



计算方法

刘景铨

计算机软件新技术国家重点实验室
南京大学



奇异值 (singular values)

通知：第5次作业将于今天稍后发布

对于一般的 $m \times n$ 实数矩阵 A ，可以有奇异值分解：

$$A = U S V^T$$

A 的形状： $m \times n$

U 的形状： $m \times m$

S 的形状： $m \times n$

V 的形状： $n \times n$

$$U^T U = I_m, \quad V^T V = I_n$$

观察：

$$A A^T = U S S^T U^T$$

$$A^T A = V S^T S V^T$$

$A A^T$ 和 $A^T A$ 都是实数对称阵，可做特征值分解！



SVD

定理： AA^T 和 $A^T A$ 有着相同的非零特征值。

证明： 考虑 AA^T 的特征值 $\lambda \neq 0$ 和特征向量 u : $AA^T u = \lambda u$

$$A^T A (A^T u) = A^T (AA^T u) = A^T (\lambda u) = \lambda (A^T u)$$

可见，只要 $A^T u \neq 0$ ，则 $A^T u$ 为 $A^T A$ 的一个特征向量，特征值也是 λ 。不妨假设 $A^T u = 0$ ，则 $AA^T u = A(A^T u) = 0$ ，与 $\lambda \neq 0$ 矛盾。

类似地，考虑 $A^T A$ 的特征值 $\lambda \neq 0$ 和特征向量 v : $A^T A v = \lambda v$ ，也有

$$AA^T (Av) = A(A^T A v) = A(\lambda v) = \lambda (Av).$$

同理有 $Av \neq 0$ ，因此 Av 为 AA^T 的一个特征向量，特征值也是 λ 。



SVD

AA^T 和 $A^T A$ 有着相同的非零特征值

把它们记作 $\{\lambda_i\}$ ，另外对 AA^T 和 $A^T A$ 进行谱分解，可得到正规正交的向量 $\{\mathbf{u}_i\}, \{\mathbf{v}_i\}$ 作为它们的特征向量：

$$AA^T \mathbf{u}_i = \lambda_i \mathbf{u}_i$$

$$A^T A \mathbf{v}_i = \lambda_i \mathbf{v}_i$$

把 $\{\mathbf{u}_i\}$ 组合成矩阵 \mathbf{U} ， $\{\mathbf{v}_i\}$ 组合成矩阵 \mathbf{V} ，则有 $\mathbf{U}^T \mathbf{U} = \mathbf{I}_m$ ， $\mathbf{V}^T \mathbf{V} = \mathbf{I}_n$

回顾之前的证明：

- 若 \mathbf{u} 为 AA^T 的一个特征向量，则 $A^T \mathbf{u}$ 为 $A^T A$ 的一个特征向量，且有着相同的特征值；
- 若 \mathbf{v} 为 $A^T A$ 的一个特征向量，则 $A \mathbf{v}$ 为 AA^T 的一个特征向量，且有着相同的特征值



SVD

回顾之前的证明:

- 若 \mathbf{u} 为 $\mathbf{A}\mathbf{A}^T$ 的一个特征向量, 则 $\mathbf{A}^T\mathbf{u}$ 为 $\mathbf{A}^T\mathbf{A}$ 的一个特征向量, 且有着相同的特征值;
- 若 \mathbf{v} 为 $\mathbf{A}^T\mathbf{A}$ 的一个特征向量, 则 $\mathbf{A}\mathbf{v}$ 为 $\mathbf{A}\mathbf{A}^T$ 的一个特征向量, 且有着相同的特征值

$$\mathbf{A}\mathbf{A}^T\mathbf{u}_i = \lambda_i\mathbf{u}_i$$

$$\mathbf{A}^T\mathbf{A}\mathbf{v}_i = \lambda_i\mathbf{v}_i$$

$$\mathbf{A}^T\mathbf{u}_i = \sqrt{\lambda_i}\mathbf{v}_i$$

$$\mathbf{A}\mathbf{v}_i = \sqrt{\lambda_i}\mathbf{u}_i$$

不失一般性地假设 $m > n$, 则有

$$\begin{aligned} \mathbf{U}^T\mathbf{A}\mathbf{V} &= \mathit{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n}) \\ \Rightarrow \mathbf{A} &= \mathbf{U}\mathbf{S}\mathbf{V}^T \end{aligned}$$

\mathbf{S} 的对角线元素被称作奇异值(singular values), 通常记作 $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n \geq 0$

从几何上看, 矩阵 \mathbf{A} 的作用为: \mathbb{R}^n 上的旋转变换、伸缩变换和 \mathbb{R}^m 上的旋转变换



SVD与Rayleigh quotient

回顾特征值的min-max刻画, 记 $R_B(x) := \frac{x^T B x}{x^T x}$
对于对称的矩阵 B , $R_B(x)$ 在特征值处取到极值

回顾 $\|A\|_2 := \max_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2}$, $cond_2(A) := \|A\|_2 \cdot \|A^{-1}\|_2$
对于对称正定的矩阵 A ,

$$\|A\|_2 = \lambda_{max}(A)$$
$$cond_2(A) = \lambda_{max}(A) / \lambda_{min}(A)$$

对于一般的矩阵呢?

$$\|A\|_2 = \sqrt{\lambda_{max}(A^T A)} = \sigma_{max}(A)$$
$$cond_2(A) = \frac{\sqrt{\lambda_{max}(A^T A)}}{\sqrt{\lambda_{min}(A^T A)}} = \frac{\sigma_{max}(A)}{\sigma_{min}(A)}$$



拉格朗日乘数法(选讲)

对于一般的矩阵 A , $\max_{\|x\|_2=1} \|Ax\|_2$ 什么时候取到极值?

除了通过特征值的min-max刻画, 还可以通过拉格朗日乘数法, 把带约束的最优化问题转化为不带约束的:

$$L(\lambda) := \sup_x x^T A^T A x - \lambda(x^T x - 1)$$

对 x 求梯度也可得 $A^T A x = \lambda x$

因此也可得到 $\max_{\|x\|_2=1} \|Ax\|_2 = \sqrt{\lambda_{max}(A^T A)}$



SVD与low rank factorization

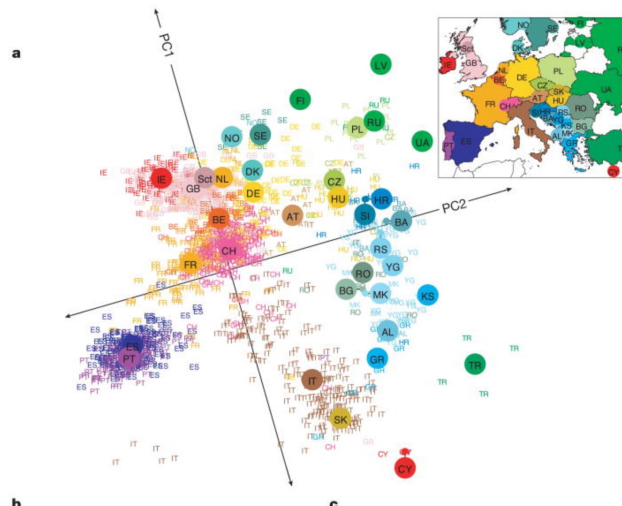
$$A = \sum_{i=1}^{\min(m,n)} \sigma_i \mathbf{u}_i \mathbf{v}_i^T$$

考虑如下的近似 $\tilde{A} := \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T$

Eckart-Young定理: \tilde{A} 是所有列空间的维度最多为 r 的矩阵中, 能同时最小化 $\|\tilde{A} - A\|_2$ 和 $\|\tilde{A} - A\|_F$ 的矩阵

Many applications:

- Principal component analysis
 - Data visualization
 - [Genomes can encode geography](#)
 - Eigenfaces
- Recommender systems
- Latent semantic analysis





SVD与伪逆pseudoinverse (选讲)

$$A = \sum_{i=1}^{\min(m,n)} \sigma_i u_i v_i^T$$

$$A^+ = VS^+U^T = \sum_{i:\sigma_i \neq 0} \frac{1}{\sigma_i} v_i u_i^T$$

如果 A 可逆, 则 $A^+ = A^{-1}$

如果 A 是overdetermined, 则 A^+b 为 $Ax = b$ 的最小二乘解

如果 A 是underdetermined, 则 A^+b 为 $Ax = b$ 中2范数最小的最小二乘解



Random walk on graphs

给定图 $G = (V, E)$

图上的随机游走:

- 从一个给定的顶点出发
- 接下来, 每一步都从当前顶点, 移动到一个均匀随机选取的邻居
- 不断重复

这个随机过程的“长期表现”是怎么样的呢?

1. 重复 t 步之后, 当前顶点是某个顶点 u 的概率是多少?
2. 是否存在一个极限的随机分布, 随机游走会收敛到它? (稳态)
3. 多久才会收敛? (混合时间, mixing time)
4. 从点 s 出发, 多久才会到达点 t ? (hitting time)
5. 多久才会遍历每个顶点至少一次? (遍历时间, cover time)



Random walk on graphs

这个随机过程的“长期表现”是怎么样的呢？

1. 重复 t 步之后，当前顶点是某个顶点 u 的概率是多少？
2. 是否存在一个极限的随机分布，随机游走会收敛到它？（稳态）
3. 多久才会收敛？(混合时间，mixing time)
4. 从点 s 出发，多久才会到达点 t ? (hitting time)
5. 多久才会遍历每个顶点至少一次？(遍历时间，cover time)

前3个问题可以通过纯概率的方法(coupling)来研究；我们将讨论如何利用与幂迭代的联系，通过马尔可夫链(Markov chains)转移矩阵的特征值来研究。

后2个问题我们将留到之后讨论“图与电阻电路网络”的时候



幂迭代与无向图上的随机游走

令 A 为无向图的邻接矩阵, D 为度数的对角矩阵. (注: 因为是无向图, 所以 A 是对称的)

记 $p_t(v)$ 为时间 t 在点 v 的概率, 则有

$$p_{t+1}(v) = \sum_{u:uv \in E} p_t(u) \cdot \frac{1}{\deg(u)}$$

可以写出矩阵的形式: $\vec{p}_{t+1} = \vec{p}_t(D^{-1}A)$, 进而有 $\vec{p}_t = \vec{p}_0(D^{-1}A)^t$; \vec{p}_t 为行向量

转移矩阵 $P := D^{-1}A$

要研究 P^t , 只需要研究 $W = D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ 的幂迭代, 因为:

$$W = D^{\frac{1}{2}} P D^{-\frac{1}{2}}$$

对于实数对称矩阵 $W = VDV^T$ 的幂迭代:

$$W^k = VD^kV^T = \sum_{1 \leq i \leq n} \lambda_i^k v_i v_i^T$$

W 的特征值是怎样分布的? 与邻接矩阵的谱相关



图的谱

在传统的算法课上，无向图的邻接矩阵往往只是作为一个数据结构

在接下来的几节课里面，会把邻接矩阵作为矩阵研究：特征值，行列式，秩等等，并把这些代数性质与大家习惯的组合性质相联系

这节课，首先给大家介绍随机游走与马尔可夫链的基本概念



Markov chains

考虑一般的有向图上的随机游走，记 X_t 为时间 t 随机游走所处的状态
转移矩阵 $P_{i,j} = \Pr[X_{t+1} = j | X_t = i]$

记 $p_t(i)$ 为时间 t 在状态 i 的概率. 那么, 对于所有 $0 \leq j \leq n - 1$,

- $p_{t+1}(j) = \sum_{i=0}^{n-1} p_t(i) \cdot P_{i,j}$
- 更一般地, $\overrightarrow{p_{t+1}} = \overrightarrow{p_t} \cdot P$, 并且 $\overrightarrow{p_t} = \overrightarrow{p_0} \cdot P^t$.

注意: 转移只与时间 t 在各个状态的概率相关, 与过去无关



Irreducible Markov chains

对于有限的马尔可夫链，如果它对应的有向图是强连通的，则称它为不可约的(irreducible)。



Aperiodic Markov chains

给定一个马尔可夫链，其中状态 i 的周期定义为

$$period(i) := \gcd\{t \mid P_{i,i}^t > 0\}$$

- 如果一个状态 i 的周期 $period(i) = 1$ ，则称它为非周期性的(aperiodic)
- 如果马尔可夫链中每一个状态都是非周期性的话，则称该马尔可夫链是非周期性的；否则称它是周期性的。

例子：二分图，3-cycle，self-loops



Irreducible & aperiodic Markov chains

引理：如果一个马尔可夫链是不可约，且非周期的话，则一定存在足够大的常数 T ，使得 $(P^t)_{i,j} > 0 \forall i, j, \forall t \geq T$ 。

证明思路：首先证明 $(P^t)_{i,i} > 0$ ，之后使用不可约的性质。



稳态分布 (stationary distribution)

如果分布 $\vec{\pi}$ 满足 $\vec{\pi} = \vec{\pi} \cdot P$, 则称 $\vec{\pi}$ 为稳态分布

稳态分布是一个“平衡态” / 不动点

特别地: $\vec{\pi} = \vec{\pi} \cdot P^t, \forall t$

注意: 如果极限分布存在, 则一定是一个稳态分布



概率分布的距离度量与收敛

要研究收敛，需要一个距离的度量；这里我们使用的是TV距离(total variation distance)。给定概率分布 \vec{p} 与 \vec{q} ，它们的TV距离定义为

$$d_{TV}(\vec{p}, \vec{q}) := \frac{1}{2} \|p - q\|_1 = \frac{1}{2} \sum_{i=1}^n |p_i - q_i|$$

注意 $0 \leq d_{TV}(\vec{p}, \vec{q}) \leq 1$

如果 $\lim_{t \rightarrow \infty} d_{TV}(\vec{p}_t, \vec{q}) = 0$ ，则说 $\{\vec{p}_t\}$ 收敛到 \vec{q}

(选讲) 也可以选用2范数

$$\left\| \frac{p}{\pi} - 1 \right\|_{2,\pi}^2 := \sum_{i=1}^n \pi_i \left(\frac{p_i}{\pi_i} - 1 \right)^2$$

由Cauchy-Schwarz不等式,

$$\sum_{i=1}^n |p_i - \pi_i| = \sum_{i=1}^n \pi_i \left| \frac{p_i}{\pi_i} - 1 \right| \leq \sqrt{\sum_{i=1}^n \pi_i \left| \frac{p_i}{\pi_i} - 1 \right|^2} = \left\| \frac{p}{\pi} - 1 \right\|_{2,\pi}$$

这意味着，如果2-norm 是小的，则1-norm 也是小的(但反过来不一定)。



回归时间 (Return time)

从状态 i 出发，第一次返回到状态 i 的时间，称之为状态 i 的回归时间，定义为

$$H_i := \min\{t \geq 1 \mid X_t = i, X_0 = i\}$$

期望回归时间定义为 $h_i := E[H_i]$.



Fundamental Theorem of Markov Chains

马尔可夫链的基本定理

对于有限，不可约，非周期的马尔可夫链，有以下事实：

1. 存在一个稳态分布 $\vec{\pi}$.
2. 随着 $t \rightarrow \infty$, \vec{p}_t 都会收敛到 $\vec{\pi}$, 无论从什么样的 \vec{p}_0 开始.
3. 稳态分布是唯一的
4. $\pi(i) = \frac{1}{h_i}$



Intuition

马尔可夫链的基本定理的一些直观理解

对于有限, 不可约, 非周期的马尔可夫链, 有以下事实:

1. 存在一个稳态分布 $\vec{\pi}$.
2. 随着 $t \rightarrow \infty$, \vec{p}_t 都会收敛到 $\vec{\pi}$, 无论从什么样的 \vec{p}_0 开始.
3. 稳态分布是唯一的
4. $\pi(i) = \frac{1}{h_i}$

考虑两个随机游走, 当它们相遇之后, 就变得“无法区别”了

回忆之前的引理: 存在足够大的常数 T , 使得 $(P^t)_{i,j} > 0 \forall i, j, \forall t \geq T$

这意味着无论它们现在在何方, 经过一段时间之后, 都有非零的概率相遇

最终会相遇的概率是1, 此时将收敛到同一个稳态分布

通过概率的coupling argument, 可以把上述想法变成严格的证明

$\pi(i) = \frac{1}{h_i}$: 期望回归时间为 h_i



Pagerank

考虑一个描述网页超链接的有向图

点集：网页

边集：如果网页 i 到网页 j 存在超链接，则 $i \rightarrow j$

搜索引擎想要给网页一个关于其内容质量/重要程度的打分

Pagerank的想法：

- 如果一个页面被很多人转载/引用，它是比较重要的；衡量标准：存在超链接指向它；
- 如果一个页面被其它重要的页面所指向，它本身可能也是比较重要的；



Pagerank

一个迭代算法:

- 一开始每个页面都有pagerank分数为 $1/n$
- 接下来的**每一步**，每个页面都会把自己的分数平均分开，并发送给它所指向的邻居
- 每个页面会更新自己的分数为，它所收到的分数的和
- 不断重复直到收敛



Pagerank

Pagerank分数收敛到的平衡态，正是随机游走的稳态分布。

$$Pagerank_{t+1}(j) = \sum_{i:i \rightarrow j \in E} Pagerank_t(i) / \deg_{out}(i)$$

$$\text{令 } P_{i,j} = \begin{cases} \frac{1}{\deg_{out}(i)}, & \text{if } ij \in E \\ 0, & \text{otherwise} \end{cases}, \text{ 则有 } \overrightarrow{Pagerank_{t+1}} = \overrightarrow{Pagerank_t} \cdot P$$

Pagerank满足与随机游走相同的迭代关系

- 当图是有限，不可约（强连通），非周期的时候，Pagerank分数是唯一的。（马尔可夫链的基本定理）
- Pagerank分数只与图的结构有关，与初始值无关。
- Pagerank分数将会是随机游走期望回归时间的倒数。



Pagerank

实践中的图不一定是不可约（强连通），非周期的。

可以作以下修改：固定 ϵ ，把 $1 - \epsilon$ 的pagerank分数分给邻居， ϵ 的pagerank分数均匀分给所有点

对应的随机游走则是：

以概率 $1 - \epsilon$ ，走到一个均匀随机的邻居；

以概率 ϵ ，走到一个均匀随机的顶点。

可以证明，这个修改将保证图是强连通且非周期的而且不会改变网页的相对分数