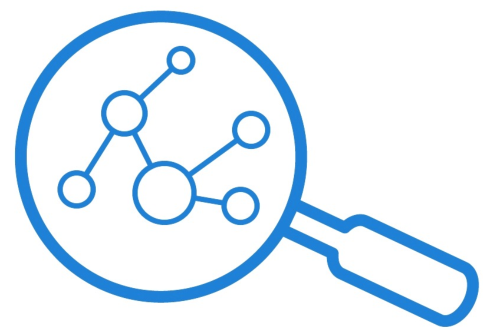
数据科学与大数据技术的数学基础

第八讲



计算机学院
余皓然

2024/5/13

课程内容

Part1 随机化方法
一致性哈希 布隆过滤器 CM Sketch方法 最小哈希 欧氏距离下的相似搜索 Jaccard相似度下的相似搜索
Part2 谱分析方法
主成分分析 奇异值分解 谱图论
Part3 最优化方法
压缩感知

主成分分析

主成分的代数性质

代数分析

主成分分析（Principal Component Analysis, PCA）

给定$\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m \in \mathbb{R}^n$及维数$k \geq 1$，求标准正交向量组$\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k \in \mathbb{R}^n$从而最大化

$$\frac{1}{m}\sum_{i=1}^{m}\left(\langle \boldsymbol{x}_i, \boldsymbol{v}_1\rangle^2 + \cdots + \langle \boldsymbol{x}_i, \boldsymbol{v}_k\rangle^2\right).$$

如何进一步用线性代数工具改写问题（如写成矩阵/向量相乘形式）？

代数分析

主成分分析（Principal Component Analysis, PCA）

给定$\boldsymbol{x}_1,\ldots,\boldsymbol{x}_m\in\mathbb{R}^n$及维数$k\geq 1$，求标准正交向量组$\boldsymbol{v}_1,\ldots,\boldsymbol{v}_k\in\mathbb{R}^n$从而最大化

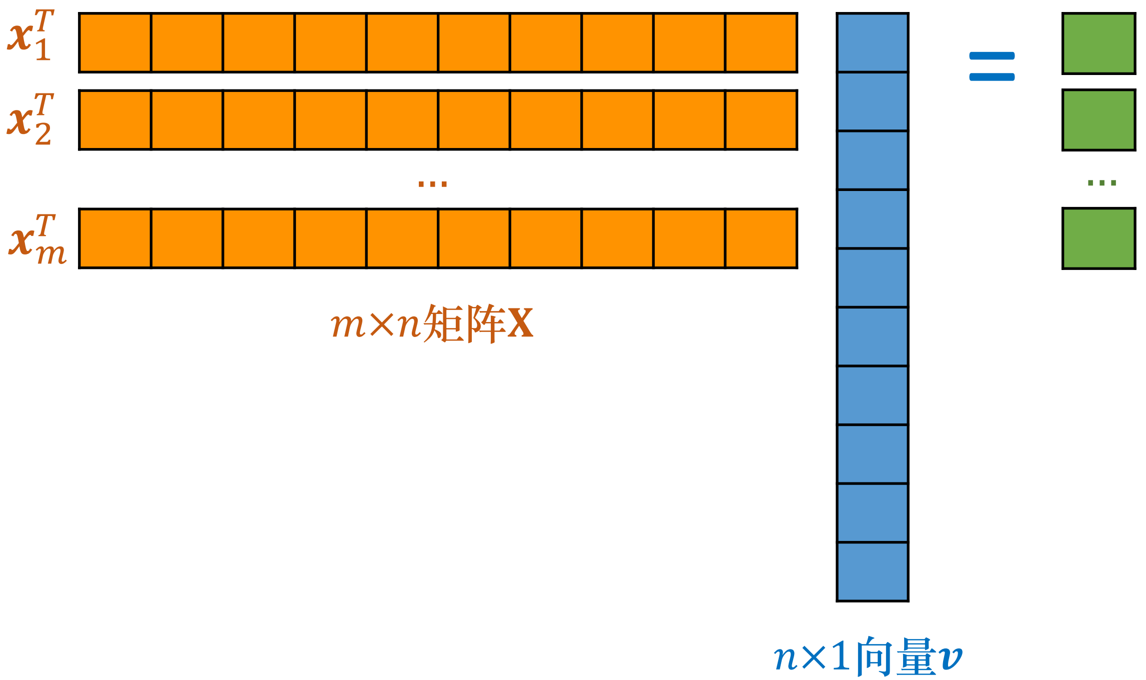
$$\frac{1}{m}\sum_{i=1}^{m}\left(\langle\boldsymbol{x}_i,\boldsymbol{v}_1\rangle^2+\cdots+\langle\boldsymbol{x}_i,\boldsymbol{v}_k\rangle^2\right).$$

如何进一步用线性代数工具改写问题（如写成矩阵/向量相乘形式）？

先分析$k=1$的情况（科研中常用思路：先分析简化问题、寻找规则，再求拓展）

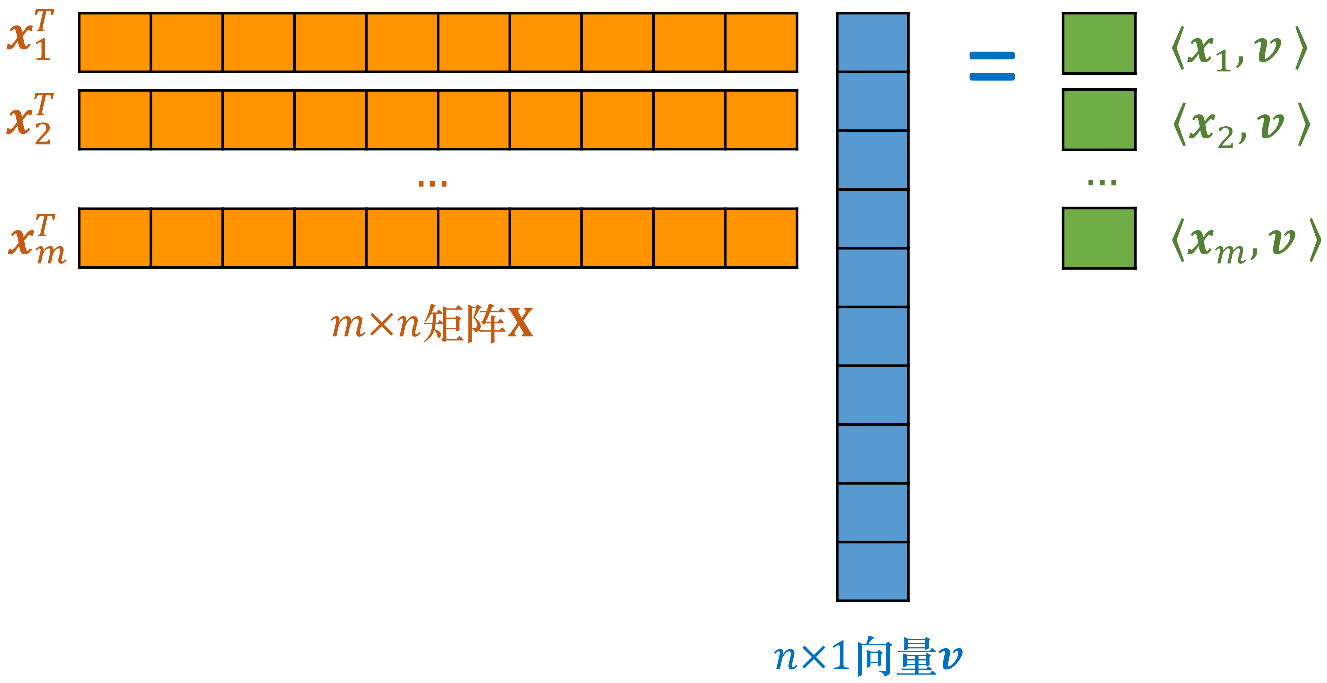
$k = 1$情况

给定$\boldsymbol{x}_1, \dots, \boldsymbol{x}_m \in \mathbb{R}^n$，求**单位向量**$\boldsymbol{v} \in \mathbb{R}^n$从而最大化$\frac{1}{m}\sum_{i=1}^{m}\langle \boldsymbol{x}_i, \boldsymbol{v} \rangle^2$.



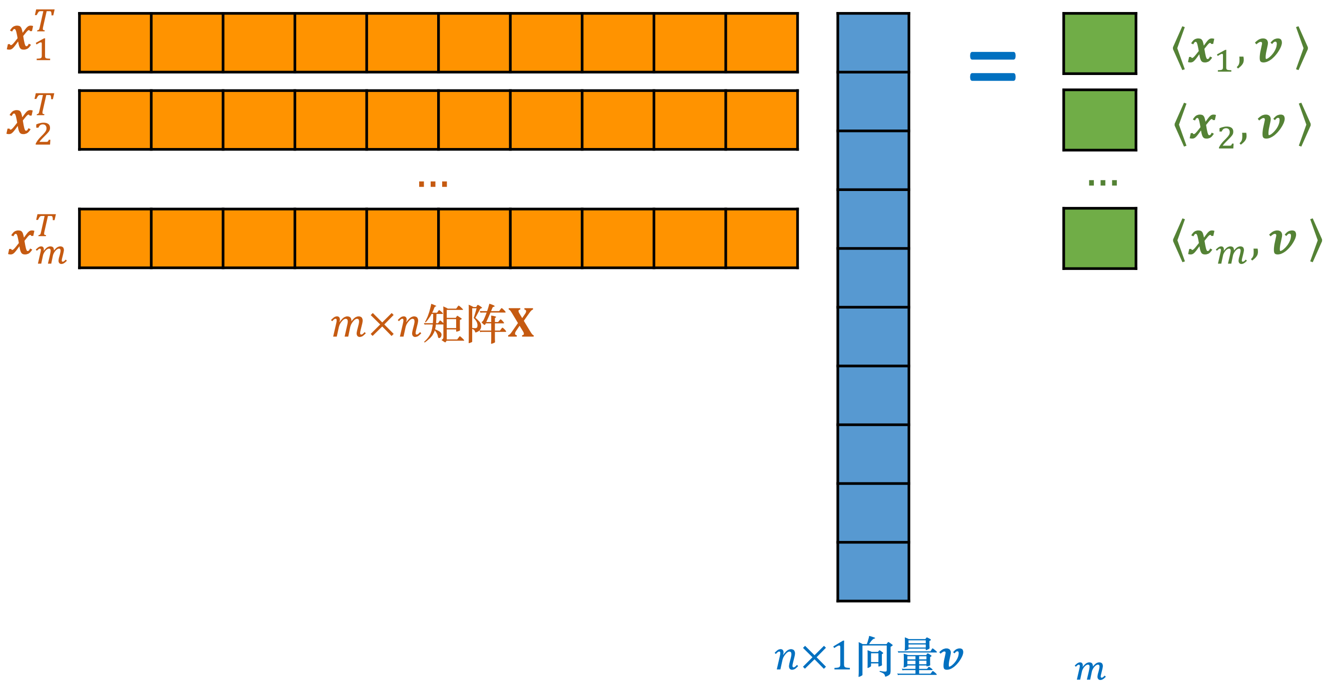
$k = 1$情况

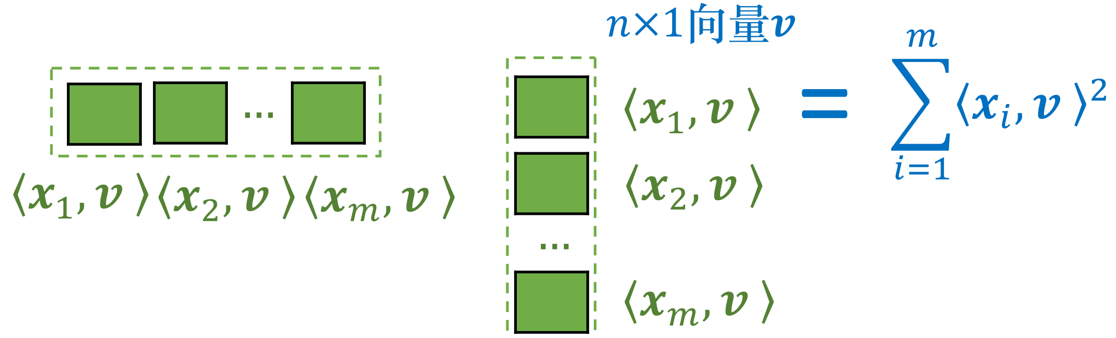
给定$\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m \in \mathbb{R}^n$，求**单位向量**$\boldsymbol{v} \in \mathbb{R}^n$从而最大化$\frac{1}{m}\sum_{i=1}^{m}\langle \boldsymbol{x}_i, \boldsymbol{v} \rangle^2$.



$k=1$情况

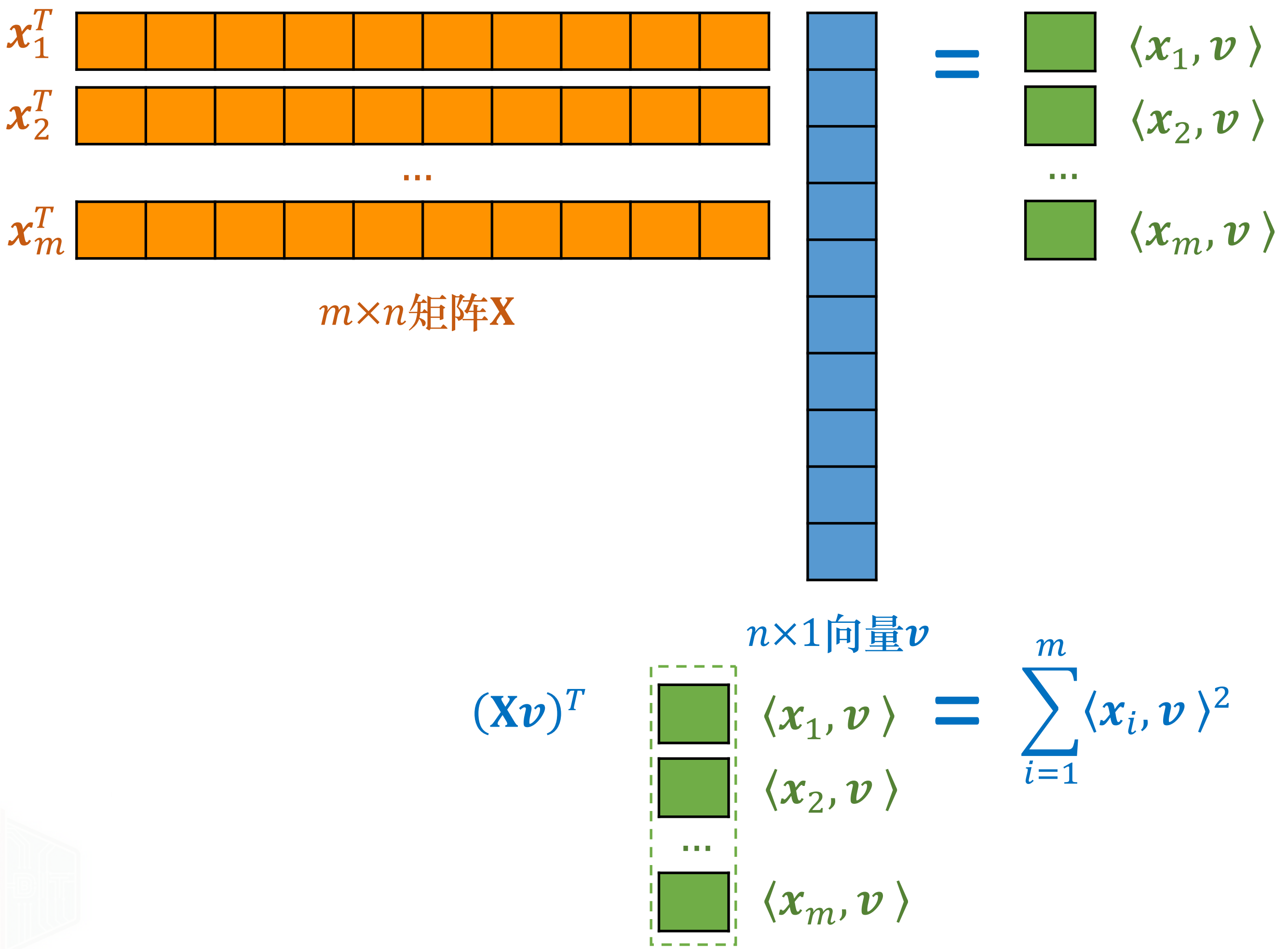
给定$\boldsymbol{x}_1,\dots,\boldsymbol{x}_m\in\mathbb{R}^n$，求**单位向量**$\boldsymbol{v}\in\mathbb{R}^n$从而最大化$\frac{1}{m}\sum_{i=1}^{m}\langle\boldsymbol{x}_i,\boldsymbol{v}\rangle^2$.





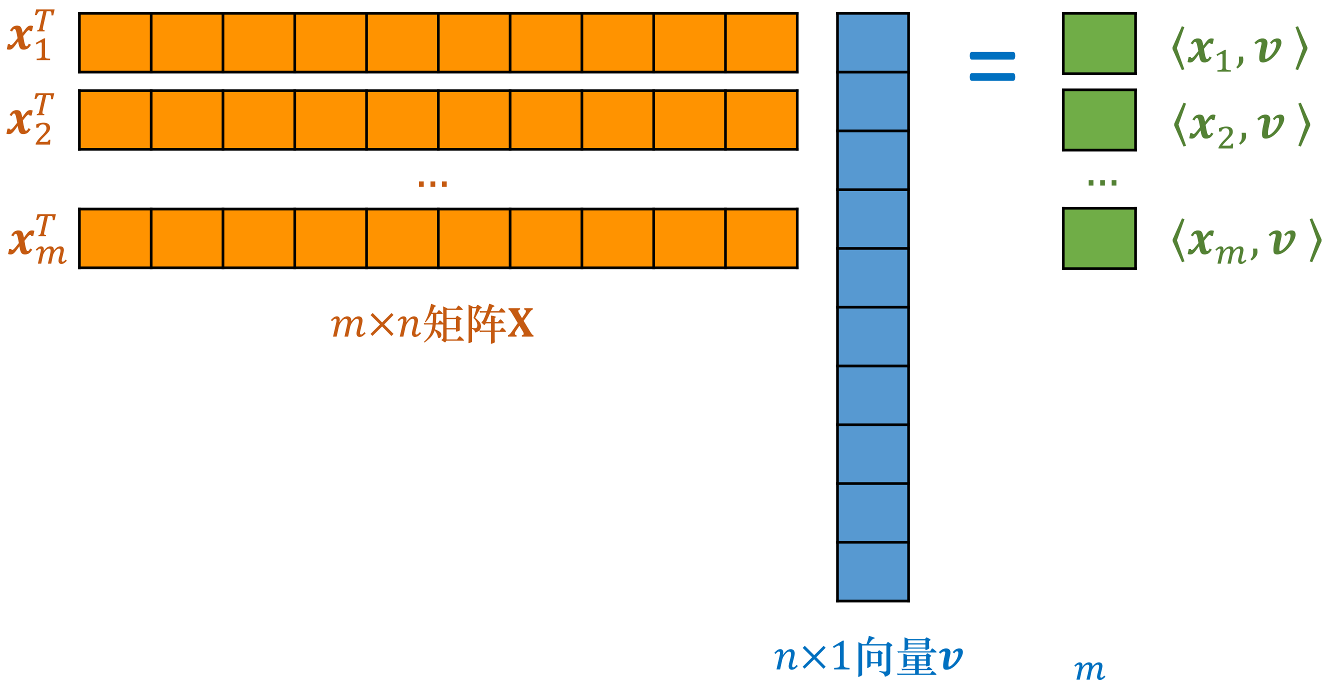
$k = 1$情况

给定$\boldsymbol{x}_1, \dots, \boldsymbol{x}_m \in \mathbb{R}^n$，求单位向量$\boldsymbol{v} \in \mathbb{R}^n$从而最大化$\frac{1}{m}\sum_{i=1}^{m}\langle \boldsymbol{x}_i, \boldsymbol{v} \rangle^2$.



$k=1$情况

给定$\boldsymbol{x}_1,\dots,\boldsymbol{x}_m\in\mathbb{R}^n$，求单位向量$\boldsymbol{v}\in\mathbb{R}^n$从而最大化$\frac{1}{m}\sum_{i=1}^{m}\langle\boldsymbol{x}_i,\boldsymbol{v}\rangle^2$.



$$(\mathbf{X}\boldsymbol{v})^T(\mathbf{X}\boldsymbol{v})=\sum_{i=1}^{m}\langle\boldsymbol{x}_i,\boldsymbol{v}\rangle^2$$

$k=1$情况

给定$\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m \in \mathbb{R}^n$，求单位向量$\boldsymbol{v} \in \mathbb{R}^n$从而最大化$\boldsymbol{v}^T\mathbf{X}^T\mathbf{X}\boldsymbol{v}$.

略去系数$\frac{1}{m}$

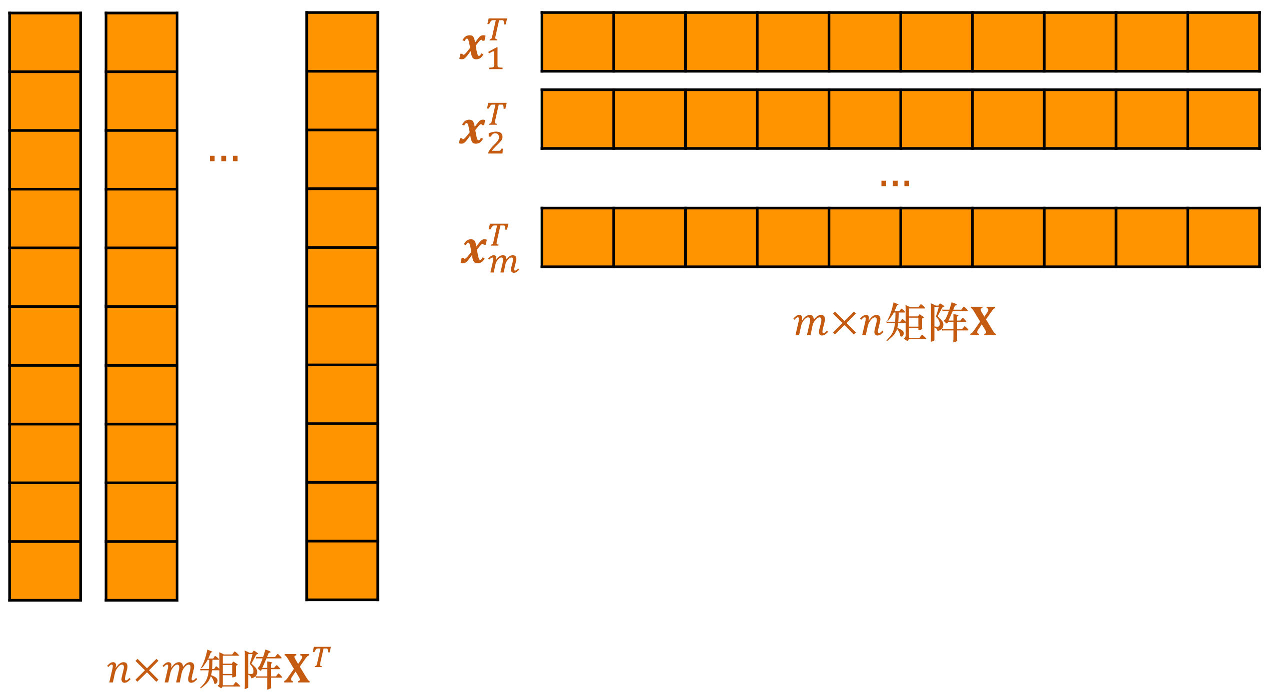
分析$n \times n$矩阵$\mathbf{X}^T\mathbf{X}$

是对称矩阵，因为$(\mathbf{X}^T\mathbf{X})^T = \mathbf{X}^T\mathbf{X}$

$k = 1$情况

给定$\boldsymbol{x}_1, \dots, \boldsymbol{x}_m \in \mathbb{R}^n$，求**单位向量**$\boldsymbol{v} \in \mathbb{R}^n$从而最大化$\boldsymbol{v}^T\mathbf{X}^T\mathbf{X}\boldsymbol{v}$.

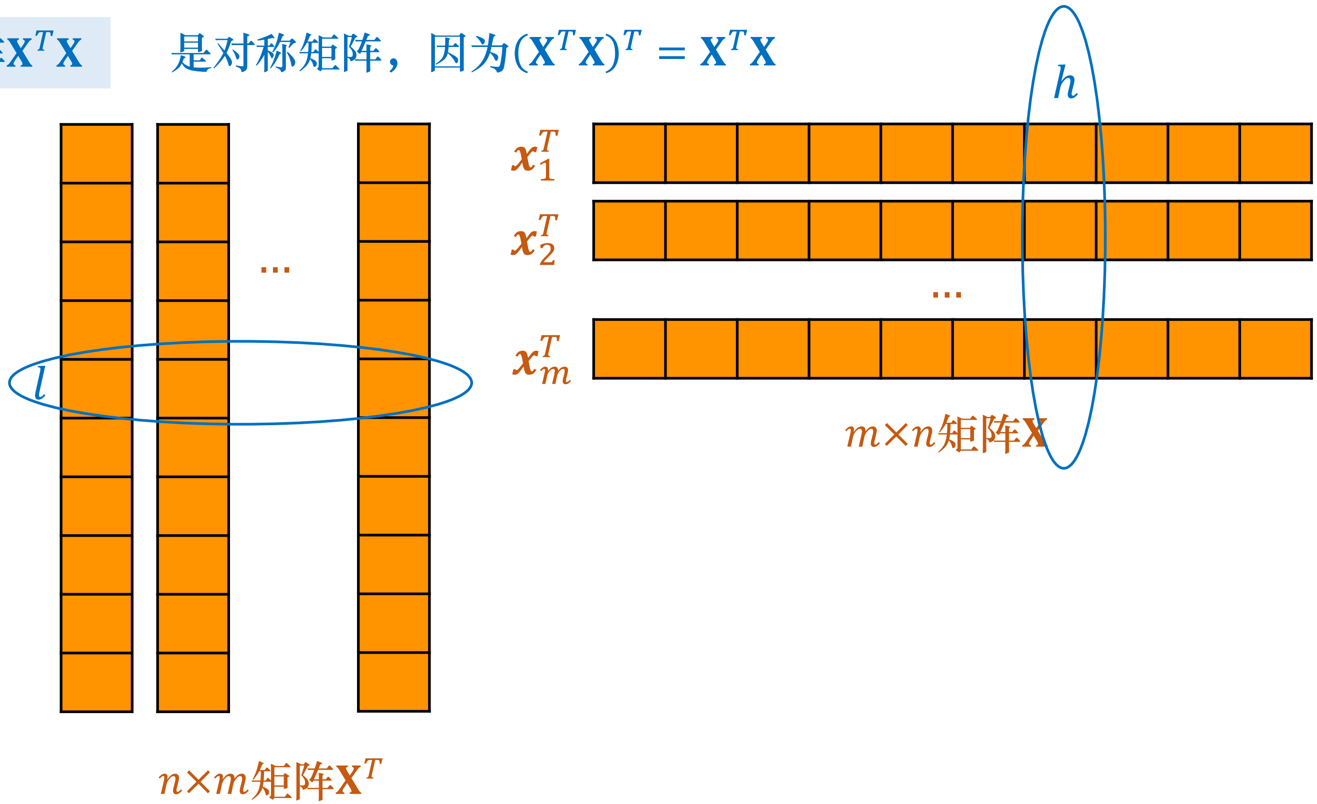
分析$n \times n$矩阵$\mathbf{X}^T\mathbf{X}$　是对称矩阵，因为$(\mathbf{X}^T\mathbf{X})^T = \mathbf{X}^T\mathbf{X}$



$k=1$情况

给定$\boldsymbol{x}_1,\ldots,\boldsymbol{x}_m\in\mathbb{R}^n$，求单位向量$\boldsymbol{v}\in\mathbb{R}^n$从而最大化$\boldsymbol{v}^T\mathbf{X}^T\mathbf{X}\boldsymbol{v}$.

分析$n\times n$矩阵$\mathbf{X}^T\mathbf{X}$　是对称矩阵，因为$(\mathbf{X}^T\mathbf{X})^T=\mathbf{X}^T\mathbf{X}$



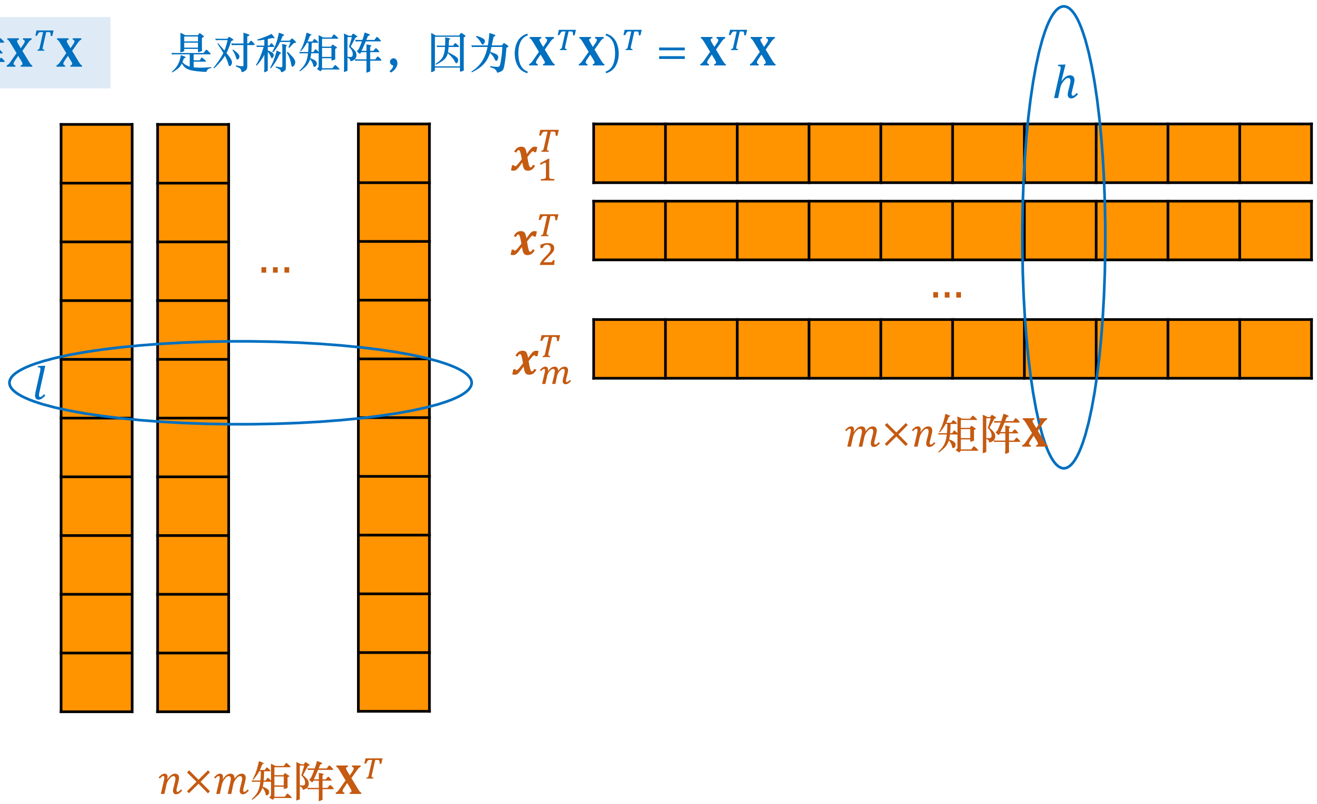
假设每个数据$\boldsymbol{x}_i$对应一个文件i而且$\boldsymbol{x}_{ij}$表示编号为j的词出现情况

矩阵$\mathbf{X}^T\mathbf{X}$的(l,h)元素的含义?

$k = 1$情况

给定$\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m \in \mathbb{R}^n$，求**单位向量**$\boldsymbol{v} \in \mathbb{R}^n$从而最大化$\boldsymbol{v}^T\mathbf{X}^T\mathbf{X}\boldsymbol{v}$.

分析$n{\times}n$矩阵$\mathbf{X}^T\mathbf{X}$　是对称矩阵，因为$(\mathbf{X}^T\mathbf{X})^T = \mathbf{X}^T\mathbf{X}$



假设每个数据$\boldsymbol{x}_i$对应一个文件i而且$\boldsymbol{x}_{ij}$表示编号为j的词出现情况

矩阵$\mathbf{X}^T\mathbf{X}$的(l,h)元素的含义：有多少次编号为l的词与编号为h的词同时在一个文件中出现过

$k = 1$情况

给定$\boldsymbol{x}_1, \dots, \boldsymbol{x}_m \in \mathbb{R}^n$，求单位向量$\boldsymbol{v} \in \mathbb{R}^n$从而最大化$\boldsymbol{v}^T\mathbf{X}^T\mathbf{X}\boldsymbol{v}$.

将$n{\times}n$矩阵$\mathbf{X}^T\mathbf{X}$称为（预处理后的）向量$\boldsymbol{x}_1, \dots, \boldsymbol{x}_m$的协方差矩阵

为书写方便，记矩阵$\mathbf{X}^T\mathbf{X}$为矩阵$\mathbf{A}$

先分析如何在特殊情况选择单位向量$\boldsymbol{v}$最大化$\boldsymbol{v}^T\mathbf{A}\boldsymbol{v}$

$k = 1$情况：A满足特殊条件

$$\max_{\|\boldsymbol{v}\|=1} \boldsymbol{v}^T \mathbf{A} \boldsymbol{v}$$

假设**A**是对角矩阵 $\begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}$ 且满足 $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq 0$

$k=1$情况：A满足特殊条件

$\max_{\|\boldsymbol{v}\|=1} \boldsymbol{v}^T \mathbf{A} \boldsymbol{v}$ 假设A是对角矩阵 $\begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}$ 且满足 $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq 0$

易得$\boldsymbol{v}^T \mathbf{A} \boldsymbol{v} = (v_1\lambda_1, v_2\lambda_2, \dots, v_n\lambda_n) \begin{pmatrix} v_1 \\ v_2 \\ \dots \\ v_n \end{pmatrix} = \sum_{i=1}^n v_i^2 \lambda_i$

问题简化为：

$$\max \sum_{i=1}^n v_i^2 \lambda_i$$

$$\text{s.t.} \sum_{i=1}^n v_i^2 = 1$$

$k=1$情况：A满足特殊条件

$\max_{\|\boldsymbol{v}\|=1} \boldsymbol{v}^T \mathbf{A} \boldsymbol{v}$

假设A是对角矩阵 $\begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}$ 且满足 $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq 0$

易得$\boldsymbol{v}^T \mathbf{A} \boldsymbol{v} = (v_1\lambda_1, v_2\lambda_2, \dots, v_n\lambda_n) \begin{pmatrix} v_1 \\ v_2 \\ \dots \\ v_n \end{pmatrix} = \sum_{i=1}^n v_i^2 \lambda_i$

问题简化为：

$$\max \sum_{i=1}^n v_i^2 \lambda_i$$
$$\text{s.t.} \sum_{i=1}^n v_i^2 = 1$$

解为$\boldsymbol{v}^* = (1,0,\dots,0)$或$(-1,0,\dots,0)$

（通过将v_i^2换成非负变量易得）

下一步：拓展到更一般化的A

$k=1$情况：A满足特殊条件

$$\max_{\|\boldsymbol{v}\|=1} \boldsymbol{v}^T \mathbf{A} \boldsymbol{v}$$

假设A可写成$\mathbf{QDQ}^T$，其中Q是正交矩阵、D是对角矩阵 $\begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}$ 且满足 $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq 0$

例：

$$\underset{\mathbf{A}}{\begin{bmatrix} \frac{3}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{3}{2} \end{bmatrix}} = \underset{\mathbf{Q}}{\begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}} \underset{\mathbf{D}}{\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}} \underset{\mathbf{Q}^T}{\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}}$$

正交矩阵即所有列向量构成标准正交向量组（列向量模都为1且彼此内积为0）的（实数）方阵

$k=1$情况：A满足特殊条件

$$\max_{\|v\|=1} v^T A v$$

假设A可写成$\mathbf{QDQ}^T$，其中Q是正交矩阵、D是对角矩阵$\begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}$且满足 $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq 0$

例：

$$\underset{\mathbf{A}}{\begin{bmatrix} \frac{3}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{3}{2} \end{bmatrix}} = \underset{\mathbf{Q}}{\begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}} \underset{\mathbf{D}}{\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}} \underset{\mathbf{Q}^T}{\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}}$$

$$\begin{aligned} I = BB^{-1} &= BIB^{-1} \\ &= B(AB)B^{-1} && \text{since } AB = I \\ &= BAI = BA. \end{aligned}$$

正交矩阵即所有列向量构成**标准正交向量组**（列向量模都为1且彼此内积为0）的（实数）**方阵**

- 有$\mathbf{Q}^T\mathbf{Q} = \mathbf{I}$（展开矩阵、利用标准正交向量定义可得）
- 有$\mathbf{QQ}^T = \mathbf{I}$（由$\mathbf{Q}^T\mathbf{Q} = \mathbf{I}$及逆矩阵性质可得）
- $\det(\mathbf{Q}) = 1 \text{or} -1$（利用$\det(\mathbf{Q}^T)\det(\mathbf{Q}) = \det(\mathbf{Q})^2 = \det(\mathbf{I}) = 1$得）
- Q的行向量也构成标准正交向量组，即**$\mathbf{Q}^T$也是正交矩阵**（由$\mathbf{QQ}^T = \mathbf{I}$及标准正交向量组定义得）
- 对任意$n\times 1$向量$\boldsymbol{v}$有$\|\mathbf{Q}\boldsymbol{v}\|^2 = \|\boldsymbol{v}\|^2$（由$\|\mathbf{Q}\boldsymbol{v}\|^2 = (\mathbf{Q}\boldsymbol{v})^T\mathbf{Q}\boldsymbol{v}$以及$\mathbf{Q}^T\mathbf{Q} = \mathbf{I}$可得）

$k=1$情况：A满足特殊条件

$\max_{\|v\|=1} \boldsymbol{v}^T \mathbf{A} \boldsymbol{v}$

假设A可写成$\mathbf{QDQ}^T$，其中Q是正交矩阵、D是对角矩阵$\begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}$且满足$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$

例：

$$\underset{\mathbf{A}}{\begin{bmatrix} \frac{3}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{3}{2} \end{bmatrix}} = \underset{\mathbf{Q}}{\begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}} \underset{\mathbf{D}}{\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}} \underset{\mathbf{Q}^T}{\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}}$$

正交矩阵即所有列向量构成**标准正交向量组**（列向量模都为1且彼此内积为0）的（实数）**方阵**

- 有$\mathbf{Q}^T\mathbf{Q} = \mathbf{I}$（展开矩阵、利用标准正交向量定义可得）
- 有$\mathbf{Q}\mathbf{Q}^T = \mathbf{I}$（由$\mathbf{Q}^T\mathbf{Q} = \mathbf{I}$及逆矩阵性质可得）
- $\det(\mathbf{Q}) = 1 \text{or} -1$（利用$\det(\mathbf{Q}^T)\det(\mathbf{Q}) = \det(\mathbf{Q})^2 = \det(\mathbf{I}) = 1$得）
- Q的行向量也构成标准正交向量组，即**$\mathbf{Q}^T$也是正交矩阵**（由$\mathbf{Q}\mathbf{Q}^T = \mathbf{I}$及标准正交向量组定义得）
- 对任意$n\times1$向量$\boldsymbol{v}$有$\|\mathbf{Q}\boldsymbol{v}\|^2 = \|\boldsymbol{v}\|^2$（由$\|\mathbf{Q}\boldsymbol{v}\|^2 = (\mathbf{Q}\boldsymbol{v})^T\mathbf{Q}\boldsymbol{v}$以及$\mathbf{Q}^T\mathbf{Q} = \mathbf{I}$可得）

求解$\max_{\|v\|=1} \boldsymbol{v}^T \mathbf{QDQ}^T \boldsymbol{v}$

$k=1$情况：A满足特殊条件

$$\max_{\|v\|=1} v^T A v$$

假设A可写成QDQ^T，其中Q是正交矩阵、D是对角矩阵$\begin{bmatrix}\lambda_1 & 0 & \dots & 0\\ 0 & \lambda_2 & \dots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \dots & \lambda_n\end{bmatrix}$且满足$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq 0$

可将$v^T QDQ^T v$整理为$\sum_{i=1}^{n} \lambda_i (v^T q_i)^2$（通过将矩阵Q展成$[q_1, q_2, \ldots, q_n]$）

$k=1$情况：A满足特殊条件

$$\max_{\|\boldsymbol{v}\|=1} \boldsymbol{v}^T \mathbf{A} \boldsymbol{v}$$

假设A可写成$\mathbf{QDQ}^T$，其中Q是正交矩阵、D是对角矩阵 $\begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}$ 且满足 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$

可将$\boldsymbol{v}^T\mathbf{QDQ}^T\boldsymbol{v}$整理为$\sum_{i=1}^{n}\lambda_i(\boldsymbol{v}^T\boldsymbol{q}_i)^2$（通过将矩阵Q展成$[\boldsymbol{q}_1, \boldsymbol{q}_2, \dots, \boldsymbol{q}_n]$）

目标函数简化为：$$\max \sum_{i=1}^{n} \lambda_i(\boldsymbol{v}^T\boldsymbol{q}_i)^2$$

约束条件整理为：$$\text{s.t.} \sum_{i=1}^{n} (\boldsymbol{v}^T\boldsymbol{q}_i)^2 = \underline{\|\boldsymbol{v}^T\mathbf{Q}\|^2 = \|\boldsymbol{v}\|^2} = 1$$

$\mathbf{Q}^T$是正交矩阵时有$\|\mathbf{Q}^T\boldsymbol{v}\|^2 = \|\boldsymbol{v}\|^2$

$k=1$情况：A满足特殊条件

$$\max_{\|\boldsymbol{v}\|=1} \boldsymbol{v}^T \boldsymbol{A}\boldsymbol{v}$$

假设A可写成$\mathbf{QDQ}^T$，其中Q是正交矩阵、D是对角矩阵 $\begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}$ 且满足 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$

可将$\boldsymbol{v}^T\mathbf{QDQ}^T\boldsymbol{v}$整理为$\sum_{i=1}^{n}\lambda_i(\boldsymbol{v}^T\boldsymbol{q}_i)^2$（通过将矩阵Q展成$[\boldsymbol{q}_1, \boldsymbol{q}_2, \dots, \boldsymbol{q}_n]$）

目标函数简化为：$$\max \sum_{i=1}^{n}\lambda_i(\boldsymbol{v}^T\boldsymbol{q}_i)^2$$

约束条件整理为：$$\text{s.t.} \sum_{i=1}^{n}(\boldsymbol{v}^T\boldsymbol{q}_i)^2 = 1$$

前述A为对角矩阵时的结果

$$\max \sum_{i=1}^{n}\lambda_i v_i^2$$

$$\text{s.t.} \sum_{i=1}^{n} v_i^2 = 1$$

$\boldsymbol{v}^* = (1,0,\dots,0)$或$(-1,0,\dots,0)$

$k=1$情况：A满足特殊条件

$\max\limits_{\|\boldsymbol{v}\|=1} \boldsymbol{v}^T A \boldsymbol{v}$

假设A可写成QDQ^T，其中Q是正交矩阵、D是对角矩阵 $\begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}$ 且满足 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$

可将$\boldsymbol{v}^T QDQ^T \boldsymbol{v}$整理为$\sum_{i=1}^{n} \lambda_i (\boldsymbol{v}^T \boldsymbol{q}_i)^2$（通过将矩阵Q展成$[\boldsymbol{q}_1, \boldsymbol{q}_2, \dots, \boldsymbol{q}_n]$）

目标函数简化为：$\max \sum_{i=1}^{n} \lambda_i (\boldsymbol{v}^T \boldsymbol{q}_i)^2$

约束条件整理为：$\text{s.t.} \sum_{i=1}^{n} (\boldsymbol{v}^T \boldsymbol{q}_i)^2 = 1$

（最优解之一）

$\boldsymbol{v}^*$使$\boldsymbol{v}^T [\boldsymbol{q}_1, \boldsymbol{q}_2, \dots, \boldsymbol{q}_n] = (1,0,\dots,0)$

前述A为对角矩阵时的结果

$$\max \sum_{i=1}^{n} \lambda_i v_i^2$$

$$\text{s.t.} \sum_{i=1}^{n} v_i^2 = 1$$

$\boldsymbol{v}^* = (1,0,\dots,0)$或$(-1,0,\dots,0)$

$k=1$情况：A满足特殊条件

$$\max_{\|\boldsymbol{v}\|=1} \boldsymbol{v}^T A \boldsymbol{v}$$

假设A可写成QDQ^T，其中Q是正交矩阵、D是对角矩阵 $\begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}$ 且满足 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$

可将$\boldsymbol{v}^T QDQ^T \boldsymbol{v}$整理为$\sum_{i=1}^{n} \lambda_i (\boldsymbol{v}^T \boldsymbol{q}_i)^2$（通过将矩阵Q展成$[\boldsymbol{q}_1, \boldsymbol{q}_2, \dots, \boldsymbol{q}_n]$）

目标函数简化为：

$$\max \sum_{i=1}^{n} \lambda_i (\boldsymbol{v}^T \boldsymbol{q}_i)^2$$

约束条件整理为：

$$\text{s.t.} \sum_{i=1}^{n} (\boldsymbol{v}^T \boldsymbol{q}_i)^2 = 1$$

前述A为对角矩阵时的结果

$$\max \sum_{i=1}^{n} \lambda_i v_i^2$$

$$\text{s.t.} \sum_{i=1}^{n} v_i^2 = 1$$

$\boldsymbol{v}^* = (1,0,\dots,0)$或$(-1,0,\dots,0)$

(最优解之一)

$\boldsymbol{v}^*$使$\boldsymbol{v}^T [\boldsymbol{q}_1, \boldsymbol{q}_2, \dots, \boldsymbol{q}_n] = (1,0,\dots,0)$

解为$\boldsymbol{v}^* = \boldsymbol{q}_1$（矩阵Q的第一列）

如果所有的A都可写成QDQ^T的形式、Q和D满足前述条件，即可完成对$k=1$时的分析

$k=1$情况：A满足特殊条件

对$\mathbf{A}=\mathbf{X}^T\mathbf{X}$，是否总可写成$\mathbf{QDQ}^T$？$\mathbf{Q}$是正交矩阵、$\mathbf{D}$是$\begin{bmatrix}\lambda_1 & 0 & \dots & 0\\ 0 & \lambda_2 & \dots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \dots & \lambda_n\end{bmatrix}$ 且 $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq 0$

$k=1$情况：A满足特殊条件

对$\mathbf{A}=\mathbf{X}^T\mathbf{X}$，是否总可写成$\mathbf{QDQ}^T$？$\mathbf{Q}$是正交矩阵、$\mathbf{D}$是$\begin{bmatrix}\lambda_1 & 0 & \dots & 0\\ 0 & \lambda_2 & \dots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \dots & \lambda_n\end{bmatrix}$ 且 $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq 0$

$\mathbf{A}=\mathbf{X}^T\mathbf{X}$是对称矩阵

线性代数知识

实对称矩阵的正交对角化（Orthogonal Diagonalization of Real Symmetric Matrices）

若$\mathbf{A}$是实对称矩阵，存在正交矩阵$\mathbf{P}$和对角矩阵$\mathbf{D}$使$\mathbf{A}=\mathbf{PDP}^T$。

$$\begin{bmatrix} | & | & & | \\ \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \\ | & | & & | \end{bmatrix} \begin{bmatrix}\lambda_1 & 0 & \dots & 0\\ 0 & \lambda_2 & \dots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \dots & \lambda_n\end{bmatrix} \begin{bmatrix} - & \mathbf{q}_1^T & - \\ - & \mathbf{q}_2^T & - \\ & \vdots & \\ - & \mathbf{q}_n^T & - \end{bmatrix}$$

$k=1$情况：A满足特殊条件

对$\mathbf{A}=\mathbf{X}^T\mathbf{X}$，是否总可写成$\mathbf{QDQ}^T$？ Q是正交矩阵、D是$\begin{bmatrix}\lambda_1 & 0 & \dots & 0\\ 0 & \lambda_2 & \dots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \dots & \lambda_n\end{bmatrix}$ 且 $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq 0$

$\mathbf{A}=\mathbf{X}^T\mathbf{X}$是对称矩阵

线性代数知识

实对称矩阵的正交对角化（Orthogonal Diagonalization of Real Symmetric Matrices）

若A是实对称矩阵，存在正交矩阵P和对角矩阵D使$\mathbf{A}=\mathbf{PDP}^T$。

易由特征向量及特征值定义得（如证$\mathbf{A}\boldsymbol{p}_1=\mathbf{PDP}^T\boldsymbol{p}_1=\lambda_1\boldsymbol{p}_1$）：

P的列向量为A的特征向量、D的对角线元素为A的特征值

$k=1$情况：A满足特殊条件

对$\mathbf{A}=\mathbf{X}^T\mathbf{X}$，是否总可写成$\mathbf{Q}\mathbf{D}\mathbf{Q}^T$？Q是正交矩阵、D是$\begin{bmatrix}\lambda_1 & 0 & \dots & 0\\ 0 & \lambda_2 & \dots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \dots & \lambda_n\end{bmatrix}$且 $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq 0$

$\mathbf{A}=\mathbf{X}^T\mathbf{X}$是对称矩阵

线性代数知识

实对称矩阵的正交对角化（Orthogonal Diagonalization of Real Symmetric Matrices）

若A是实对称矩阵，存在正交矩阵P和对角矩阵D使$\mathbf{A}=\mathbf{P}\mathbf{D}\mathbf{P}^T$。

易由特征向量及特征值定义得（如证$\mathbf{A}\boldsymbol{p}_1=\mathbf{P}\mathbf{D}\mathbf{P}^T\boldsymbol{p}_1=\lambda_1\boldsymbol{p}_1$）：

P的列向量为A的特征向量、D的对角线元素为A的特征值

还需要对角矩阵D的元素全为非负且按大小排列

$k=1$情况：$\mathbf{A}$满足特殊条件

对$\mathbf{A}=\mathbf{X}^T\mathbf{X}$，是否总可写成$\mathbf{QDQ}^T$？$\mathbf{Q}$是正交矩阵、$\mathbf{D}$是$\begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}$ 且 $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq 0$

$\mathbf{A}=\mathbf{X}^T\mathbf{X}$是对称矩阵

线性代数知识

实对称矩阵的正交对角化（Orthogonal Diagonalization of Real Symmetric Matrices）

若$\mathbf{A}$是实对称矩阵，存在正交矩阵$\mathbf{P}$和对角矩阵$\mathbf{D}$使$\mathbf{A}=\mathbf{PDP}^T$。

易由特征向量及特征值定义得（如证$\mathbf{A}\boldsymbol{p}_1=\mathbf{PDP}^T\boldsymbol{p}_1=\lambda_1\boldsymbol{p}_1$）：

$\mathbf{P}$的列向量为$\mathbf{A}$的特征向量、$\mathbf{D}$的对角线元素为$\mathbf{A}$的特征值

还需要对角矩阵$\mathbf{D}$的元素全为非负且按大小排列

证明对称矩阵$\mathbf{X}^T\mathbf{X}$是半正定矩阵，即对任意向量$\boldsymbol{u}\in\mathbb{R}^n$有$\boldsymbol{u}^T\mathbf{X}^T\mathbf{X}\boldsymbol{u}\geq 0$。

再利用半正定矩阵的所有特征值非负的性质证明对角矩阵$\mathbf{D}$的元素全为非负。

$k=1$情况：A满足特殊条件

对$\mathbf{A}=\mathbf{X}^T\mathbf{X}$，是否总可写成$\mathbf{QDQ}^T$？Q是正交矩阵、D是$\begin{bmatrix}\lambda_1 & 0 & \dots & 0\\ 0 & \lambda_2 & \dots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \dots & \lambda_n\end{bmatrix}$ 且 $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq 0$

$\mathbf{A}=\mathbf{X}^T\mathbf{X}$是对称矩阵

线性代数知识

实对称矩阵的正交对角化（Orthogonal Diagonalization of Real Symmetric Matrices）

若A是实对称矩阵，存在正交矩阵P和对角矩阵D使$\mathbf{A}=\mathbf{PDP}^T$。

易由特征向量及特征值定义得（如证$\mathbf{A}\boldsymbol{p}_1=\mathbf{PDP}^T\boldsymbol{p}_1=\lambda_1\boldsymbol{p}_1$）：

P的列向量为A的特征向量、D的对角线元素为A的特征值

还需要对角矩阵D的元素全为非负且按大小排列

证明对称矩阵$\mathbf{X}^T\mathbf{X}$是半正定矩阵，即对任意向量$\boldsymbol{u}\in\mathbb{R}^n$有$\boldsymbol{u}^T\mathbf{X}^T\mathbf{X}\boldsymbol{u}\geq 0$。

也可取$\boldsymbol{u}=\boldsymbol{p}_i$（P的第$i$列向量）此时$\boldsymbol{u}^T\mathbf{X}^T\mathbf{X}\boldsymbol{u}=\boldsymbol{p}_i^T\lambda_i\boldsymbol{p}_i=\lambda_i$（利用标准正交向量性质或特征向量定义），说明对所有特征值有$\lambda_i\geq 0$。

$k=1$情况：A满足特殊条件

对$\mathbf{A}=\mathbf{X}^T\mathbf{X}$，是否总可写成$\mathbf{QDQ}^T$？$\mathbf{Q}$是正交矩阵、$\mathbf{D}$是$\begin{bmatrix}\lambda_1 & 0 & \dots & 0\\ 0 & \lambda_2 & \dots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \dots & \lambda_n\end{bmatrix}$且 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$

$\mathbf{A}=\mathbf{X}^T\mathbf{X}$是对称矩阵

线性代数知识

实对称矩阵的正交对角化（Orthogonal Diagonalization of Real Symmetric Matrices）

若$\mathbf{A}$是实对称矩阵，存在正交矩阵$\mathbf{P}$和对角矩阵$\mathbf{D}$使$\mathbf{A}=\mathbf{PDP}^T$。

易由特征向量及特征值定义得（如证$\mathbf{A}\boldsymbol{p}_1=\mathbf{PDP}^T\boldsymbol{p}_1=\lambda_1\boldsymbol{p}_1$）：

$\mathbf{P}$的列向量为$\mathbf{A}$的特征向量、$\mathbf{D}$的对角线元素为$\mathbf{A}$的特征值

还需要对角矩阵$\mathbf{D}$的元素按大小排列

$k=1$情况：A满足特殊条件

对$\mathbf{A}=\mathbf{X}^T\mathbf{X}$，是否总可写成$\mathbf{QDQ}^T$？$\mathbf{Q}$是正交矩阵、$\mathbf{D}$是$\begin{bmatrix}\lambda_1 & 0 & \dots & 0\\ 0 & \lambda_2 & \dots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \dots & \lambda_n\end{bmatrix}$ 且 $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq 0$

$\mathbf{A}=\mathbf{X}^T\mathbf{X}$是对称矩阵

线性代数知识

实对称矩阵的正交对角化（Orthogonal Diagonalization of Real Symmetric Matrices）

若A是实对称矩阵，存在正交矩阵P和对角矩阵D使$\mathbf{A}=\mathbf{PDP}^T$。

易由特征向量及特征值定义得（如证$\mathbf{A}\boldsymbol{p}_1=\mathbf{PDP}^T\boldsymbol{p}_1=\lambda_1\boldsymbol{p}_1$）：

P的列向量为A的特征向量、D的对角线元素为A的特征值

还需要对角矩阵D的元素按大小排列

同时调整特征值及相应特征向量的排序，结果不变

$$\begin{bmatrix} | & | & & | \\ \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \\ | & | & & | \end{bmatrix}\begin{bmatrix}\lambda_1 & 0 & \dots & 0\\ 0 & \lambda_2 & \dots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \dots & \lambda_n\end{bmatrix}\begin{bmatrix} \text{---} & \mathbf{q}_1^T & \text{---} \\ \text{---} & \mathbf{q}_2^T & \text{---} \\ & \vdots & \\ \text{---} & \mathbf{q}_n^T & \text{---} \end{bmatrix}$$

$k=1$情况：A满足特殊条件

对$\mathbf{A}=\mathbf{X}^T\mathbf{X}$，是否总可写成$\mathbf{QDQ}^T$？$\mathbf{Q}$是正交矩阵、$\mathbf{D}$是$\begin{bmatrix}\lambda_1 & 0 & \dots & 0\\ 0 & \lambda_2 & \dots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \dots & \lambda_n\end{bmatrix}$ 且 $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq 0$

$\mathbf{A}=\mathbf{X}^T\mathbf{X}$是对称矩阵

线性代数知识

实对称矩阵的正交对角化（Orthogonal Diagonalization of Real Symmetric Matrices）

若$\mathbf{A}$是实对称矩阵，存在正交矩阵$\mathbf{P}$和对角矩阵$\mathbf{D}$使$\mathbf{A}=\mathbf{PDP}^T$。

易由特征向量及特征值定义得（如证$\mathbf{A}\boldsymbol{p}_1=\mathbf{PDP}^T\boldsymbol{p}_1=\lambda_1\boldsymbol{p}_1$）：

$\mathbf{P}$的列向量为$\mathbf{A}$的特征向量、$\mathbf{D}$的对角线元素为$\mathbf{A}$的特征值

还需要对角矩阵$\mathbf{D}$的元素按大小排列

例如，$[\boldsymbol{p}_1,\boldsymbol{p}_2]\begin{bmatrix}\lambda_1 & 0\\ 0 & \lambda_2\end{bmatrix}\begin{bmatrix}\boldsymbol{p}_1^T\\ \boldsymbol{p}_2^T\end{bmatrix}=[\boldsymbol{p}_2,\boldsymbol{p}_1]\begin{bmatrix}\lambda_2 & 0\\ 0 & \lambda_1\end{bmatrix}\begin{bmatrix}\boldsymbol{p}_2^T\\ \boldsymbol{p}_1^T\end{bmatrix}$

回顾

$\max \sum_{i=1}^{n} \lambda_i (\boldsymbol{v}^T \boldsymbol{q}_i)^2 \quad \mathrm{s.t.} \sum_{i=1}^{n} (\boldsymbol{v}^T \boldsymbol{q}_i)^2 = 1$

最优解之一为$\boldsymbol{v}^* = \boldsymbol{q}_1$（矩阵$\mathbf{Q}$的第一列）

$k = 1$情况

主成分分析（Principal Component Analysis, PCA）

给定$\boldsymbol{x}_1, \dots, \boldsymbol{x}_m \in \mathbb{R}^n$及维数$k \geq 1$，求标准正交向量组$\boldsymbol{v}_1, \dots, \boldsymbol{v}_k \in \mathbb{R}^n$从而最大化

$$\frac{1}{m} \sum_{i=1}^{m} \left(\langle \boldsymbol{x}_i, \boldsymbol{v}_1 \rangle^2 + \cdots + \langle \boldsymbol{x}_i, \boldsymbol{v}_k \rangle^2 \right).$$

主成分分析问题的最优解

当$k = 1$，最大化目标函数的向量$\boldsymbol{v}^*$为矩阵$\mathbf{X}^T\mathbf{X}$的最大特征值对应的单位特征向量。

如果在对$\mathbf{X}^T\mathbf{X}$做正交对角化时将$\mathbf{X}^T\mathbf{X}$最大特征值排在对角矩阵的第一位，那么$\boldsymbol{v}^*$是正交矩阵的第一个列向量。

$k=1$情况

主成分分析（Principal Component Analysis, PCA）

给定$\boldsymbol{x}_1,\dots,\boldsymbol{x}_m\in\mathbb{R}^n$及维数$k\geq 1$，求标准正交向量组$\boldsymbol{v}_1,\dots,\boldsymbol{v}_k\in\mathbb{R}^n$从而最大化

$$\frac{1}{m}\sum_{i=1}^{m}\left(\langle\boldsymbol{x}_i,\boldsymbol{v}_1\rangle^2+\cdots+\langle\boldsymbol{x}_i,\boldsymbol{v}_k\rangle^2\right).$$

主成分分析问题的最优解

当$k=1$，最大化目标函数的向量$\boldsymbol{v}^*$为矩阵$\mathbf{X}^T\mathbf{X}$的最大特征值对应的单位特征向量。

如果在对$\mathbf{X}^T\mathbf{X}$做正交对角化时将$\mathbf{X}^T\mathbf{X}$最大特征值排在对角矩阵的第一位，那么$\boldsymbol{v}^*$是正交矩阵的第一个列向量。

如何推广到$k\geq 2$的情况?

$k=2$情况

给定$\boldsymbol{x}_1,\ldots,\boldsymbol{x}_m\in\mathbb{R}^n$，求标准正交向量$\boldsymbol{v}_1,\boldsymbol{v}_2\in\mathbb{R}^n$最大化$\frac{1}{m}\sum_{i=1}^{m}(\langle\boldsymbol{x}_i,\boldsymbol{v}_1\rangle^2+\langle\boldsymbol{x}_i,\boldsymbol{v}_2\rangle^2)$.

$k = 2$情况

给定$\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m \in \mathbb{R}^n$，求**标准正交向量**$\boldsymbol{v}_1, \boldsymbol{v}_2 \in \mathbb{R}^n$最大化$\frac{1}{m}\sum_{i=1}^{m}(\langle \boldsymbol{x}_i, \boldsymbol{v}_1\rangle^2 + \langle \boldsymbol{x}_i, \boldsymbol{v}_2\rangle^2)$.

目标函数改写为（略去$\frac{1}{m}$）：$\boldsymbol{v}_1^T\mathbf{X}^T\mathbf{X}\boldsymbol{v}_1 + \boldsymbol{v}_2^T\mathbf{X}^T\mathbf{X}\boldsymbol{v}_2$.

$k=2$情况

给定$\boldsymbol{x}_1,\ldots,\boldsymbol{x}_m\in\mathbb{R}^n$，求标准正交向量$\boldsymbol{v}_1,\boldsymbol{v}_2\in\mathbb{R}^n$最大化$\frac{1}{m}\sum_{i=1}^{m}(\langle\boldsymbol{x}_i,\boldsymbol{v}_1\rangle^2+\langle\boldsymbol{x}_i,\boldsymbol{v}_2\rangle^2)$.

目标函数改写为（略去$\frac{1}{m}$）：$\boldsymbol{v}_1^T\mathbf{X}^T\mathbf{X}\boldsymbol{v}_1+\boldsymbol{v}_2^T\mathbf{X}^T\mathbf{X}\boldsymbol{v}_2$.

考虑特殊情况：$\mathbf{X}^T\mathbf{X}$是对角矩阵 $\begin{bmatrix}\lambda_1 & 0 & \dots & 0\\ 0 & \lambda_2 & \dots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \dots & \lambda_n\end{bmatrix}$ 且满足 $\lambda_1\geq\lambda_2\geq\ldots\geq\lambda_n\geq 0$

$k=2$情况

给定$\boldsymbol{x}_1,\dots,\boldsymbol{x}_m\in\mathbb{R}^n$，求**标准正交向量**$\boldsymbol{v}_1,\boldsymbol{v}_2\in\mathbb{R}^n$最大化$\frac{1}{m}\sum_{i=1}^m(\langle\boldsymbol{x}_i,\boldsymbol{v}_1\rangle^2+\langle\boldsymbol{x}_i,\boldsymbol{v}_2\rangle^2)$.

目标函数改写为（略去$\frac{1}{m}$）：$\boldsymbol{v}_1^T\mathbf{X}^T\mathbf{X}\boldsymbol{v}_1+\boldsymbol{v}_2^T\mathbf{X}^T\mathbf{X}\boldsymbol{v}_2$.

考虑特殊情况：$\mathbf{X}^T\mathbf{X}$是对角矩阵 $\begin{bmatrix}\lambda_1 & 0 & \dots & 0\\ 0 & \lambda_2 & \dots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \dots & \lambda_n\end{bmatrix}$ 且满足 $\lambda_1\geq\lambda_2\geq\dots\geq\lambda_n\geq 0$

问题简化为：

$$\max\sum_{i=1}^n(v_{1i}^2+v_{2i}^2)\lambda_i$$
$$\text{s.t.}\sum_{i=1}^n v_{1i}^2=1,\sum_{i=1}^n v_{2i}^2=1.$$

$k=2$情况

给定$\boldsymbol{x}_1,\ldots,\boldsymbol{x}_m \in \mathbb{R}^n$，求标准正交向量$\boldsymbol{v}_1,\boldsymbol{v}_2 \in \mathbb{R}^n$最大化$\frac{1}{m}\sum_{i=1}^{m}(\langle \boldsymbol{x}_i,\boldsymbol{v}_1\rangle^2+\langle \boldsymbol{x}_i,\boldsymbol{v}_2\rangle^2)$.

目标函数改写为（略去$\frac{1}{m}$）：$\boldsymbol{v}_1^T\mathbf{X}^T\mathbf{X}\boldsymbol{v}_1+\boldsymbol{v}_2^T\mathbf{X}^T\mathbf{X}\boldsymbol{v}_2$.

考虑特殊情况：$\mathbf{X}^T\mathbf{X}$是对角矩阵 $\begin{bmatrix}\lambda_1 & 0 & \ldots & 0\\ 0 & \lambda_2 & \ldots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \ldots & \lambda_n\end{bmatrix}$ 且满足 $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq 0$

问题简化为：

$$\max \sum_{i=1}^{n}(v_{1i}^2+v_{2i}^2)\lambda_i$$

$$\text{s.t.} \sum_{i=1}^{n} v_{1i}^2=1, \sum_{i=1}^{n} v_{2i}^2=1, \sum_{i=1}^{n} v_{1i}v_{2i}=0.$$

$k=2$情况

给定$\boldsymbol{x}_1,\ldots,\boldsymbol{x}_m\in\mathbb{R}^n$，求标准正交向量$\boldsymbol{v}_1,\boldsymbol{v}_2\in\mathbb{R}^n$最大化$\frac{1}{m}\sum_{i=1}^m(\langle\boldsymbol{x}_i,\boldsymbol{v}_1\rangle^2+\langle\boldsymbol{x}_i,\boldsymbol{v}_2\rangle^2)$.

目标函数改写为（略去$\frac{1}{m}$）：$\boldsymbol{v}_1^T\mathbf{X}^T\mathbf{X}\boldsymbol{v}_1+\boldsymbol{v}_2^T\mathbf{X}^T\mathbf{X}\boldsymbol{v}_2$.

考虑特殊情况：$\mathbf{X}^T\mathbf{X}$是对角矩阵 $\begin{bmatrix}\lambda_1 & 0 & \cdots & 0\\ 0 & \lambda_2 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \lambda_n\end{bmatrix}$ 且满足 $\lambda_1\geq\lambda_2\geq\ldots\geq\lambda_n\geq 0$

问题简化为：

$$\max\sum_{i=1}^n(v_{1i}^2+v_{2i}^2)\lambda_i$$

$$\text{s.t.}\sum_{i=1}^n v_{1i}^2=1,\sum_{i=1}^n v_{2i}^2=1,\sum_{i=1}^n v_{1i}v_{2i}=0.$$

最优解之一为$\boldsymbol{v}_1^*=(1,0,\ldots,0),\boldsymbol{v}_2^*=(0,1,\ldots,0)$

当λ_i都不同时，可先证明$\boldsymbol{v}_1^*$和$\boldsymbol{v}_2^*$满足：

$\exists j,\left(v_{1j}^*\right)^2=1$且$v_{1i}^*=0,\forall i\neq j;\exists k\neq j,(v_{2k}^*)^2=1$且$v_{2i}^*=0,\forall i\neq k$

（用拉格朗日乘子法可证）

一般情况

主成分分析（Principal Component Analysis, PCA）

给定$\boldsymbol{x}_1, \dots, \boldsymbol{x}_m \in \mathbb{R}^n$及维数$k \geq 1$，求标准正交向量组$\boldsymbol{v}_1, \dots, \boldsymbol{v}_k \in \mathbb{R}^n$从而最大化

$$\frac{1}{m}\sum_{i=1}^{m}\left(\langle \boldsymbol{x}_i, \boldsymbol{v}_1\rangle^2 + \cdots + \langle \boldsymbol{x}_i, \boldsymbol{v}_k\rangle^2\right).$$

主成分分析问题的最优解

最大化目标函数的$\boldsymbol{v}_1^*, \dots, \boldsymbol{v}_k^*$为矩阵$\mathbf{X}^T\mathbf{X}$的最大的$k$个特征值对应的$k$个单位特征向量。

如果在对$\mathbf{X}^T\mathbf{X}$做正交对角化时将$\mathbf{X}^T\mathbf{X}$的特征值在对角矩阵中按序排列，那么$\boldsymbol{v}_1^*, \dots, \boldsymbol{v}_k^*$是正交矩阵的前$k$个列向量。

（一些材料直接将主成分定义为$\mathbf{X}^T\mathbf{X}$的最大的k个特征值对应的单位特征向量。根据以上分析，这k个单位特征向量其实对应一个最优化问题的最优解）

一般情况

主成分分析（Principal Component Analysis, PCA）

给定$\boldsymbol{x}_1, \dots, \boldsymbol{x}_m \in \mathbb{R}^n$及维数$k \geq 1$，求标准正交向量组$\boldsymbol{v}_1, \dots, \boldsymbol{v}_k \in \mathbb{R}^n$从而最大化

$$\frac{1}{m}\sum_{i=1}^{m}\left(\langle \boldsymbol{x}_i, \boldsymbol{v}_1\rangle^2 + \cdots + \langle \boldsymbol{x}_i, \boldsymbol{v}_k\rangle^2\right).$$

主成分分析问题的最优解

最大化目标函数的$\boldsymbol{v}_1^*, \dots, \boldsymbol{v}_k^*$为矩阵$\mathbf{X}^T\mathbf{X}$的最大的$k$个特征值对应的$k$个单位特征向量。

因为单位特征向量不唯一（如当$\lambda_1 = \lambda_2$时…），$\boldsymbol{v}_1^*, \dots, \boldsymbol{v}_k^*$不唯一

近似误差

主成分分析（Principal Component Analysis, PCA）

给定$\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m \in \mathbb{R}^n$及维数$k \geq 1$，求标准正交向量组$\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k \in \mathbb{R}^n$从而最大化

$$\frac{1}{m}\sum_{i=1}^{m}\left(\langle \boldsymbol{x}_i, \boldsymbol{v}_1\rangle^2 + \cdots + \langle \boldsymbol{x}_i, \boldsymbol{v}_k\rangle^2\right).$$

主成分分析问题的最优解

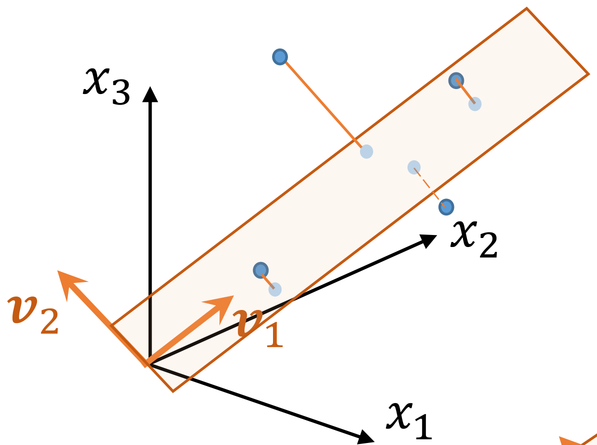
最大化目标函数的$\boldsymbol{v}_1^*, \ldots, \boldsymbol{v}_k^*$为矩阵$\mathbf{X}^T\mathbf{X}$的最大的$k$个特征值对应的$k$个单位特征向量。

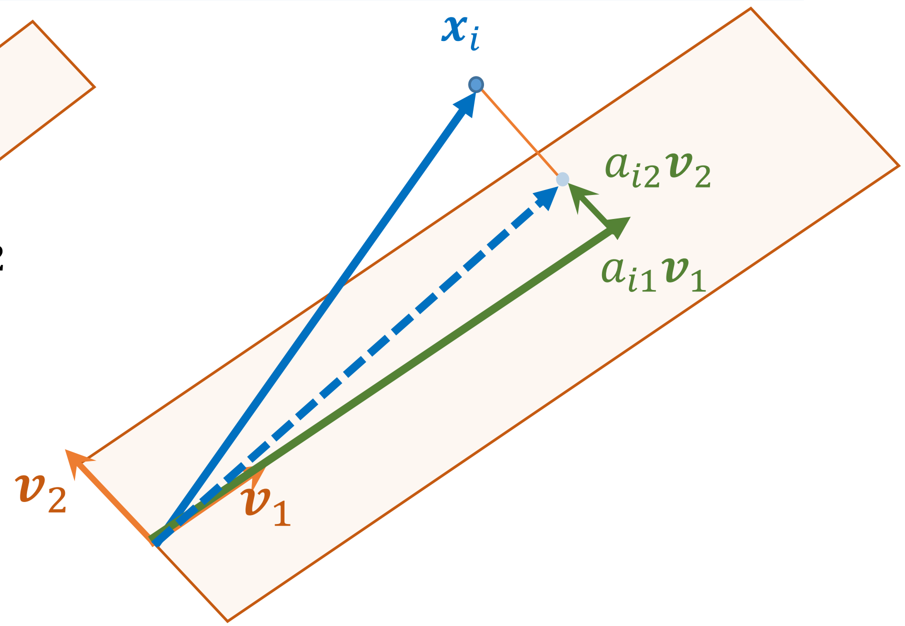
当$k < n$时，如何刻画近似误差？如何定义？如何计算？

近似误差

上一讲内容回顾

将（已完成预处理的）m个n维向量$\boldsymbol{x}_1,\ldots,\boldsymbol{x}_m\in\mathbb{R}^n$近似为$\sum_{j=1}^{k}a_{ij}\,\boldsymbol{v}_j, i=1,\ldots,m$。





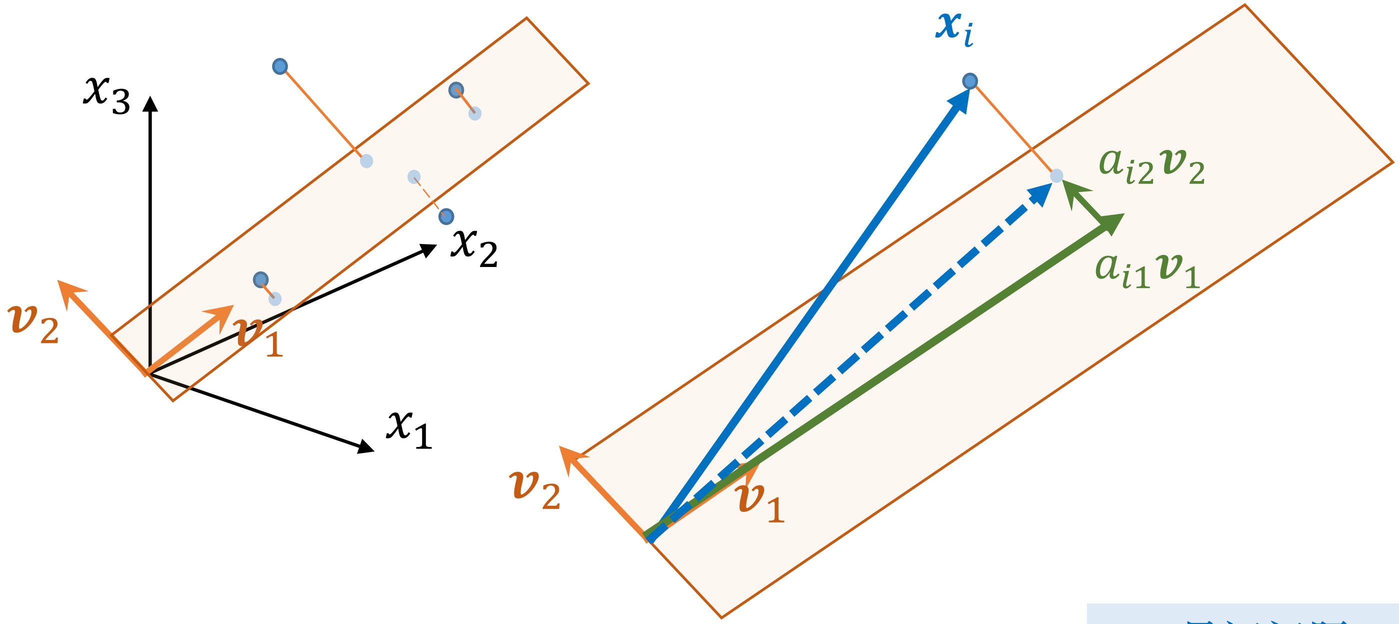
$$\operatorname{argmin}\frac{1}{m}\sum_{i=1}^{m}(\textit{distance between } \boldsymbol{x}_i \textit{ and } k-\textit{dimensional subspace spanned by } \boldsymbol{v}_1,\ldots,\boldsymbol{v}_k)^2$$

$$\operatorname{argmax}\frac{1}{m}\sum_{i=1}^{m}(\langle\boldsymbol{x}_i,\boldsymbol{v}_1\rangle^2+\cdots+\langle\boldsymbol{x}_i,\boldsymbol{v}_k\rangle^2)$$

近似误差

上一讲内容回顾

将（已完成预处理的）m个n维向量$\boldsymbol{x}_1, \dots, \boldsymbol{x}_m \in \mathbb{R}^n$近似为$\sum_{j=1}^{k} a_{ij}\,\boldsymbol{v}_j, i = 1, \dots, m$。



$$\operatorname{argmin} \frac{1}{m}\sum_{i=1}^{m}(\textit{distance between } \boldsymbol{x}_i \textit{ and } k-\textit{dimensional subspace spanned by } \boldsymbol{v}_1, \dots, \boldsymbol{v}_k)^2$$

最初问题：
最小化近似误差

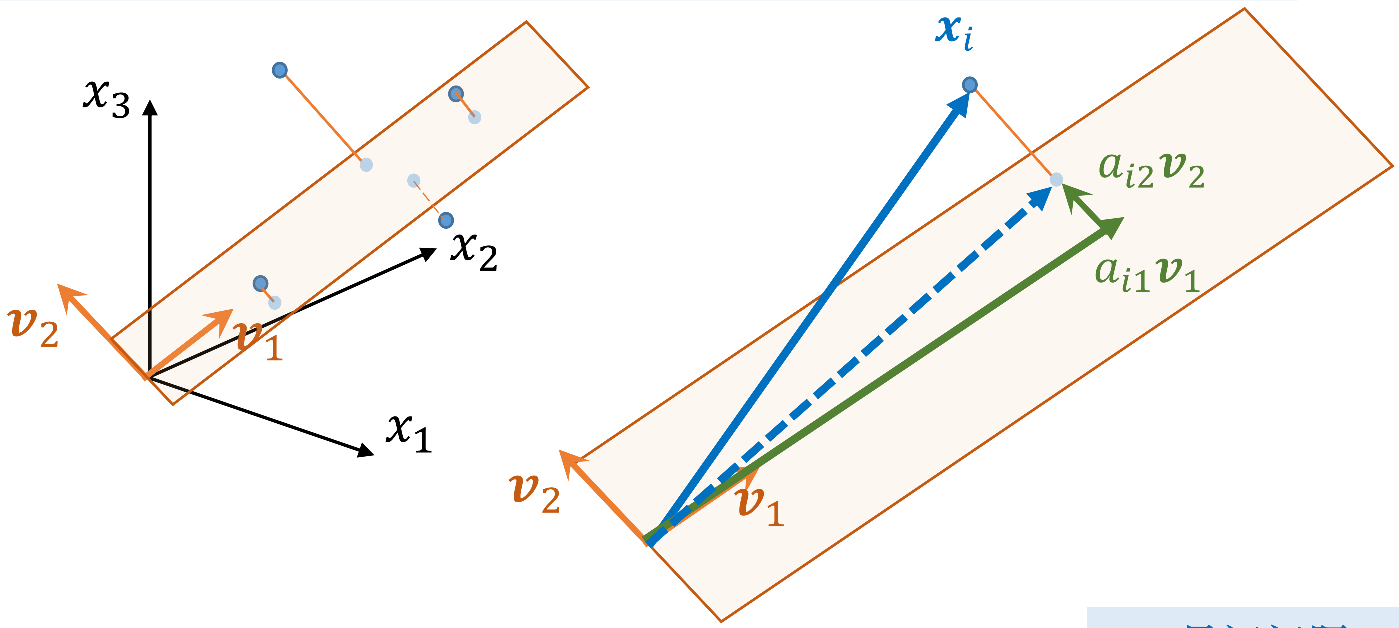
$$\operatorname{argmax} \frac{1}{m}\sum_{i=1}^{m}(\langle \boldsymbol{x}_i, \boldsymbol{v}_1\rangle^2 + \cdots + \langle \boldsymbol{x}_i, \boldsymbol{v}_k\rangle^2)$$

本讲中的目标函数

近似误差

上一讲内容回顾

将（已完成预处理的）m个n维向量$\boldsymbol{x}_1, \dots, \boldsymbol{x}_m \in \mathbb{R}^n$近似为$\sum_{j=1}^{k} a_{ij}\, \boldsymbol{v}_j, i = 1, \dots, m$。



$$\operatorname{argmin} \frac{1}{m}\sum_{i=1}^{m}(\textit{distance between } \boldsymbol{x}_i \textit{ and } k - \textit{dimensional subspace spanned by } \boldsymbol{v}_1, \dots, \boldsymbol{v}_k)^2$$

最初问题：
最小化近似误差

$$\operatorname{argmax} \frac{1}{m}\sum_{i=1}^{m}(\langle \boldsymbol{x}_i, \boldsymbol{v}_1\rangle^2 + \cdots + \langle \boldsymbol{x}_i, \boldsymbol{v}_k\rangle^2)$$

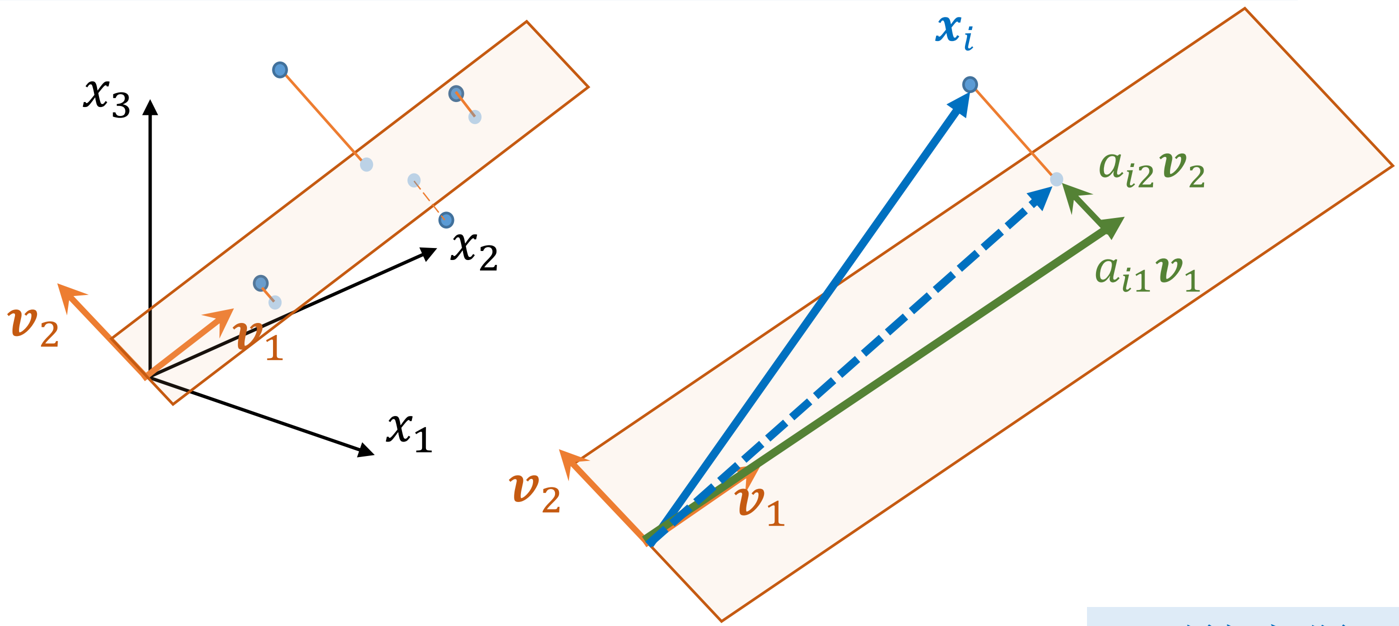
本讲中的目标函数

近似误差 $\frac{1}{m}\sum_{i=1}^{m}(\textit{distance between } \boldsymbol{x}_i \textit{ and } k - \textit{dimensional subspace spanned by } \boldsymbol{v}_1, \dots, \boldsymbol{v}_k)^2$

近似误差

上一讲内容回顾

将（已完成预处理的）m个n维向量$\boldsymbol{x}_1,\dots,\boldsymbol{x}_m \in \mathbb{R}^n$近似为$\sum_{j=1}^{k} a_{ij}\,\boldsymbol{v}_j, i=1,\dots,m$。



$$\operatorname{argmin}\frac{1}{m}\sum_{i=1}^{m}(\textit{distance between } \boldsymbol{x}_i \textit{ and } k-\textit{dimensional subspace spanned by } \boldsymbol{v}_1,\dots,\boldsymbol{v}_k)^2$$

最初问题：
最小化近似误差

$$\operatorname{argmax}\frac{1}{m}\sum_{i=1}^{m}(\langle\boldsymbol{x}_i,\boldsymbol{v}_1\rangle^2+\cdots+\langle\boldsymbol{x}_i,\boldsymbol{v}_k\rangle^2)$$

本讲中的目标函数

近似误差

$$\frac{1}{m}\sum_{i=1}^{m}\|\boldsymbol{x}_i\|_2^2-\frac{1}{m}\sum_{i=1}^{m}(\langle\boldsymbol{x}_i,\boldsymbol{v}_1\rangle^2+\cdots+\langle\boldsymbol{x}_i,\boldsymbol{v}_k\rangle^2)$$

近似误差

分析思路：用矩阵$\mathbf{X}^T\mathbf{X}$的特征值刻画结果

近似误差 $$\frac{1}{m}\sum_{i=1}^{m}\|\boldsymbol{x}_i\|_2^2-\frac{1}{m}\sum_{i=1}^{m}\left(\langle\boldsymbol{x}_i,\boldsymbol{v}_1\rangle^2+\cdots+\langle\boldsymbol{x}_i,\boldsymbol{v}_k\rangle^2\right)$$

近似误差

分析思路：用矩阵$\mathbf{X}^T\mathbf{X}$的特征值刻画结果

近似误差 $\frac{1}{m}\sum_{i=1}^{m}\|\boldsymbol{x}_i\|_2^2-\frac{1}{m}\sum_{i=1}^{m}(\langle\boldsymbol{x}_i,\boldsymbol{v}_1\rangle^2+\cdots+\langle\boldsymbol{x}_i,\boldsymbol{v}_k\rangle^2)$

定义$\mathbf{V}$为由列向量$\boldsymbol{v}_1,\dots,\boldsymbol{v}_k$构成的$n\times k$矩阵

分析思路：用矩阵$\mathbf{X}^T\mathbf{X}$的特征值刻画结果

近似误差

近似误差 $\frac{1}{m}\sum_{i=1}^{m}\|\boldsymbol{x}_i\|_2^2-\frac{1}{m}\sum_{i=1}^{m}(\underbrace{\langle\boldsymbol{x}_i,\boldsymbol{v}_1\rangle}_{\text{标量}}^2+\cdots+\langle\boldsymbol{x}_i,\boldsymbol{v}_k\rangle^2)$

定义$\mathbf{V}$为由列向量$\boldsymbol{v}_1,\ldots,\boldsymbol{v}_k$构成的$n\times k$矩阵

$$=\frac{1}{m}\sum_{i=1}^{m}\|\underbrace{\boldsymbol{x}_i}_{n\times1\text{列向量}}\|_2^2-\frac{1}{m}\sum_{i=1}^{m}\|\underbrace{\mathbf{V}^T\boldsymbol{x}_i}_{k\times1\text{列向量}}\|_2^2$$

近似误差

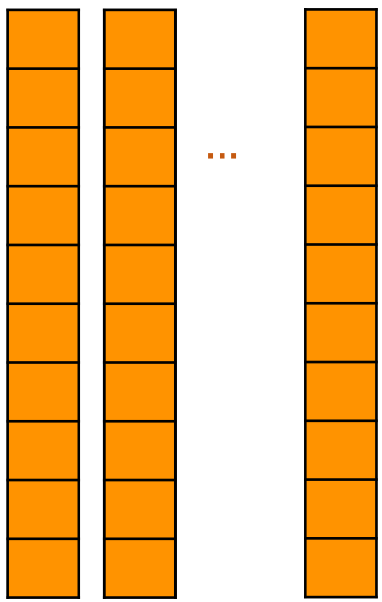
分析思路：用矩阵$\mathbf{X}^T\mathbf{X}$的特征值刻画结果

近似误差 $\frac{1}{m}\sum_{i=1}^{m}\|\boldsymbol{x}_i\|_2^2-\frac{1}{m}\sum_{i=1}^{m}(\langle\boldsymbol{x}_i,\boldsymbol{v}_1\rangle^2+\cdots+\langle\boldsymbol{x}_i,\boldsymbol{v}_k\rangle^2)$

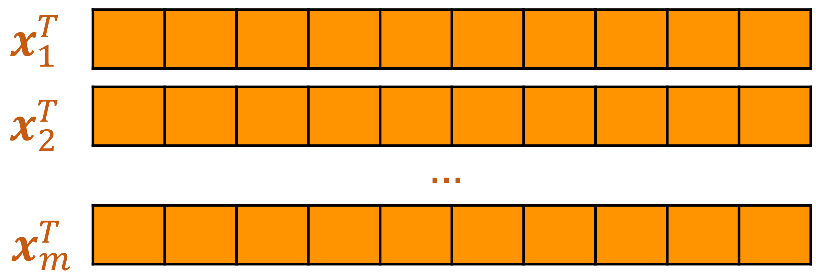
定义$\mathbf{V}$为由列向量$\boldsymbol{v}_1,\ldots,\boldsymbol{v}_k$构成的$n\times k$矩阵

$$=\frac{1}{m}\sum_{i=1}^{m}\underbrace{\|\boldsymbol{x}_i\|_2^2}_{n\times1\text{列向量}}-\frac{1}{m}\sum_{i=1}^{m}\underbrace{\|\mathbf{V}^T\boldsymbol{x}_i\|_2^2}_{k\times1\text{列向量}}$$

分析$n\times n$矩阵$\mathbf{X}^T\mathbf{X}$



$n\times m$矩阵$\mathbf{X}^T$



$m\times n$矩阵$\mathbf{X}$

$$\sum_{i=1}^{m}\|\boldsymbol{x}_i\|_2^2=\mathbf{tr}(\mathbf{X}^T\mathbf{X})$$

（展开$\mathbf{X}^T\mathbf{X}$中对角线元素可得）

迹：方阵对角线上元素之和

$$\mathrm{tr}(\mathbf{A})=\sum_{i=1}^{n}a_{ii}$$

近似误差

分析思路：用矩阵$\mathbf{X}^T\mathbf{X}$的特征值刻画结果

近似误差 $\frac{1}{m}\sum_{i=1}^{m}\|\boldsymbol{x}_i\|_2^2-\frac{1}{m}\sum_{i=1}^{m}(\langle\boldsymbol{x}_i,\boldsymbol{v}_1\rangle^2+\cdots+\langle\boldsymbol{x}_i,\boldsymbol{v}_k\rangle^2)$

定义$\mathbf{V}$为由列向量$\boldsymbol{v}_1,\ldots,\boldsymbol{v}_k$构成的$n\times k$矩阵

$$=\frac{1}{m}\sum_{i=1}^{m}\|\underbrace{\boldsymbol{x}_i}_{n\times 1\text{列向量}}\|_2^2-\frac{1}{m}\sum_{i=1}^{m}\|\underbrace{\mathbf{V}^T\boldsymbol{x}_i}_{k\times 1\text{列向量}}\|_2^2$$

$$=\frac{1}{m}\mathbf{tr}(\underbrace{\mathbf{X}^T\mathbf{X}}_{n\times n\text{方阵}})-\frac{1}{m}\mathbf{tr}(\underbrace{\mathbf{V}^T\mathbf{X}^T\mathbf{X}\mathbf{V}}_{k\times k\text{方阵}})$$

Given any $n\times n$ real or complex matrix $\mathbf{A}$, there is

$$\boxed{\operatorname{tr}(\mathbf{A})=\sum_{i=1}^{n}\lambda_i}$$

近似误差

分析思路：用矩阵$\mathbf{X}^T\mathbf{X}$的特征值刻画结果

近似误差 $\frac{1}{m}\sum_{i=1}^{m}\|\boldsymbol{x}_i\|_2^2-\frac{1}{m}\sum_{i=1}^{m}(\langle\boldsymbol{x}_i,\boldsymbol{v}_1\rangle^2+\cdots+\langle\boldsymbol{x}_i,\boldsymbol{v}_k\rangle^2)$

定义$\mathbf{V}$为由列向量$\boldsymbol{v}_1,\dots,\boldsymbol{v}_k$构成的$n\times k$矩阵

$$=\frac{1}{m}\sum_{i=1}^{m}\|\boldsymbol{x}_i\|_2^2-\frac{1}{m}\sum_{i=1}^{m}\|\mathbf{V}^T\boldsymbol{x}_i\|_2^2$$

$$=\frac{1}{m}\mathbf{tr}(\underbrace{\mathbf{X}^T\mathbf{X}}_{n\times n\text{方阵}})-\frac{1}{m}\mathbf{tr}(\underbrace{\mathbf{V}^T\mathbf{X}^T\mathbf{X}\mathbf{V}}_{k\times k\text{方阵}})$$

$$=\frac{1}{m}\sum_{i=1}^{n}\lambda_i-\frac{1}{m}\mathbf{tr}(\mathbf{V}^T\mathbf{Q}\mathbf{D}\mathbf{Q}^T\mathbf{V})$$

Given any $n\times n$ real or complex matrix $\mathbf{A}$, there is

$$\boxed{\operatorname{tr}(\mathbf{A})=\sum_{i=1}^{n}\lambda_i}$$

$\mathbf{V}^\mathrm{T}\mathbf{Q}\mathbf{D}\mathbf{Q}^\mathrm{T}\mathbf{V}$为 $\begin{bmatrix}\lambda_1 & 0 & \dots & 0\\ 0 & \lambda_2 & \dots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \dots & \lambda_k\end{bmatrix}$ 可利用标准正交向量组性质得

$k\times k$方阵，不是矩阵**D**

近似误差

分析思路：用矩阵$\mathbf{X}^T\mathbf{X}$的特征值刻画结果

近似误差 $\frac{1}{m}\sum_{i=1}^{m}\|\boldsymbol{x}_i\|_2^2 - \frac{1}{m}\sum_{i=1}^{m}(\langle\boldsymbol{x}_i,\boldsymbol{v}_1\rangle^2 + \cdots + \langle\boldsymbol{x}_i,\boldsymbol{v}_k\rangle^2)$

定义$\mathbf{V}$为由列向量$\boldsymbol{v}_1,\ldots,\boldsymbol{v}_k$构成的$n\times k$矩阵

$$= \frac{1}{m}\sum_{i=1}^{m}\|\boldsymbol{x}_i\|_2^2 - \frac{1}{m}\sum_{i=1}^{m}\|\mathbf{V}^T\boldsymbol{x}_i\|_2^2$$

$$= \frac{1}{m}\mathbf{tr}(\underbrace{\mathbf{X}^T\mathbf{X}}_{n\times n\text{方阵}}) - \frac{1}{m}\mathbf{tr}(\underbrace{\mathbf{V}^T\mathbf{X}^T\mathbf{X}\mathbf{V}}_{k\times k\text{方阵}})$$

$$= \frac{1}{m}\sum_{i=1}^{n}\lambda_i - \frac{1}{m}\mathbf{tr}(\mathbf{V}^T\mathbf{Q}\mathbf{D}\mathbf{Q}^T\mathbf{V})$$

$$= \frac{1}{m}\sum_{i=1}^{n}\lambda_i - \frac{1}{m}\sum_{i=1}^{k}\lambda_i$$

$$= \frac{1}{m}\sum_{i=k+1}^{n}\lambda_i$$

近似误差

主成分分析（Principal Component Analysis, PCA）

给定$\boldsymbol{x}_1, \dots, \boldsymbol{x}_m \in \mathbb{R}^n$及维数$k \geq 1$，求标准正交向量组$\boldsymbol{v}_1, \dots, \boldsymbol{v}_k \in \mathbb{R}^n$从而最大化

$$\frac{1}{m}\sum_{i=1}^{m}\left(\langle \boldsymbol{x}_i, \boldsymbol{v}_1\rangle^2 + \cdots + \langle \boldsymbol{x}_i, \boldsymbol{v}_k\rangle^2\right).$$

主成分分析问题的最优解

最大化目标函数的$\boldsymbol{v}_1^*, \dots, \boldsymbol{v}_k^*$为矩阵$\mathbf{X}^T\mathbf{X}$的最大的$k$个特征值对应的$k$个单位特征向量。

近似误差

$$\frac{1}{m}\sum_{i=1}^{m}(distance\ between\ \boldsymbol{x}_i\ and\ k-dimensional\ subspace\ spanned\ by\ \boldsymbol{v}_1, \dots, \boldsymbol{v}_k)^2$$

等于矩阵$\mathbf{X}^T\mathbf{X}$的最小的$n-k$个特征值之和

该值越小，近似误差越小

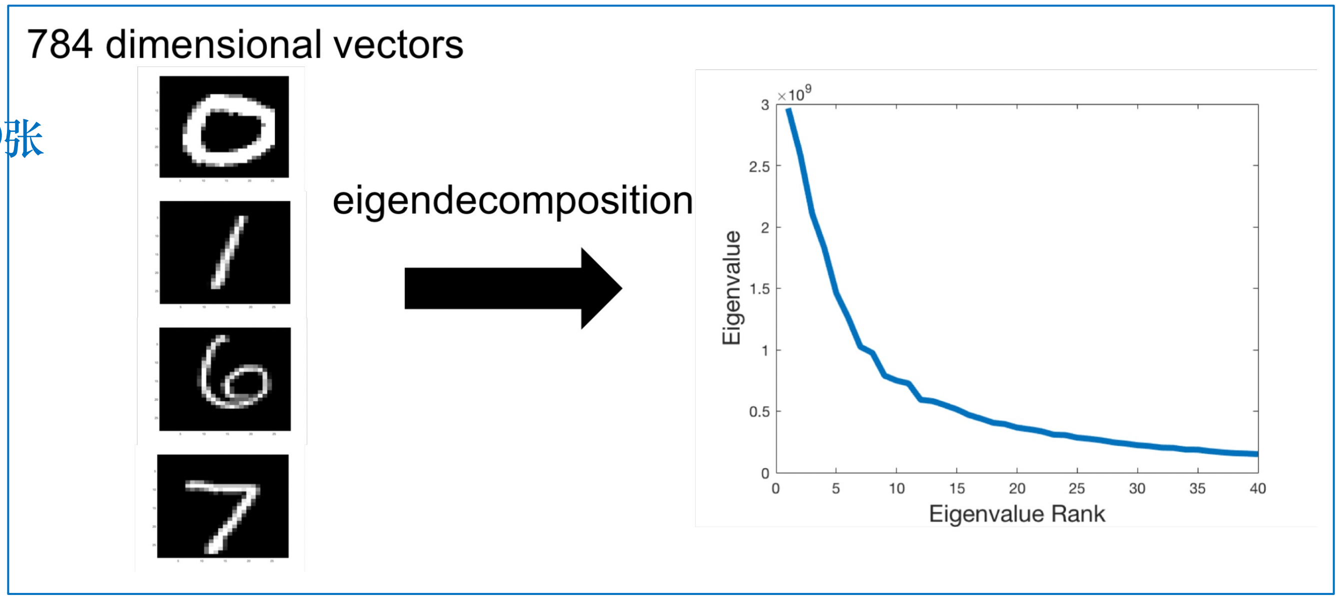
近似误差

近似误差 $\frac{1}{m}\sum_{i=1}^{m}(distance\ between\ \boldsymbol{x}_i\ and\ k-dimensional\ subspace\ spanned\ by\ \boldsymbol{v}_1,\ldots,\boldsymbol{v}_k)^2$

等于矩阵$\mathbf{X}^T\mathbf{X}$的最小的$n-k$个特征值之和

该值越小，近似误差越小

如有60000张



图例取自Cameron Musco <COMPSCI 514 - Algorithms for Data Science>

近似误差

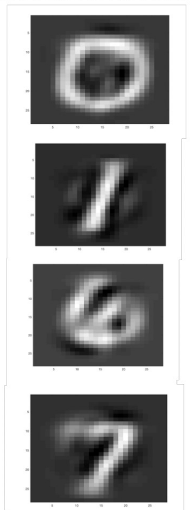
近似误差 $\frac{1}{m}\sum_{i=1}^{m}(distance\ between\ \boldsymbol{x}_i\ and\ k-dimensional\ subspace\ spanned\ by\ \boldsymbol{v}_1,\ldots,\boldsymbol{v}_k)^2$

等于矩阵$\mathbf{X}^T\mathbf{X}$的最小的$n-k$个特征值之和

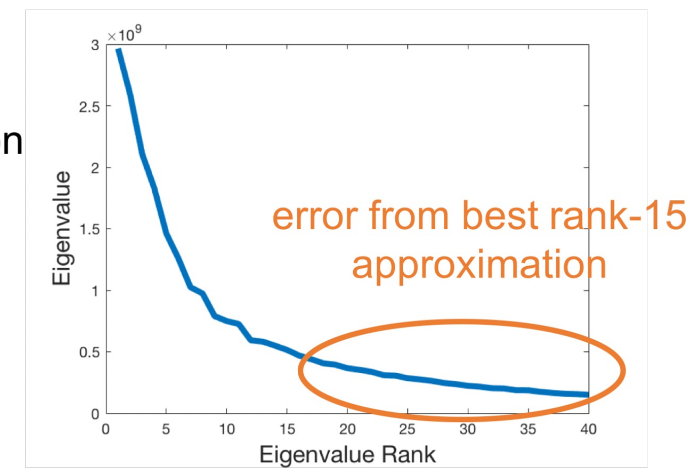
该值越小，近似误差越小

784 dimensional vectors

如有60000张



eigendecomposition



图例取自Cameron Musco <COMPSCI 514 - Algorithms for Data Science>

主成分分析

幂迭代法

幂迭代法

主成分分析问题的最优解

最大化目标函数的$\boldsymbol{v}_1^*, \dots, \boldsymbol{v}_k^*$为矩阵$\mathbf{X}^T\mathbf{X}$的最大的$k$个特征值对应的$k$个单位特征向量。

如何算矩阵的特征向量?

（1）奇异值分解（Singular Value Decomposition）（下一讲内容）

（2）幂迭代法（Power Iteration Method）

幂迭代法

计算矩阵最大特征值对应的单位特征向量（即第一主成分）：

幂迭代法

0 随机选择一个单位向量$\boldsymbol{u}_0 \in \mathbb{R}^n$; 计算$\mathbf{A} = \mathbf{X}^T\mathbf{X}$

1 for i=1, 2, ...

2 计算$\boldsymbol{u}_i = \boldsymbol{A}\boldsymbol{u}_{i-1}$

3 检查终止条件：若$\frac{\boldsymbol{u}_i}{\|\boldsymbol{u}_i\|} \approx \frac{\boldsymbol{u}_{i-1}}{\|\boldsymbol{u}_{i-1}\|}$，结束算法、输出$\frac{\boldsymbol{u}_i}{\|\boldsymbol{u}_i\|}$

确保是单位向量

为什么可以输出A的最大特征值对应的单位特征向量?

分析 $\boldsymbol{u}_i=\mathbf{A}^i\boldsymbol{u}_0$ 的几何意义

幂迭代法的几何解释

分析 $\boldsymbol{u}_i=\mathbf{A}^i\boldsymbol{u}_0$ 的几何意义

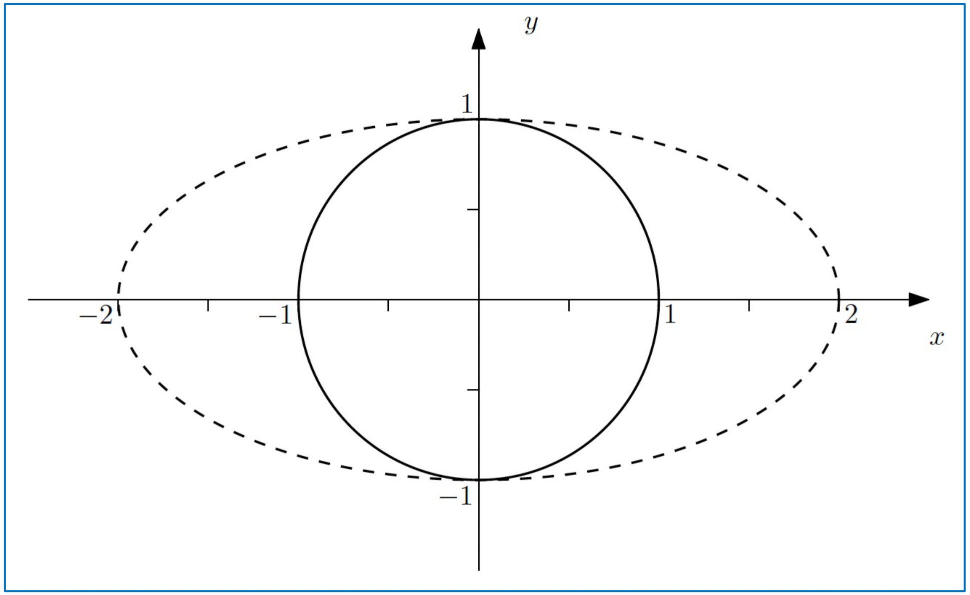
先考虑特殊情况：假设**A**是对角矩阵 $\begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}$ 且满足 $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq 0$

幂迭代法的几何解释

分析 $\boldsymbol{u}_i=\mathbf{A}^i\boldsymbol{u}_0$ 的几何意义

先考虑特殊情况：假设**A**是对角矩阵 $\begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}$ 且满足 $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq 0$

若矩阵**A**为 $\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$ 设$\boldsymbol{u}_0 = (x, y)^T$



$\boldsymbol{u}_0 = (x, y)^T$有可能位置：实线圆

$\boldsymbol{u}_1 = (2x, y)^T$所有可能位置：虚线椭圆

示例取自Tim Roughgarden and Gregory Valiant <CS 168 - The Modern Algorithmic Toolbox>

幂迭代法的几何解释

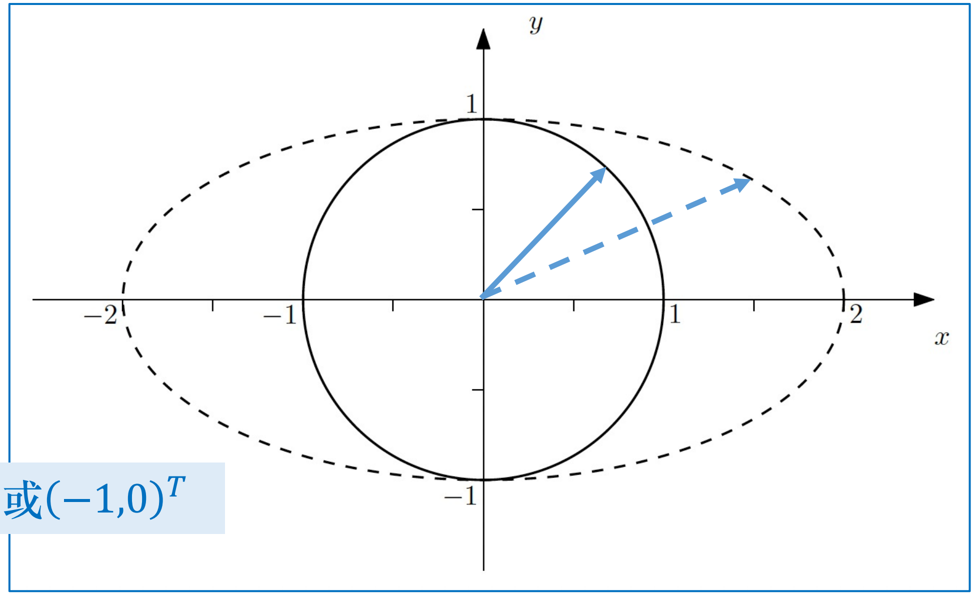
分析 $\boldsymbol{u}_i = \mathbf{A}^i \boldsymbol{u}_0$ 的几何意义

先考虑特殊情况：假设**A**是对角矩阵 $\begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}$ 且满足 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$

若矩阵**A**为 $\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$ 设 $\boldsymbol{u}_0 = (x, y)^T$

$\boldsymbol{u}_0 = (x, y)^T$有可能位置：实线圆

$\boldsymbol{u}_1 = (2x, y)^T$所有可能位置：虚线椭圆



特征值2对应的单位特征向量为$(1,0)^T$或$(-1,0)^T$

幂迭代法的几何解释

分析 $\boldsymbol{u}_i=\mathbf{A}^i\boldsymbol{u}_0$ 的几何意义

再考虑一般情况：**A**可写成$\mathbf{QDQ}^T$，**Q**是正交矩阵、**D**是$\begin{bmatrix}\lambda_1 & 0 & \dots & 0\\ 0 & \lambda_2 & \dots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \dots & \lambda_n\end{bmatrix}$且 $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq 0$

若矩阵**A**为 $\begin{pmatrix}\frac{3}{2} & \frac{1}{2}\\ \frac{1}{2} & \frac{3}{2}\end{pmatrix} = \underbrace{\begin{pmatrix}\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{pmatrix}}_{\text{rotate back } 45^\circ} \cdot \underbrace{\begin{pmatrix}2 & 0\\ 0 & 1\end{pmatrix}}_{\text{stretch}} \cdot \underbrace{\begin{pmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{pmatrix}}_{\text{rotate clockwise } 45^\circ}$ 此**A**仅为示例，未必由 $\mathbf{X}^T\mathbf{X}$得到

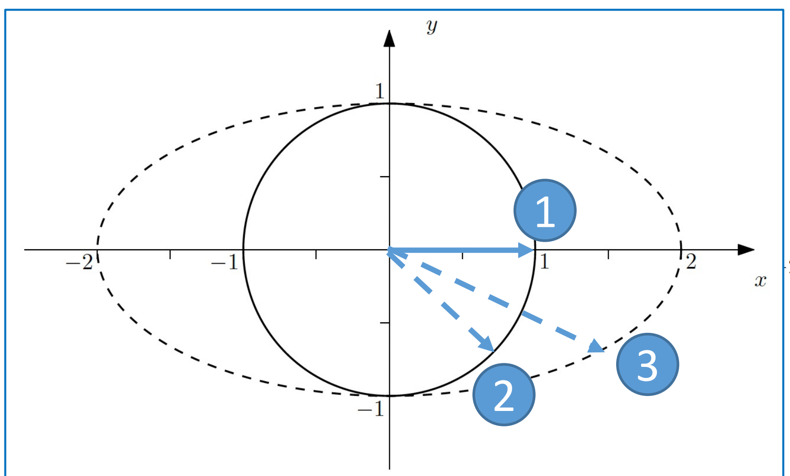
示例修改自Tim Roughgarden and Gregory Valiant <CS 168 - The Modern Algorithmic Toolbox>

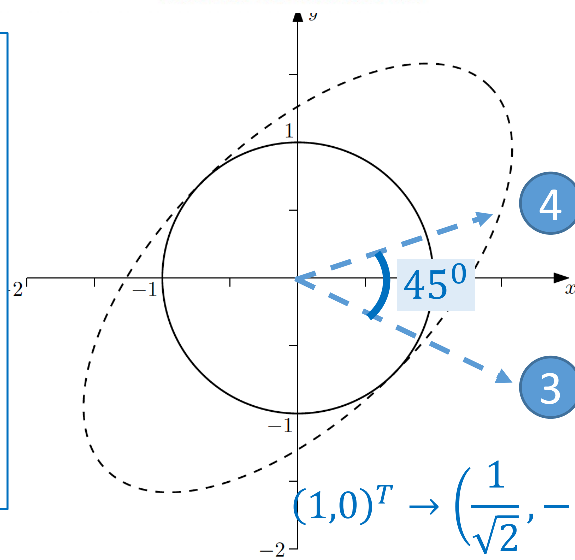
幂迭代法的几何解释

分析 $\boldsymbol{u}_i$=$\mathbf{A}^i\boldsymbol{u}_0$ 的几何意义

再考虑一般情况：**A**可写成**QDQ**^{T}，**Q**是正交矩阵、**D**是 $\begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}$ 且 $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq 0$

若矩阵**A**为 $\begin{pmatrix} \frac{3}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{3}{2} \end{pmatrix} = \underbrace{\begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}}_{\text{rotate back } 45^\circ} \cdot \underbrace{\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}}_{\text{stretch}} \cdot \underbrace{\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}}_{\text{rotate clockwise } 45^\circ}$ 此**A**仅为示例，未必由 $\mathbf{X}^T\mathbf{X}$得到





$(1,0)^T \rightarrow \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)^T \rightarrow \left(\sqrt{2}, -\frac{1}{\sqrt{2}}\right)^T \rightarrow \left(\frac{3}{2}, \frac{1}{2}\right)^T$

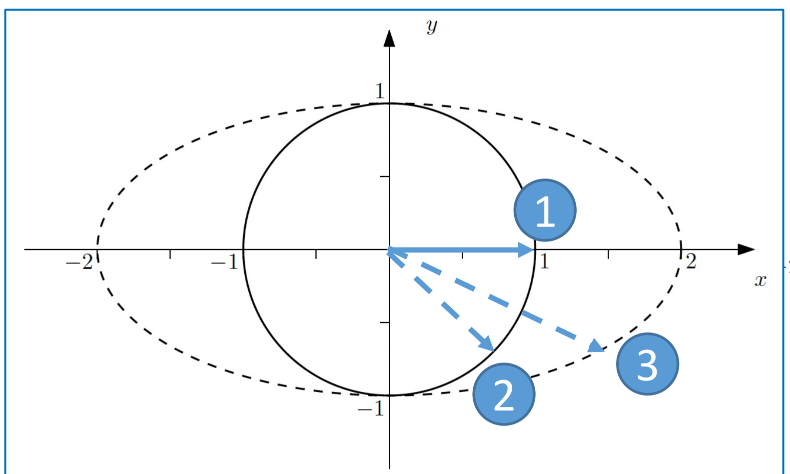
示例修改自Tim Roughgarden and Gregory Valiant <CS 168 - The Modern Algorithmic Toolbox>

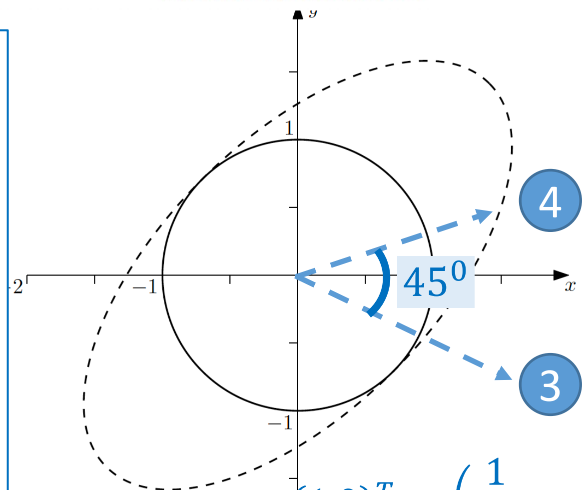
幂迭代法的几何解释

分析 $\boldsymbol{u}_i=\mathbf{A}^i\boldsymbol{u}_0$ 的几何意义

再考虑一般情况：A可写成$\mathbf{QDQ}^T$，Q是正交矩阵、D是 $\begin{bmatrix}\lambda_1 & 0 & \dots & 0\\ 0 & \lambda_2 & \dots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \dots & \lambda_n\end{bmatrix}$ 且 $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq 0$

若矩阵A为 $\begin{pmatrix}\frac{3}{2} & \frac{1}{2}\\ \frac{1}{2} & \frac{3}{2}\end{pmatrix} = \underbrace{\begin{pmatrix}\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{pmatrix}}_{\text{rotate back } 45^\circ} \cdot \underbrace{\begin{pmatrix}2 & 0\\ 0 & 1\end{pmatrix}}_{\text{stretch}} \cdot \underbrace{\begin{pmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{pmatrix}}_{\text{rotate clockwise } 45^\circ}$ **此A仅为示例，未必由 $\mathbf{X}^T\mathbf{X}$得到**





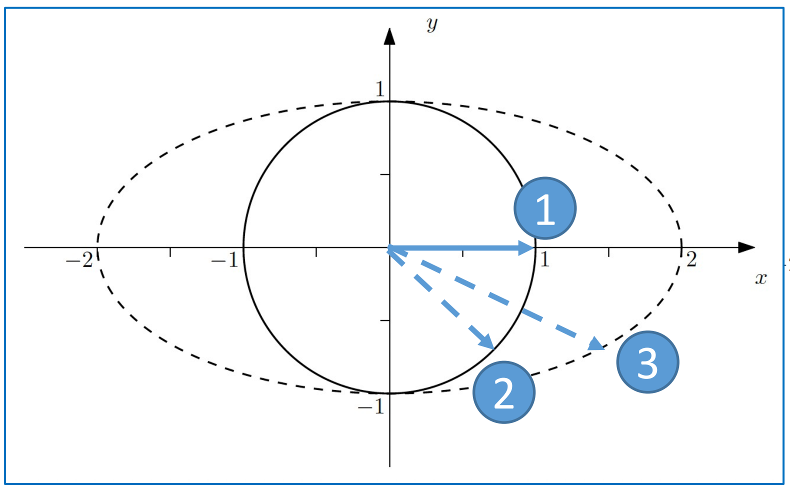
$(1,0)^T \rightarrow \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)^T \rightarrow \left(\sqrt{2}, -\frac{1}{\sqrt{2}}\right)^T \rightarrow \left(\frac{3}{2}, \frac{1}{2}\right)^T$

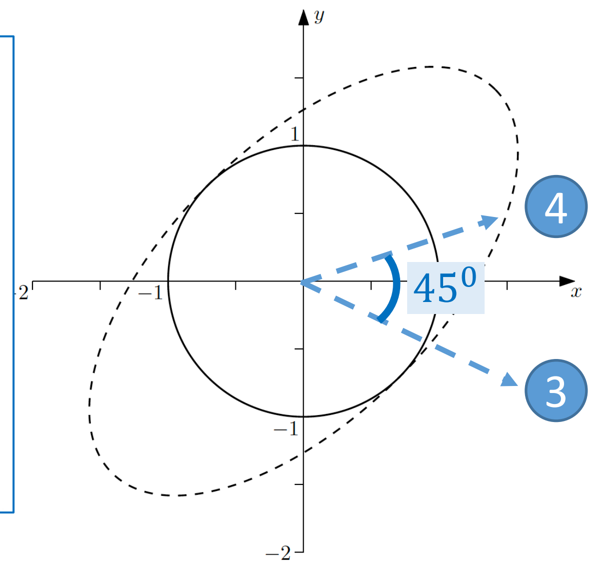
特征值2对应的单位特征向量为$\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)^T$ 或 $\left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)^T$

幂迭代法的几何解释

分析 $\boldsymbol{u}_i$=$\mathbf{A}^i\boldsymbol{u}_0$ 的几何意义

再考虑一般情况：**A**可写成**QDQ**^{T}，**Q**是正交矩阵、**D**是 $\begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}$ 且 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$



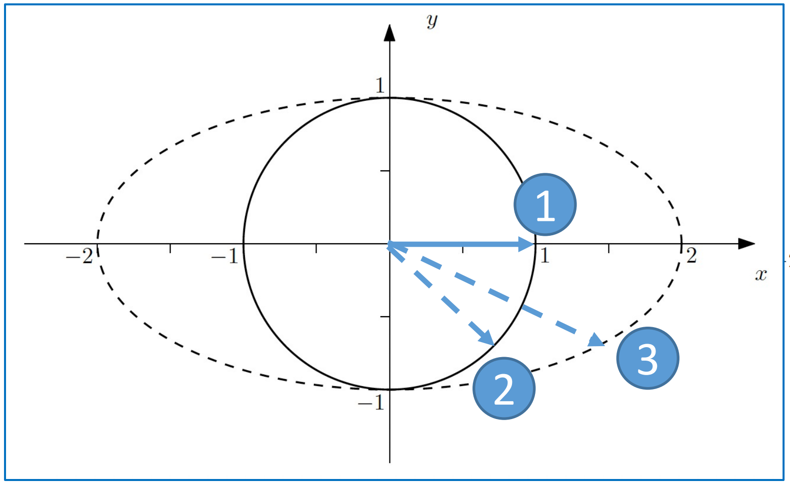


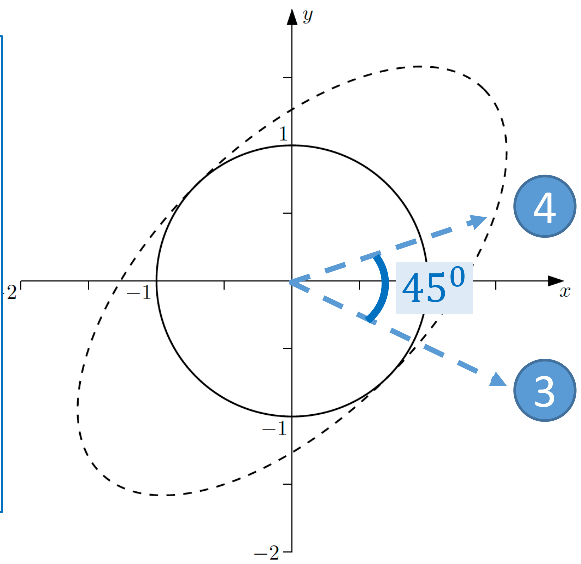
以上仅分析了$i = 1$的情况，$i > 1$时？

幂迭代法的几何解释

分析 $\boldsymbol{u}_i$=$\mathbf{A}^i\boldsymbol{u}_0$ 的几何意义

再考虑一般情况：**A**可写成$\mathbf{QDQ}^T$，**Q**是正交矩阵、**D**是$\begin{bmatrix}\lambda_1 & 0 & \dots & 0\\ 0 & \lambda_2 & \dots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \dots & \lambda_n\end{bmatrix}$且 $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq 0$





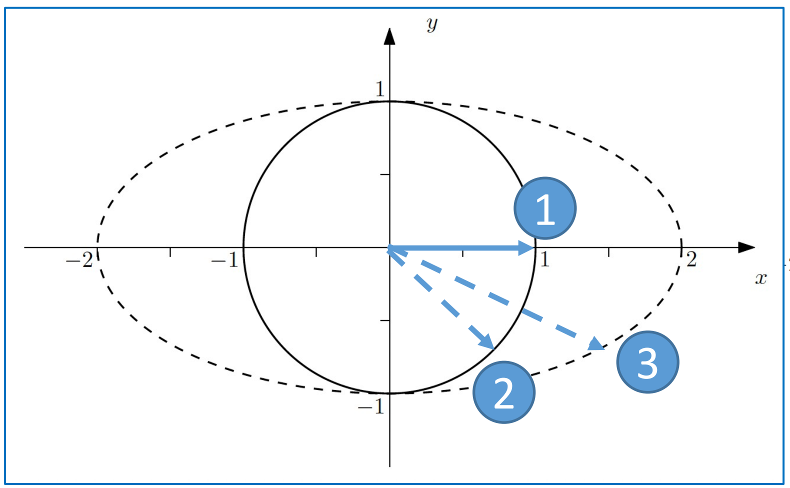
以上仅分析了$i=1$的情况，$i>1$时？

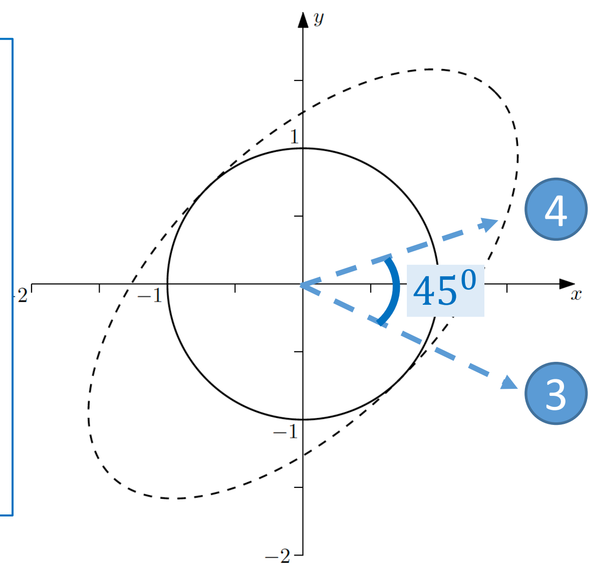
可证明，若$\mathbf{A}=\mathbf{QDQ}^T$，有$\mathbf{A}^i=\mathbf{QD}^i\mathbf{Q}^T$（用正交矩阵的性质$\mathbf{Q}^T\mathbf{Q}=\mathbf{I}$）

幂迭代法的几何解释

分析 $\boldsymbol{u}_i$=$\mathbf{A}^i\boldsymbol{u}_0$ 的几何意义

再考虑一般情况：$\mathbf{A}$可写成$\mathbf{QDQ}^T$，$\mathbf{Q}$是正交矩阵、$\mathbf{D}$是$\begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}$且 $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq 0$





以上仅分析了$i = 1$的情况，$i > 1$时？

可证明，若$\mathbf{A} = \mathbf{QDQ}^T$，有$\mathbf{A}^i = \mathbf{QD}^i\mathbf{Q}^T$（用正交矩阵的性质$\mathbf{Q}^T\mathbf{Q} = \mathbf{I}$）

当i很大时，$\frac{\boldsymbol{u}_i}{\|\boldsymbol{u}_i\|}$将趋近于$\mathbf{A}$的最大特征值对应的向量

幂迭代法的证明

计算矩阵最大特征值对应的单位特征向量（即第一主成分）：

幂迭代法

0 随机选择一个单位向量$\boldsymbol{u}_0 \in \mathbb{R}^n$; 计算$\mathbf{A} = \mathbf{X}^T\mathbf{X}$

1 for i=1, 2, ...

2 　计算$\boldsymbol{u}_i = \boldsymbol{A}\boldsymbol{u}_{i-1}$

3 　检查终止条件：若$\frac{\boldsymbol{u}_i}{\|\boldsymbol{u}_i\|} \approx \frac{\boldsymbol{u}_{i-1}}{\|\boldsymbol{u}_{i-1}\|}$，结束算法、输出$\frac{\boldsymbol{u}_i}{\|\boldsymbol{u}_i\|}$

用$\boldsymbol{q}_1$表示**A**的最大特征值对应的单位特征向量，如何证明当i很大时$\left|\left\langle\frac{\boldsymbol{u}_i}{\|\boldsymbol{u}_i\|}, \boldsymbol{q}_1\right\rangle\right|$趋近于1?

幂迭代法的证明

证明思路：
（1）用$\boldsymbol{q}_1, \boldsymbol{q}_2, \dots, \boldsymbol{q}_n$表示$\mathbf{A}^i\boldsymbol{u}_0$
（2）用$\lambda_1, \lambda_2, \dots$刻画目标数值

用$\boldsymbol{q}_1$表示$\mathbf{A}$的最大特征值对应的单位特征向量，证明当i很大时$\left|\left\langle \frac{\mathbf{A}^i\boldsymbol{u}_0}{\|\mathbf{A}^i\boldsymbol{u}_0\|}, \boldsymbol{q}_1 \right\rangle\right|$趋近于1

用$\boldsymbol{q}_1, \boldsymbol{q}_2, \dots, \boldsymbol{q}_n \in \mathbb{R}^n$表示$\mathbf{A}$的$n$个单位特征向量（对应$\lambda_1 \geq \lambda_2 \geq \cdots \lambda_n$），是一组标准正交向量

幂迭代法的证明

证明思路：

（1）用$\boldsymbol{q}_1, \boldsymbol{q}_2, \dots, \boldsymbol{q}_n$表示$\mathbf{A}^i\boldsymbol{u}_0$

（2）用$\lambda_1, \lambda_2, \dots$刻画目标数值

用$\boldsymbol{q}_1$表示$\mathbf{A}$的最大特征值对应的单位特征向量，证明当i很大时$\left|\left\langle \frac{\mathbf{A}^i\boldsymbol{u}_0}{\|\mathbf{A}^i\boldsymbol{u}_0\|}, \boldsymbol{q}_1 \right\rangle\right|$趋近于1

用$\boldsymbol{q}_1, \boldsymbol{q}_2, \dots, \boldsymbol{q}_n \in \mathbb{R}^n$表示$\mathbf{A}$的$n$个单位特征向量（对应$\lambda_1 \geq \lambda_2 \geq \cdots \lambda_n$），是一组标准正交向量

因为$\boldsymbol{u}_0 \in \mathbb{R}^n$，一定可将其写成$\boldsymbol{q}_1, \boldsymbol{q}_2, \dots, \boldsymbol{q}_n$的线性组合，将结果记为$\boldsymbol{u}_0 = \sum_{j=1}^{n} c_j \boldsymbol{q}_j$

幂迭代法的证明

证明思路：

（1）用$\boldsymbol{q}_1, \boldsymbol{q}_2, \dots, \boldsymbol{q}_n$表示$\mathbf{A}^i\boldsymbol{u}_0$

（2）用$\lambda_1, \lambda_2, \dots$刻画目标数值

用$\boldsymbol{q}_1$表示$\mathbf{A}$的最大特征值对应的单位特征向量，证明当i很大时$\left|\left\langle \frac{\mathbf{A}^i\boldsymbol{u}_0}{\|\mathbf{A}^i\boldsymbol{u}_0\|}, \boldsymbol{q}_1 \right\rangle\right|$趋近于1

用$\boldsymbol{q}_1, \boldsymbol{q}_2, \dots, \boldsymbol{q}_n \in \mathbb{R}^n$表示$\mathbf{A}$的$n$个单位特征向量（对应$\lambda_1 \ge \lambda_2 \ge \cdots \lambda_n$），是一组标准正交向量

因为$\boldsymbol{u}_0 \in \mathbb{R}^n$，一定可将其写成$\boldsymbol{q}_1, \boldsymbol{q}_2, \dots, \boldsymbol{q}_n$的线性组合，将结果记为$\boldsymbol{u}_0 = \sum_{j=1}^{n} c_j \boldsymbol{q}_j$

可以将$\mathbf{A}$写成$\mathbf{QDQ}^T$，其中$\mathbf{Q}$的各列是$\boldsymbol{q}_1, \boldsymbol{q}_2, \dots, \boldsymbol{q}_n$，$\mathbf{D}$是由$\lambda_1, \lambda_2, \dots, \lambda_n$确定的对角矩阵

幂迭代法的证明

证明思路：
（1）用$\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_n$表示$\mathbf{A}^i\boldsymbol{u}_0$
（2）用$\lambda_1, \lambda_2, \ldots$刻画目标数值

用$\boldsymbol{q}_1$表示$\mathbf{A}$的最大特征值对应的单位特征向量，证明当i很大时$\left|\left\langle \frac{\mathbf{A}^i\boldsymbol{u}_0}{\|\mathbf{A}^i\boldsymbol{u}_0\|}, \boldsymbol{q}_1 \right\rangle\right|$趋近于1

用$\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_n \in \mathbb{R}^n$表示$\mathbf{A}$的$n$个单位特征向量（对应$\lambda_1 \geq \lambda_2 \geq \cdots \lambda_n$），是一组标准正交向量

因为$\boldsymbol{u}_0 \in \mathbb{R}^n$，一定可将其写成$\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_n$的线性组合，将结果记为$\boldsymbol{u}_0 = \sum_{j=1}^{n} c_j \boldsymbol{q}_j$

可以将$\mathbf{A}$写成$\mathbf{QDQ}^T$，其中$\mathbf{Q}$的各列是$\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_n$，$\mathbf{D}$是由$\lambda_1, \lambda_2, \ldots, \lambda_n$确定的对角矩阵

根据前面结果，有$\mathbf{A}^i = \mathbf{QD}^i\mathbf{Q}^T$

幂迭代法的证明

证明思路：
（1）用$\boldsymbol{q}_1, \boldsymbol{q}_2, \dots, \boldsymbol{q}_n$表示$\mathbf{A}^i\boldsymbol{u}_0$
（2）用$\lambda_1, \lambda_2, \dots$刻画目标数值

用$\boldsymbol{q}_1$表示$\mathbf{A}$的最大特征值对应的单位特征向量，证明当i很大时$\left|\left\langle \frac{\mathbf{A}^i\boldsymbol{u}_0}{\|\mathbf{A}^i\boldsymbol{u}_0\|}, \boldsymbol{q}_1 \right\rangle\right|$趋近于1

用$\boldsymbol{q}_1, \boldsymbol{q}_2, \dots, \boldsymbol{q}_n \in \mathbb{R}^n$表示$\mathbf{A}$的$n$个单位特征向量（对应$\lambda_1 \geq \lambda_2 \geq \cdots \lambda_n$），是一组标准正交向量

因为$\boldsymbol{u}_0 \in \mathbb{R}^n$，一定可将其写成$\boldsymbol{q}_1, \boldsymbol{q}_2, \dots, \boldsymbol{q}_n$的线性组合，将结果记为$\boldsymbol{u}_0 = \sum_{j=1}^{n} c_j \boldsymbol{q}_j$

可以将$\mathbf{A}$写成$\mathbf{QDQ}^T$，其中$\mathbf{Q}$的各列是$\boldsymbol{q}_1, \boldsymbol{q}_2, \dots, \boldsymbol{q}_n$，$\mathbf{D}$是由$\lambda_1, \lambda_2, \dots, \lambda_n$确定的对角矩阵

根据前面结果，有$\mathbf{A}^i = \mathbf{QD}^i\mathbf{Q}^T$

可以计算$\mathbf{A}^i\boldsymbol{u}_0 = \mathbf{QD}^i\mathbf{Q}^T\boldsymbol{u}_0 = \mathbf{QD}^i \begin{bmatrix} c_1 \\ c_2 \\ \dots \\ c_n \end{bmatrix} = \mathbf{Q} \begin{bmatrix} c_1\lambda_1^i \\ c_2\lambda_2^i \\ \dots \\ c_n\lambda_n^i \end{bmatrix} = \sum_{j=1}^{n} \lambda_j^i c_j \boldsymbol{q}_j$

幂迭代法的证明

证明思路：
（1）用$\boldsymbol{q}_1, \boldsymbol{q}_2, \dots, \boldsymbol{q}_n$表示$\mathbf{A}^i\boldsymbol{u}_0$
（2）用$\lambda_1, \lambda_2, \dots$刻画目标数值

用$\boldsymbol{q}_1$表示$\mathbf{A}$的最大特征值对应的单位特征向量，证明当i很大时$\left|\left\langle \frac{\mathbf{A}^i\boldsymbol{u}_0}{\|\mathbf{A}^i\boldsymbol{u}_0\|}, \boldsymbol{q}_1 \right\rangle\right|$趋近于1

用$\boldsymbol{q}_1, \boldsymbol{q}_2, \dots, \boldsymbol{q}_n \in \mathbb{R}^n$表示$\mathbf{A}$的$n$个单位特征向量（对应$\lambda_1 \geq \lambda_2 \geq \cdots \lambda_n$），是一组标准正交向量

因为$\boldsymbol{u}_0 \in \mathbb{R}^n$，一定可将其写成$\boldsymbol{q}_1, \boldsymbol{q}_2, \dots, \boldsymbol{q}_n$的线性组合，将结果记为$\boldsymbol{u}_0 = \sum_{j=1}^n c_j \boldsymbol{q}_j$

可以将$\mathbf{A}$写成$\mathbf{QDQ}^T$，其中$\mathbf{Q}$的各列是$\boldsymbol{q}_1, \boldsymbol{q}_2, \dots, \boldsymbol{q}_n$，$\mathbf{D}$是由$\lambda_1, \lambda_2, \dots, \lambda_n$确定的对角矩阵

根据前面结果，有$\mathbf{A}^i = \mathbf{QD}^i\mathbf{Q}^T$

可以计算$\mathbf{A}^i\boldsymbol{u}_0 = \mathbf{QD}^i\mathbf{Q}^T\boldsymbol{u}_0 = \mathbf{QD}^i \begin{bmatrix} c_1 \\ c_2 \\ \dots \\ c_n \end{bmatrix} = \mathbf{Q} \begin{bmatrix} c_1\lambda_1^i \\ c_2\lambda_2^i \\ \dots \\ c_n\lambda_n^i \end{bmatrix} = \sum_{j=1}^n \lambda_j^i c_j \boldsymbol{q}_j$

可以计算$\|\mathbf{A}^i\boldsymbol{u}_0\| = \sqrt{(\mathbf{A}^i\boldsymbol{u}_0)^T\mathbf{A}^i\boldsymbol{u}_0} = \sqrt{\sum_{j=1}^n \left(\lambda_j^i c_j\right)^2}$ 且$\langle \mathbf{A}^i\boldsymbol{u}_0, \boldsymbol{q}_1 \rangle = \lambda_1^i c_1$

得到$\left\langle \frac{\mathbf{A}^i\boldsymbol{u}_0}{\|\mathbf{A}^i\boldsymbol{u}_0\|}, \boldsymbol{q}_1 \right\rangle = \frac{\lambda_1^i c_1}{\sqrt{\sum_{j=1}^n \left(\lambda_j^i c_j\right)^2}}$

幂迭代法的证明

证明思路：
（1）用$\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_n$表示$\mathbf{A}^i \boldsymbol{u}_0$
（2）用$\lambda_1, \lambda_2, \ldots$刻画目标数值

用$\boldsymbol{q}_1$表示$\mathbf{A}$的最大特征值对应的单位特征向量，证明当i很大时$\left|\left\langle \frac{\mathbf{A}^i \boldsymbol{u}_0}{\|\mathbf{A}^i \boldsymbol{u}_0\|}, \boldsymbol{q}_1 \right\rangle\right|$趋近于1

$$\left|\left\langle \frac{\mathbf{A}^i \boldsymbol{u}_0}{\|\mathbf{A}^i \boldsymbol{u}_0\|}, \boldsymbol{q}_1 \right\rangle\right| = \frac{\lambda_1^i |c_1|}{\sqrt{\sum_{j=1}^n \left(\lambda_j^i c_j\right)^2}} = \frac{\lambda_1^i |c_1|}{\sqrt{\left(\lambda_1^i c_1\right)^2 + \sum_{j\geq 2}^n \left(\lambda_j^i c_j\right)^2}}$$

利用λ_1是最大特征值

幂迭代法的证明

证明思路：

（1）用$\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_n$表示$\mathbf{A}^i\boldsymbol{u}_0$

（2）用$\lambda_1, \lambda_2, \ldots$刻画目标数值

用$\boldsymbol{q}_1$表示$\mathbf{A}$的最大特征值对应的单位特征向量，证明当i很大时$\left|\left\langle \frac{\mathbf{A}^i\boldsymbol{u}_0}{\|\mathbf{A}^i\boldsymbol{u}_0\|}, \boldsymbol{q}_1 \right\rangle\right|$趋近于1

$$\left|\left\langle \frac{\mathbf{A}^i\boldsymbol{u}_0}{\|\mathbf{A}^i\boldsymbol{u}_0\|}, \boldsymbol{q}_1 \right\rangle\right| = \frac{\lambda_1^i|c_1|}{\sqrt{\sum_{j=1}^n \left(\lambda_j^i c_j\right)^2}} = \frac{\lambda_1^i|c_1|}{\sqrt{\left(\lambda_1^i c_1\right)^2 + \sum_{j\geq 2}^n \left(\lambda_j^i c_j\right)^2}}$$

$$\geq \frac{\lambda_1^i|c_1|}{\sqrt{\left(\lambda_1^i c_1\right)^2 + \lambda_2^{2i}\sum_{j\geq 2}^n \left(c_j\right)^2}}$$

放缩成关于λ_1^i与λ_2^i的式子

证明思路：
（1）用$\boldsymbol{q}_1, \boldsymbol{q}_2, \dots, \boldsymbol{q}_n$表示$\mathbf{A}^i\boldsymbol{u}_0$
（2）用$\lambda_1, \lambda_2, \dots$刻画目标数值

幂迭代法的证明

用$\boldsymbol{q}_1$表示$\mathbf{A}$的最大特征值对应的单位特征向量，证明当i很大时$\left|\left\langle \frac{\mathbf{A}^i\boldsymbol{u}_0}{\|\mathbf{A}^i\boldsymbol{u}_0\|}, \boldsymbol{q}_1 \right\rangle\right|$趋近于1

$$\left|\left\langle \frac{\mathbf{A}^i\boldsymbol{u}_0}{\|\mathbf{A}^i\boldsymbol{u}_0\|}, \boldsymbol{q}_1 \right\rangle\right| = \frac{\lambda_1^i|c_1|}{\sqrt{\sum_{j=1}^n (\lambda_j^i c_j)^2}} = \frac{\lambda_1^i|c_1|}{\sqrt{(\lambda_1^i c_1)^2 + \sum_{j\ge 2}^n (\lambda_j^i c_j)^2}}$$

$$\ge \frac{\lambda_1^i|c_1|}{\sqrt{(\lambda_1^i c_1)^2 + \lambda_2^{2i}\sum_{j\ge 2}^n (c_j)^2}} \ge \frac{\lambda_1^i|c_1|}{\sqrt{(\lambda_1^i c_1)^2 + \lambda_2^{2i}}}$$

利用$\|\boldsymbol{u}_0\| = \left\|\sum_{j=1}^n c_j\boldsymbol{q}_j\right\| = \sqrt{\left(\sum_{j=1}^n c_j\boldsymbol{q}_j\right)^T \sum_{j=1}^n c_j\boldsymbol{q}_j} = 1$可得$\sum_{j\ge 2}^n (c_j)^2 \le \sum_{j\ge 1}^n (c_j)^2 = 1$

幂迭代法的证明

证明思路：
（1）用$\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_n$表示$\mathbf{A}^i\boldsymbol{u}_0$
（2）用$\lambda_1, \lambda_2, \ldots$刻画目标数值

用$\boldsymbol{q}_1$表示$\mathbf{A}$的最大特征值对应的单位特征向量，证明当i很大时$\left|\left\langle \frac{\mathbf{A}^i\boldsymbol{u}_0}{\|\mathbf{A}^i\boldsymbol{u}_0\|}, \boldsymbol{q}_1 \right\rangle\right|$趋近于1

$$\left|\left\langle \frac{\mathbf{A}^i\boldsymbol{u}_0}{\|\mathbf{A}^i\boldsymbol{u}_0\|}, \boldsymbol{q}_1 \right\rangle\right| = \frac{\lambda_1^i|c_1|}{\sqrt{\sum_{j=1}^n \left(\lambda_j^i c_j\right)^2}} = \frac{\lambda_1^i|c_1|}{\sqrt{\left(\lambda_1^i c_1\right)^2 + \sum_{j\geq 2}^n \left(\lambda_j^i c_j\right)^2}}$$

$$\geq \frac{\lambda_1^i|c_1|}{\sqrt{\left(\lambda_1^i c_1\right)^2 + \lambda_2^{2i}\sum_{j\geq 2}^n \left(c_j\right)^2}} \geq \frac{\lambda_1^i|c_1|}{\sqrt{\left(\lambda_1^i c_1\right)^2 + \lambda_2^{2i}}} \geq \frac{\lambda_1^i|c_1|}{\lambda_1^i|c_1| + \lambda_2^i}$$

证明思路：

（1）用$\boldsymbol{q}_1, \boldsymbol{q}_2, \dots, \boldsymbol{q}_n$表示$\mathbf{A}^i\boldsymbol{u}_0$

（2）用$\lambda_1, \lambda_2, \dots$刻画目标数值

幂迭代法的证明

用$\boldsymbol{q}_1$表示A的最大特征值对应的单位特征向量，证明当i很大时$\left|\left\langle \frac{\mathbf{A}^i\boldsymbol{u}_0}{\|\mathbf{A}^i\boldsymbol{u}_0\|}, \boldsymbol{q}_1 \right\rangle\right|$趋近于1

$$\left|\left\langle \frac{\mathbf{A}^i\boldsymbol{u}_0}{\|\mathbf{A}^i\boldsymbol{u}_0\|}, \boldsymbol{q}_1 \right\rangle\right| = \frac{\lambda_1^i|c_1|}{\sqrt{\sum_{j=1}^n\left(\lambda_j^i c_j\right)^2}} = \frac{\lambda_1^i|c_1|}{\sqrt{\left(\lambda_1^i c_1\right)^2 + \sum_{j\geq 2}^n\left(\lambda_j^i c_j\right)^2}}$$

$$\geq \frac{\lambda_1^i|c_1|}{\sqrt{\left(\lambda_1^i c_1\right)^2 + \lambda_2^{2i}\sum_{j\geq 2}^n\left(c_j\right)^2}} \geq \frac{\lambda_1^i|c_1|}{\sqrt{\left(\lambda_1^i c_1\right)^2 + \lambda_2^{2i}}} \geq \frac{\lambda_1^i|c_1|}{\lambda_1^i|c_1| + \lambda_2^i}$$

易得：

$$\left|\left\langle \frac{\mathbf{A}^i\boldsymbol{u}_0}{\|\mathbf{A}^i\boldsymbol{u}_0\|}, \boldsymbol{q}_1 \right\rangle\right| \geq \frac{\lambda_1^i|c_1|}{\lambda_1^i|c_1| + \lambda_2^i} = \frac{1}{1 + \frac{\lambda_2^i}{\lambda_1^i|c_1|}} \geq 1 - \frac{1}{|c_1|}\left(\frac{\lambda_2}{\lambda_1}\right)^i$$

幂迭代法的证明

证明思路：
（1）用$\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_n$表示$\mathbf{A}^i \boldsymbol{u}_0$
（2）用$\lambda_1, \lambda_2, \ldots$刻画目标数值

用$\boldsymbol{q}_1$表示A的最大特征值对应的单位特征向量，证明当i很大时$\left|\left\langle \frac{\mathbf{A}^i \boldsymbol{u}_0}{\|\mathbf{A}^i \boldsymbol{u}_0\|}, \boldsymbol{q}_1 \right\rangle\right|$趋近于1

$$\left|\left\langle \frac{\mathbf{A}^i \boldsymbol{u}_0}{\|\mathbf{A}^i \boldsymbol{u}_0\|}, \boldsymbol{q}_1 \right\rangle\right| = \frac{\lambda_1^i |c_1|}{\sqrt{\sum_{j=1}^n \left(\lambda_j^i c_j\right)^2}} = \frac{\lambda_1^i |c_1|}{\sqrt{\left(\lambda_1^i c_1\right)^2 + \sum_{j\geq 2}^n \left(\lambda_j^i c_j\right)^2}}$$

$$\geq \frac{\lambda_1^i |c_1|}{\sqrt{\left(\lambda_1^i c_1\right)^2 + \lambda_2^{2i} \sum_{j\geq 2}^n \left(c_j\right)^2}} \geq \frac{\lambda_1^i |c_1|}{\sqrt{\left(\lambda_1^i c_1\right)^2 + \lambda_2^{2i}}} \geq \frac{\lambda_1^i |c_1|}{\lambda_1^i |c_1| + \lambda_2^i}$$

易得：

$$\left|\left\langle \frac{\mathbf{A}^i \boldsymbol{u}_0}{\|\mathbf{A}^i \boldsymbol{u}_0\|}, \boldsymbol{q}_1 \right\rangle\right| \geq \frac{\lambda_1^i |c_1|}{\lambda_1^i |c_1| + \lambda_2^i} = \frac{1}{1 + \frac{\lambda_2^i}{\lambda_1^i |c_1|}} \geq 1 - \frac{1}{|c_1|}\left(\frac{\lambda_2}{\lambda_1}\right)^i$$

例如，当$\boldsymbol{u}_0$恰好为其它单位特征向量$\boldsymbol{q}_j (j \neq 1)$时，幂迭代法无法收敛到$\boldsymbol{q}_1$

c_1是随机向量$\boldsymbol{u}_0$在$\boldsymbol{q}_1$方向的投影长度（$\boldsymbol{u}_0 = \sum_{j=1}^n c_j \boldsymbol{q}_j$）

可假设$\boldsymbol{u}_0$各元素都分别依标准正态分布随机独立取值（并标准化），由此可得以概率描述的$|c_1|$的下界

若有$\frac{\lambda_2}{\lambda_1} < 1$，可知随$i$增大，$\left|\left\langle \frac{\mathbf{A}^i \boldsymbol{u}_0}{\|\mathbf{A}^i \boldsymbol{u}_0\|}, \boldsymbol{q}_1 \right\rangle\right|$的下界趋近于1。此外，$\left|\left\langle \frac{\mathbf{A}^i \boldsymbol{u}_0}{\|\mathbf{A}^i \boldsymbol{u}_0\|}, \boldsymbol{q}_1 \right\rangle\right| = |\cos 夹角| \leq 1$

幂迭代法的证明

用$\boldsymbol{q}_1$表示A的最大特征值对应的单位特征向量，证明当i很大时$\left|\left\langle \frac{\mathbf{A}^i\boldsymbol{u}_0}{\|\mathbf{A}^i\boldsymbol{u}_0\|}, \boldsymbol{q}_1 \right\rangle\right|$趋近于1

$$\left|\left\langle \frac{\mathbf{A}^i\boldsymbol{u}_0}{\|\mathbf{A}^i\boldsymbol{u}_0\|}, \boldsymbol{q}_1 \right\rangle\right| = \frac{\lambda_1^i|c_1|}{\sqrt{\sum_{j=1}^n\left(\lambda_j^i c_j\right)^2}} = \frac{\lambda_1^i|c_1|}{\sqrt{\left(\lambda_1^i c_1\right)^2 + \sum_{j\geq 2}^n\left(\lambda_j^i c_j\right)^2}}$$

$$\geq \frac{\lambda_1^i|c_1|}{\sqrt{\left(\lambda_1^i c_1\right)^2 + \lambda_2^{2i}\sum_{j\geq 2}^n\left(c_j\right)^2}} \geq \frac{\lambda_1^i|c_1|}{\sqrt{\left(\lambda_1^i c_1\right)^2 + \lambda_2^{2i}}} \geq \frac{\lambda_1^i|c_1|}{\lambda_1^i|c_1| + \lambda_2^i}$$

易得:

$$\left|\left\langle \frac{\mathbf{A}^i\boldsymbol{u}_0}{\|\mathbf{A}^i\boldsymbol{u}_0\|}, \boldsymbol{q}_1 \right\rangle\right| \geq \frac{\lambda_1^i|c_1|}{\lambda_1^i|c_1| + \lambda_2^i} = \frac{1}{1 + \frac{\lambda_2^i}{\lambda_1^i|c_1|}} \geq 1 - \frac{1}{|c_1|}\left(\frac{\lambda_2}{\lambda_1}\right)^i$$

实际计算中，若$\frac{\lambda_2}{\lambda_1}$接近1，算法收敛速度很慢

c_1是随机向量$\boldsymbol{u}_0$在$\boldsymbol{q}_1$方向的投影长度（$\boldsymbol{u}_0 = \sum_{j=1}^n c_j\boldsymbol{q}_j$）

可假设$\boldsymbol{u}_0$各元素都分别依标准正态分布随机独立取值（并标准化），由此可得以概率描述的$|c_1|$的下界

若有$\frac{\lambda_2}{\lambda_1} < 1$，可知随$i$增大，$\left|\left\langle \frac{\mathbf{A}^i\boldsymbol{u}_0}{\|\mathbf{A}^i\boldsymbol{u}_0\|}, \boldsymbol{q}_1 \right\rangle\right|$的下界趋近于1。此外，$\left|\left\langle \frac{\mathbf{A}^i\boldsymbol{u}_0}{\|\mathbf{A}^i\boldsymbol{u}_0\|}, \boldsymbol{q}_1 \right\rangle\right| = |\cos\text{夹角}| \leq 1$

幂迭代法

计算矩阵最大特征值对应的单位特征向量（即第一主成分）：

幂迭代法

0 随机选择一个单位向量$\boldsymbol{u}_0 \in \mathbb{R}^n$; 计算$\mathbf{A} = \mathbf{X}^T\mathbf{X}$

1 for i=1, 2, ...

2 计算$\boldsymbol{u}_i = \mathbf{A}\boldsymbol{u}_{i-1}$

3 检查终止条件：若$\frac{\boldsymbol{u}_i}{\|\boldsymbol{u}_i\|} \approx \frac{\boldsymbol{u}_{i-1}}{\|\boldsymbol{u}_{i-1}\|}$，结束算法、输出$\frac{\boldsymbol{u}_i}{\|\boldsymbol{u}_i\|}$

在实际应用中，为加快计算速度、减少矩阵乘以向量的次数，经常先计算$\mathbf{A}^1, \mathbf{A}^2, \mathbf{A}^4, \mathbf{A}^8, \ldots$，然后得到$\mathbf{A}^1\boldsymbol{u}_0, \mathbf{A}^2\boldsymbol{u}_0, \mathbf{A}^4\boldsymbol{u}_0, \mathbf{A}^8\boldsymbol{u}_0, \ldots$。

幂迭代法

计算矩阵最大特征值对应的单位特征向量（即第一主成分）：

幂迭代法

0 随机选择一个单位向量$\boldsymbol{u}_0 \in \mathbb{R}^n$; 计算$\mathbf{A} = \mathbf{X}^T\mathbf{X}$

1 for i=1, 2, ...

2 计算$\boldsymbol{u}_i = \boldsymbol{A}\boldsymbol{u}_{i-1}$

3 检查终止条件：若$\frac{\boldsymbol{u}_i}{\|\boldsymbol{u}_i\|} \approx \frac{\boldsymbol{u}_{i-1}}{\|\boldsymbol{u}_{i-1}\|}$，结束算法、输出$\frac{\boldsymbol{u}_i}{\|\boldsymbol{u}_i\|}$

如何计算前k个主成分?

幂迭代法

计算前k个主成分：

- 用幂迭代法计算第一主成分，记为$\boldsymbol{v}_1$
- 处理数据$\mathbf{X}$：

$$\begin{bmatrix} - & \mathbf{x}_1^T & - \\ - & \mathbf{x}_2^T & - \\ & \vdots & \\ - & \mathbf{x}_m^T & - \end{bmatrix} \mapsto \begin{bmatrix} - & (\mathbf{x}_1^T - \langle \mathbf{x}_1, \mathbf{v}_1 \rangle \mathbf{v}_1^T) & - \\ - & (\mathbf{x}_2^T - \langle \mathbf{x}_2, \mathbf{v}_1 \rangle \mathbf{v}_1^T) & - \\ & \vdots & \\ - & (\mathbf{x}_m^T - \langle \mathbf{x}_m, \mathbf{v}_1 \rangle \mathbf{v}_1^T) & - \end{bmatrix}$$

剔除数据中已经由第一主成分表征的数据间的差异

- 对新的$\mathbf{X}^T\mathbf{X}$计算前$k-1$个主成分

幂迭代法

计算前k个主成分：

- 用幂迭代法计算第一主成分，记为$\boldsymbol{v}_1$
- 处理数据$\mathbf{X}$：

$$\begin{bmatrix} - & \mathbf{x}_1^T & - \\ - & \mathbf{x}_2^T & - \\ & \vdots & \\ - & \mathbf{x}_m^T & - \end{bmatrix} \mapsto \begin{bmatrix} - & (\mathbf{x}_1^T - \langle \mathbf{x}_1, \mathbf{v}_1 \rangle \mathbf{v}_1^T) & - \\ - & (\mathbf{x}_2^T - \langle \mathbf{x}_2, \mathbf{v}_1 \rangle \mathbf{v}_1^T) & - \\ & \vdots & \\ - & (\mathbf{x}_m^T - \langle \mathbf{x}_m, \mathbf{v}_1 \rangle \mathbf{v}_1^T) & - \end{bmatrix}$$

剔除数据中已经由第一主成分表征的数据间的差异

- 对新的$\mathbf{X}^T\mathbf{X}$计算前$k-1$个主成分

如何理解新$\mathbf{X}^T\mathbf{X}$的前$k-1$个主成分即为原$\mathbf{X}^T\mathbf{X}$的第二至第k个主成分？

幂迭代法

计算前k个主成分：

- 用幂迭代法计算第一主成分，记为$\boldsymbol{v}_1$
- 处理数据$\mathbf{X}$：

$$\begin{bmatrix} — & \mathbf{x}_1^T & — \\ — & \mathbf{x}_2^T & — \\ & \vdots & \\ — & \mathbf{x}_m^T & — \end{bmatrix} \mapsto \begin{bmatrix} — & (\mathbf{x}_1^T - \langle \mathbf{x}_1, \mathbf{v}_1 \rangle \mathbf{v}_1^T) & — \\ — & (\mathbf{x}_2^T - \langle \mathbf{x}_2, \mathbf{v}_1 \rangle \mathbf{v}_1^T) & — \\ & \vdots & \\ — & (\mathbf{x}_m^T - \langle \mathbf{x}_m, \mathbf{v}_1 \rangle \mathbf{v}_1^T) & — \end{bmatrix}$$

剔除数据中已经由第一主成分表征的数据间的差异

- 对新的$\mathbf{X}^T\mathbf{X}$计算前$k-1$个主成分

如何理解新$\mathbf{X}^T\mathbf{X}$的前$k-1$个主成分即为原$\mathbf{X}^T\mathbf{X}$的第二至第k个主成分?

将$\mathbf{X}^T\mathbf{X}$的所有主成分记为$\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$，它们构成标准正交向量组

对数据$\boldsymbol{x}_i$，由于$\boldsymbol{x}_i \in \mathbb{R}^n$，可将其写为$n$维空间的$n$个标准正交向量的线性组合：

$$\boldsymbol{x}_i = \sum_{j=1}^{n} \langle \boldsymbol{x}_i, \boldsymbol{v}_j \rangle \boldsymbol{v}_j = \langle \boldsymbol{x}_i, \boldsymbol{v}_1 \rangle \boldsymbol{v}_1 + \langle \boldsymbol{x}_i, \boldsymbol{v}_2 \rangle \boldsymbol{v}_2 + \cdots + \langle \boldsymbol{x}_i, \boldsymbol{v}_n \rangle \boldsymbol{v}_n$$

幂迭代法

计算前k个主成分：

- 用幂迭代法计算第一主成分，记为$\boldsymbol{v}_1$
- 处理数据$\mathbf{X}$：

$$\begin{bmatrix} — & \mathbf{x}_1^T & — \\ — & \mathbf{x}_2^T & — \\ & \vdots & \\ — & \mathbf{x}_m^T & — \end{bmatrix} \mapsto \begin{bmatrix} — & (\mathbf{x}_1^T - \langle \mathbf{x}_1, \mathbf{v}_1 \rangle \mathbf{v}_1^T) & — \\ — & (\mathbf{x}_2^T - \langle \mathbf{x}_2, \mathbf{v}_1 \rangle \mathbf{v}_1^T) & — \\ & \vdots & \\ — & (\mathbf{x}_m^T - \langle \mathbf{x}_m, \mathbf{v}_1 \rangle \mathbf{v}_1^T) & — \end{bmatrix}$$

剔除数据中已经由第一主成分表征的数据间的差异

- 对新的$\mathbf{X}^T\mathbf{X}$计算前$k-1$个主成分

如何理解新$\mathbf{X}^T\mathbf{X}$的前$k-1$个主成分即为原$\mathbf{X}^T\mathbf{X}$的第二至第k个主成分?

将$\mathbf{X}^T\mathbf{X}$的所有主成分记为$\boldsymbol{v}_1, \dots, \boldsymbol{v}_n$，它们构成标准正交向量组

对数据$\boldsymbol{x}_i$，由于$\boldsymbol{x}_i \in \mathbb{R}^n$，可将其写为$n$维空间的$n$个标准正交向量的线性组合：

$$\boldsymbol{x}_i = \sum_{j=1}^{n} \langle \boldsymbol{x}_i, \boldsymbol{v}_j \rangle \boldsymbol{v}_j = \langle \boldsymbol{x}_i, \boldsymbol{v}_1 \rangle \boldsymbol{v}_1 + \langle \boldsymbol{x}_i, \boldsymbol{v}_2 \rangle \boldsymbol{v}_2 + \cdots + \langle \boldsymbol{x}_i, \boldsymbol{v}_n \rangle \boldsymbol{v}_n$$

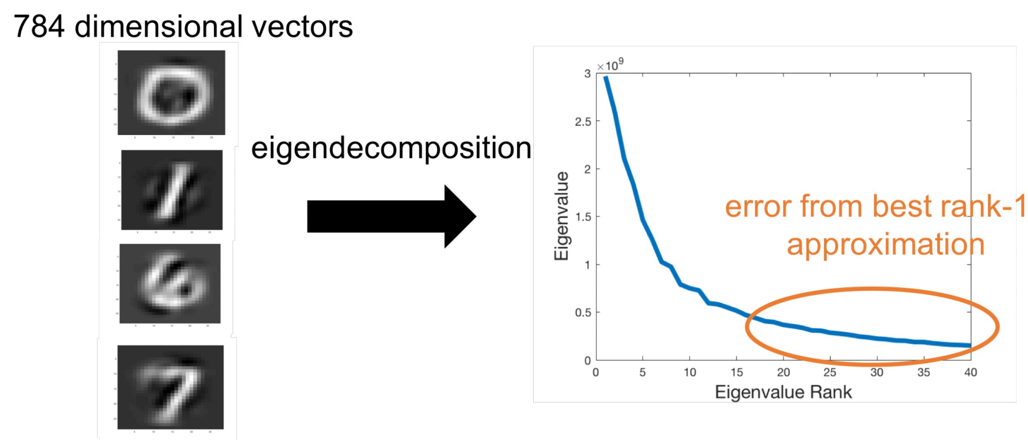
k的选择

在应用中如何选择k？即找到多少个主成分？

- 若目的是数据可视化：如$k=2, k=3$
- 若目的是数据压缩：需考虑近似精度的要求，一般要确保剩余的$n-k$个特征值都比较小

如考虑确保 $\dfrac{\sum_{i=1}^{k}\lambda_i}{\sum_{i=1}^{n}\lambda_i} \geq (1-\epsilon)$

回顾：近似误差为 $\frac{1}{m}\sum_{i=k+1}^{n}\lambda_i$



本讲小结

- 主成分的代数性质
- 求解主成分的方法

主要参考资料

Tim Roughgarden and Gregory Valiant <CS 168 - The Modern Algorithmic Toolbox> Lecture Notes

Cameron Musco <COMPSCI 514 - Algorithms for Data Science> Slides

谢谢！