalue exists, an eigenvector $v_{\max}$ associated to $\lambda_{\max}$ is called a **dominant eigenvector**.（最大特征值对应的特征向量）
  - 用一个矩阵A对一个随机选取的向量x不断地进行线性变换，即$\lim\limits_{n\to\infty} A^n x$，结果将会无限逼近矩阵$A$的主特征向量**的倍数**
  - 已知特征向量$v_{\max}$，如何计算特征值$\lambda_{\max}$：
    - 利用**Rayleigh quotient**（瑞利商）：$\lambda = \dfrac{v^\top A v}{v^\top v}$，这是为了避免用上面最开始的方法得到主特征向量倍数的结果太大，因此对结果进行了**标准化**（normalization）
      瑞利商推导过程：

$$Av = \lambda v \Rightarrow v^\top A v = v^\top \lambda v \Rightarrow v^\top A v = \lambda v^\top v \Rightarrow \lambda = \frac{v^\top A v}{v^\top v}$$

  - 最终Power Iteration的算法流程如下：

$$\begin{aligned}
&x_0 = \text{Initial Vector} \\
&\textbf{for } j = 1, 2, \dots \textbf{ do} \\
&\quad v_{j-1} = \frac{x_{j-1}}{||x_{j-1}||_2} \\
&\quad x_j = A v_{j-1} \\
&v_{\max} = \frac{x_j}{||x_j||_2} \\
&\lambda_{\max} = v_{\max}^\top A v_{\max}
\end{aligned}$$

- **Power Iteration收敛性的证明**
  - **Power Iteration收敛性定理**：
    Let A be an $n \times n$ matrix with real eigenvalues $\lambda_1, \dots, \lambda_n$ satisfying $|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \cdot \geq |\lambda_n|$. Assume that the eigenvalues of A span R^n. For almost every initial vector, Power Iteration converges to an eigenvector associated to λ_1.
  - **证明**：
    Let $v_1, \cdots, v_n$ be the eigenvectors w.r.t. $\lambda_1, \dots, \lambda_n$ respectively.
    The initial vector x_0 can be expressed as $c_1 v_1 + \cdots + c_n v_n$ where $c_1 \neq 0$.
    Applying Power Iteration yields:

$$\begin{aligned}
A x_0 &= c_1 \lambda_1 v_1 + \cdots + c_n \lambda_n v_n \\
A^2 x_0 &= c_1 \lambda_1^2 v_1 + \cdots + c_n \lambda_n^2 v_n \\
&\vdots \\
A^k x_0 &= c_1 \lambda_1^k v_1 + \cdots + c_n \lambda_n^k v_n
\end{aligned}$$

$$\lim_{k\to\infty} \frac{A^k x_0}{\lambda_1^k} = \lim_{k\to\infty} [c_1 v_1 + \cdots + c_n (\frac{\lambda_n}{\lambda_1})^k v_n] = c_1 v_1$$

- 学会如何**介绍Page Rank算法**（写一个段落来介绍这个算法，介绍它的主要思想/目的/大致思路，如果特征值不满足要求，如何调整Power Iteration方法）
  - Page Rank算法用于**衡量每个网页的相对重要性**。算法的核心是为每个网页i分配一个分数r_i，代表其重要性。如果$r_i > r_j$，则网页i比网页j更重要。**所有网页的分数需要规范化，使得所有分数之和为1。**
  - Page Rank的核心观点包括：
    - 重要的网页会从其他重要网页获得链接，**网页通过链接给其他网页“投票”**，链接越多，该网页越重要；
    - **一个网页的投票权只有一次**，如果它**链接到多个网页，则其投票权需要分割**；
    - 一个具有**高分的网页**可能具有**来自其他高分网页的链接**或有**指向高分网页的外链**；
    - 来自**重要网页的链接**比来自**不重要网页的链接**对提高网页重要性的**作用更大。**
  - 如果在算法中遇到特征值不满足要求的情况，如矩阵特征值不为1或难以计算特征值，需要调整Power iteration方法。
    - 如果**邻接矩阵的特征值不为**1：
      使邻接矩阵具有特征值1，我们可以将矩阵转化为**列随机矩阵**（column stochastic），列随机矩阵定义如下：
      - 所有条目都非负

        - 每一列的元素和都为1
        - 1总是特征值
        - 因为每一列的元素和都为1，所以其最大列和也为1，也称1-范数$\max(||A||_1) = 1$
    - 矩阵的特征值为1，但**获取的特征向量的计算成本很高**：
      根据**Perron-Frobenius定理**：如果矩阵A中的每个条目都是正的，那么其存在一个正实数λ为A的主特征值。在Page Rank构建的列随机矩阵中，这个定理保证1不仅是一个特征值，而且是最大的特征值。
      因此，我们要做的事情就是保证Page Rank构建的矩阵是
        - **列随机的**（column stochastic）
        - **符号为正的**（positive）
        - **稀疏的**（sparse）

奇异值分解（SVD）

- **SVD计算**
    - **定义**：
      令A为一个$m \times n$的矩阵，U: $\{u_1, \cdots, u_m\}$和V: $\{v_1, \cdots, v_n\}$是两个正交基集，同时有$s_1, \cdots, s_n$满足当$1 \leq i \leq n$时$Av_i = s_i u_i$，此时：
        - v_i称为A的右奇异向量
        - u_i称为A的左奇异向量
        - s_i称为A的奇异值

  $USV^\top$是A的奇异值分解（Singular Value Decomposition），其中S是一个$m \times n$的对角矩阵，其对角线上的条目是s_i，$s_i = \sqrt{\lambda_i}$
    - **求U，S，V矩阵**：
      设$m \leq n$，给定以下量：
        - $\lambda_1, \lambda_2, \cdots, \lambda_n$：$A^\top A$的特征值集合，其中$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$
        - $v_1, v_2, \cdots, v_n$是上述特征值对应的特征向量

  各个s_i和$u_i (1 \leq i \leq m)$可由以下公式计算得到：
    - $s_i = \sqrt{\lambda_i}$
    - $u_i = \begin{cases} \frac{Av_i}{s_i} & \text{if } \lambda_i \neq 0 \\ \text{an unit vector orthogonal to } u_1, u_2, \cdots, u_{i-1} & \text{otherwise} \end{cases}$
- **SVD引理、推论和性质的证明**
    - **引理1**：令A为一个$m \times n$的矩阵，$A^\top A$和$AA^\top$的特征值都是非负的
        - 证明：令v为$A^\top A$的一个单位特征向量且$A^\top Av = \lambda v$。那么$0 \leq ||Av||^2 = v^\top A^\top Av = \lambda v^\top v = \lambda$
    - **推论1**：$AA^\top$和$A^\top A$是对称的
        - 证明：$(AA^\top)^\top = (A^\top)^\top A^\top = AA^\top$。同理，$(A^\top A)^\top = A^\top A$
    - **推论2**：$u_1, u_2, \cdots, u_m$构成了$AA^\top$的组标准正交（orthonormal）特征向量集合
        - 证明：
            - 单位性（Unitary）：$u_i^\top u_i = \frac{Av_i}{s_i}^\top \frac{Av_i}{s_i} = \frac{v_i^\top A^\top Av_i}{s_i^2} = \frac{\lambda_i v_i^\top v_i}{\lambda_i} = 1$
            - 正交性（Orthogonality）：$u_i^\top u_j = \frac{Av_i}{s_i}^\top \frac{Av_j}{s_j} = \frac{v_i^\top A^\top Av_j}{s_i s_j} = \frac{\lambda v_i^\top v_j}{s_i s_j} = 0$，其中$i \neq j$
            - 特征向量（Eigenvector）：$AA^\top u_i = \frac{AA^\top Av_i}{s_i} = s_i^2 \frac{Av_i}{s_i} = \lambda_i u_i$
    - **推论3**：矩阵$A = USV^\top$的秩（rank）是S中非零条目的数量
        - 定义：
          秩（rank）在一个$m \times n$的矩阵A中代表的是其线性无关的行（列）的数量
        - 证明：
          U和V是正交的，因此它们也是可逆的，$\mathrm{rank}(A) = \mathrm{rank}(S)$，因为$S$是个对角矩阵，所以$A$的秩是$S$中非零条目的数量
    - **推论4**：如果A是一个$n \times n$的矩阵，那么$|\det(A)| = s_1 \cdots s_n$
        - 证明：
          因为$U^\top U = I$，$\det(U) = 1$或$\det(U) = -1$，同理$\det(V^\top) = \det(V) = 1$或$\det(V^\top) = \det(V) = -1$。因此$|\det(A)| = |\det(USV^\top)| = |\det(U)| \cdot |\det(S)| \cdot |\det(V^\top)| = s_1 \cdots s_n$
    - **推论5**：如果A是一个可逆的$n \times n$矩阵，那么$A^{-1} = VS^{-1}U^\top$
        - 证明：
          我们可知S矩阵同样是可逆的，根据SVD的定义，U和$V^\top$也是可逆的，拓展可得$U^{-1} = U^\top$和$(V^\top)^{-1} = V$，最终得到$A^{-1} = (USV^\top)^{-1} = (V^\top)^{-1} S^{-1} U^{-1} = VS^{-1}U^\top$
    - **推论6**：$m \times n$的矩阵A可以被改写成rank为1的矩阵的和，即

$$A = \sum_{i=1}^{r} s_i u_i v_i^\top$$

      其中r是A的秩，且u_i和v_i分别是U和V的第i列

- 证明:

$$
\begin{aligned}
A = USV^\top &= U \begin{bmatrix} s_1 & & \\ & \ddots & \\ & & s_r \end{bmatrix} V^\top \\
&= U(\begin{bmatrix} s_1 & & \\ & & \\ & & \end{bmatrix} + \begin{bmatrix} & & \\ & s_2 & \\ & & \end{bmatrix} + \cdots + \begin{bmatrix} & & \\ & & \\ & & s_r \end{bmatrix})V^\top \\
&= s_1 u_1 v_1^\top + s_2 u_2 v_2^\top + \cdots + s_r u_r v_r^\top
\end{aligned}
$$