

第四章 矩阵分解

第 9 讲 具有特殊结构和性质的矩阵

黄定江

DaSE @ ECNU

djhuang@dase.ecnu.edu.cn

1 9.1 数学中常见的具有特殊结构的矩阵

2 9.2 数据科学中常见的矩阵

1 9.1 数学中常见的具有特殊结构的矩阵

2 9.2 数据科学中常见的矩阵

9.1.1 方阵

方阵

若矩阵 $\mathbf{A} \in \mathbb{R}^{n \times n}$ ，即行数与列数都等于 n 的叫做 n 阶方阵。

方阵是非常特殊的矩阵。在线性代数中很多讨论都是针对方阵进行的。

与方阵相关的内容

- 只有方阵才可以计算行列式。
- 只有方阵才可能有逆矩阵。且方阵有逆矩阵当且仅当方阵满秩。
- 只有方阵才有伴随矩阵。
- 只有方阵才有特征值，特征向量等概念。

9.1.2 对称矩阵和正半定矩阵

对称矩阵

以主对角线为对称轴，对应各元素相等的矩阵。即矩阵 \mathbf{A} 是对称矩阵，当且仅当

$$\mathbf{A}^T = \mathbf{A}$$

正半定矩阵

如果对称矩阵 $\mathbf{A} \in \mathbb{R}^{n \times n}$ 且对任意 $\mathbf{x} \in \mathbb{R}^n$ 有

$$\mathbf{x}\mathbf{A}^T\mathbf{x} \geq 0$$

则称 \mathbf{A} 为正半定矩阵，记为 $\mathbf{A} \succeq 0$ 。进一步，若对任意的 $0 \neq \mathbf{x} \in \mathbb{R}^n$ 有

$$\mathbf{x}\mathbf{A}^T\mathbf{x} > 0$$

则称 \mathbf{A} 为正定矩阵，记为 $\mathbf{A} \succ 0$ 。

9.1.3 对角矩阵

对角矩阵

非对角元素都为零元素的方阵叫做对角矩阵。

$n \times n$ 的对角矩阵可以记为 $\mathbf{A} = \text{diag}(\mathbf{a}) = \text{diag}(a_1, a_2, \dots, a_n)$ 。

这里 \mathbf{a} 是 n 维向量，包含了矩阵 \mathbf{A} 的全部对角元素。

$$\mathbf{A} = \text{diag}(a_1, a_2, \dots, a_n) = \begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_n \end{pmatrix}$$

对角矩阵的幂和逆

容易验证，对于对角矩阵：

$$\mathbf{A}^k = \begin{pmatrix} a_1^k & & \\ & \ddots & \\ & & a_n^k \end{pmatrix}$$

如果 \mathbf{A} 是可逆矩阵，即矩阵 \mathbf{A} 的对角元都不为零，我们有：

$$\mathbf{A}^{-1} = \begin{pmatrix} \frac{1}{a_1} & & \\ & \ddots & \\ & & \frac{1}{a_n} \end{pmatrix}$$

9.1.4 三角矩阵

三角矩阵

三角矩阵是对角元下方或对角元上方全是零的方阵。

定义 1

若矩阵 \mathbf{A} 的所有元素满足 $i > j$, $a_{ij} = 0$, 则称 \mathbf{A} 为上三角矩阵

$$\mathbf{A} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ & \ddots & \vdots \\ & & a_{nn} \end{bmatrix}$$

定义 2

若矩阵 \mathbf{A} 的所有元素满足 $i < j$, $a_{ij} = 0$, 则称 \mathbf{A} 为下三角矩阵

$$\mathbf{A} = \begin{bmatrix} a_{11} & & \\ \vdots & \ddots & \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}$$

三角矩阵性质

- 三角矩阵主对角线上元素均非零 \iff 三角矩阵可逆
- 上三角矩阵的乘积还是上三角矩阵
- 若上三角矩阵可逆则其逆矩阵也是上三角矩阵
- 下三角矩阵的乘积还是下三角矩阵
- 若下三角矩阵可逆则其逆矩阵也是下三角矩阵
- 设矩阵 $\mathbf{A} \in \mathbb{R}^{n \times n}$ 为三角矩阵, $k \in \mathbb{Z}$, 那么矩阵 \mathbf{A}^k 主对角线上的元素 $(\mathbf{A}^k)_{ii} = (\mathbf{A}_{ii})^k, i = 1, 2, \dots, n$

9.1.5 正交矩阵

正交矩阵

正交矩阵 (orthogonal matrix) 指行向量和列向量是分别标准正交的方阵，即

$$\mathbf{A}^T \mathbf{A} = \mathbf{A} \mathbf{A}^T = \mathbf{I}$$

从定义上可以看出

$$\mathbf{A}^{-1} = \mathbf{A}^T$$

正交矩阵求逆，只需对矩阵转置即求得矩阵的逆。

正交矩阵的性质

正交性

设 $\mathbf{A} = [a_1, \dots, a_n]$, 并且 \mathbf{A} 是一个正交矩阵, 那么

$$\mathbf{a}_i^T \mathbf{a}_j = \begin{cases} 1 & \text{如果 } i = j \\ 0 & \text{如果 } i \neq j \end{cases}$$

和范数有关的性质

如果矩阵 $\mathbf{U} \in \mathbb{R}^{m \times m}$, $\mathbf{V} \in \mathbb{R}^{n \times n}$ 是正交矩阵, $\mathbf{M} \in \mathbb{R}^{m \times n}$, $\mathbf{x} \in \mathbb{R}^m$

- $\|\mathbf{U}\|_2 = 1$, $\|\mathbf{U}\|_F = \sqrt{m}$
- $\|\mathbf{U}\mathbf{x}\|_2 = \|\mathbf{x}\|_2$, $\|\mathbf{U}\mathbf{x}\|_F = \|\mathbf{x}\|_F$
- $\|\mathbf{UMV}\|_2 = \|\mathbf{M}\|_2$, $\|\mathbf{UMV}\|_F = \|\mathbf{M}\|_F$

9.1.6 Dyads (并向量或单纯矩阵或秩 1 矩阵)

定义 3

矩阵 $\mathbf{A} \in \mathbb{R}^{m \times n}$ 如果具有如下形式:

$$\mathbf{A} = \mathbf{u}\mathbf{v}^T$$

其中向量 $\mathbf{u} \in \mathbb{R}^m, \mathbf{v} \in \mathbb{R}^n$, 则称其为 *dyad*, 也称为并向量或单纯矩阵。如果 \mathbf{u} 和 \mathbf{v} 不为零, 则我们称其为秩 1 矩阵。

如果 \mathbf{v}, \mathbf{u} 具有相同维度, 则 dyad $\mathbf{A} = \mathbf{u}\mathbf{v}^T$ 就是一个方阵。

dyad 作为线性映射

一个 dyad $\mathbf{A} = \mathbf{u}\mathbf{v}^T$ 对于输入向量 $\mathbf{x} \in \mathbb{R}^n$ 有如下作用:

$$\mathbf{A}\mathbf{x} = (\mathbf{u}\mathbf{v}^T)\mathbf{x} = (\mathbf{v}^T\mathbf{x})\mathbf{u}$$

- 因为 $\mathbf{A}_{ij} = \mathbf{u}_i\mathbf{v}_j$; 所以每一行（列）是对应的列（行）的缩放，其中“缩放”由向量 \mathbf{u} (\mathbf{v}) 给出。
- 对于一个给定的 $\mathbf{A} = \mathbf{u}\mathbf{v}^T$ ，由对应的线性映射 $\mathbf{x} \rightarrow \mathbf{A}\mathbf{x}$ 可知，无论输入 \mathbf{x} 是什么，输出向量方向始终与 \mathbf{u} 相同。因此，输出向量是 \mathbf{u} 的一个缩放，并且缩放量为 $\mathbf{v}^T\mathbf{x}$ ，故取决于向量 \mathbf{v} 。
- 对于一个 dyad $\mathbf{A} = \mathbf{u}\mathbf{v}^T$ ，如果 \mathbf{u} 和 \mathbf{v} 不为零，则其秩为 1，因为它的像空间都是由 \mathbf{u} 生成的，因此把 $\mathbf{A} = \mathbf{u}\mathbf{v}^T$ 称为秩 1 矩阵。

dyad 的特征值和特征向量

方的 dyad ($m = n$) 有唯一的非零特征值 $\lambda = \mathbf{v}^T \mathbf{u}$ 与对应的特征向量 \mathbf{u} 。

dyad 的正规化

对于一个 dyad $\mathbf{A} = \mathbf{u}\mathbf{v}^T$ ，我们可以利用欧几里得范数单位化 \mathbf{u} 和 \mathbf{v} ，并且用一个系数来衡量 dyad 的大小，以此来标准化 dyad，也即任何 dyad 都可以写成如下正规化的形式：

$$\mathbf{A} = \mathbf{u}\mathbf{v}^T = (\|\mathbf{u}\|_2 \cdot \|\mathbf{v}\|_2) \frac{\mathbf{u}}{\|\mathbf{u}\|_2} \frac{\mathbf{v}^T}{\|\mathbf{v}\|_2} = \sigma \tilde{\mathbf{u}} \tilde{\mathbf{v}}^T$$

其中 $\sigma > 0$ ，并且 $\|\tilde{\mathbf{u}}\| = \|\tilde{\mathbf{v}}\| = 1$ 。

9.1.7 分块矩阵

任何矩阵都可以分成具有相容维的若干块或子矩阵的分块形式:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}$$

当 \mathbf{A} 是方阵, 并且 $\mathbf{A}_{12} = \mathbf{O}$, $\mathbf{A}_{21} = \mathbf{O}$, 那么称 \mathbf{A} 为块对角矩阵:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{O} \\ \mathbf{O} & \mathbf{A}_{22} \end{bmatrix}$$

若 \mathbf{A} 为块对角矩阵, 用 $\lambda(\mathbf{A})$ 表示 \mathbf{A} 的特征值集合, 显然, 它是 \mathbf{A}_{11} 和 \mathbf{A}_{22} 特征值集合的并集。

$$\mathbf{A} \text{ 为块对角矩阵} \implies \lambda(\mathbf{A}) = \lambda(\mathbf{A}_{11}) \cup \lambda(\mathbf{A}_{22})$$

一个块对角矩阵是可逆的, 当且仅当它的每个对角块是可逆的, 并且

$$\begin{bmatrix} \mathbf{A}_{11} & \mathbf{O} \\ \mathbf{O} & \mathbf{A}_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}_{11}^{-1} & \mathbf{O} \\ \mathbf{O} & \mathbf{A}_{22}^{-1} \end{bmatrix}$$

分块三角矩阵

分块方阵 \mathbf{A} , 如果 $\mathbf{A}_{21} = \mathbf{O}$, 称之为分块上三角矩阵; 如果 $\mathbf{A}_{12} = \mathbf{O}$, 称之为分块下三角矩阵。

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{O} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \quad \text{分块下三角矩阵}$$

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{O} & \mathbf{A}_{22} \end{bmatrix} \quad \text{分块上三角矩阵}$$

若 \mathbf{A} 为分块三角矩阵, 用 $\lambda(\mathbf{A})$ 表示 \mathbf{A} 的特征值集合, 同样有:

$$\mathbf{A} \text{ 为分块 (上或下) 三角矩阵} \implies \lambda(\mathbf{A}) = \lambda(\mathbf{A}_{11}) \cup \lambda(\mathbf{A}_{22})$$

分块三角矩阵的逆

非退化的分块三角矩阵的逆可以表示为:

$$\begin{bmatrix} \mathbf{A}_{11} & \mathbf{O} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}_{11}^{-1} & \mathbf{O} \\ -\mathbf{A}_{22}^{-1} \mathbf{A}_{21} \mathbf{A}_{11}^{-1} & \mathbf{A}_{22}^{-1} \end{bmatrix}$$
$$\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{O} & \mathbf{A}_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}_{11}^{-1} & -\mathbf{A}_{11}^{-1} \mathbf{A}_{12} \mathbf{A}_{22}^{-1} \\ \mathbf{O} & \mathbf{A}_{22}^{-1} \end{bmatrix}$$

非退化分块矩阵的逆

考虑非退化分块矩阵

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}$$

其中 \mathbf{A}_{11} 和 \mathbf{A}_{22} 是方阵并且可逆。令 $\mathbf{S}_1 = \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}$, $\mathbf{S}_2 = \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}$, 则有

$$\begin{aligned} \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}^{-1} &= \begin{pmatrix} \mathbf{S}_1^{-1} & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{S}_2^{-1} \\ -\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{S}_1^{-1} & \mathbf{S}_2^{-1} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{S}_1^{-1} & -\mathbf{S}_1^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ -\mathbf{S}_2^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{S}_2^{-1} \end{pmatrix} \end{aligned}$$

矩阵求逆引理

假设 \mathbf{A}_{11} 和 \mathbf{A}_{22} 分别是 $n_{\mathbf{A}_{11}} \times n_{\mathbf{A}_{11}}$ 和 $n_{\mathbf{A}_{22}} \times n_{\mathbf{A}_{22}}$ 阶方阵并且可逆, \mathbf{A}_{12} 和 \mathbf{A}_{21} 分别是 $n_{\mathbf{A}_{11}} \times n_{\mathbf{A}_{22}}$ 和 $n_{\mathbf{A}_{22}} \times n_{\mathbf{A}_{11}}$ 阶矩阵, 则如下等式成立:

$$(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})^{-1} = \mathbf{A}_{11}^{-1} + \mathbf{A}_{11}^{-1}\mathbf{A}_{12}(\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} \quad (1)$$

上式也即 \mathbf{S}_1 的逆的表达式。类似的也可得 \mathbf{S}_2 的逆的表达式。

Woodbury 公式

假设 \mathbf{E} 和 \mathbf{G} 分别是 $n_{\mathbf{E}} \times n_{\mathbf{E}}$ 和 $n_{\mathbf{G}} \times n_{\mathbf{G}}$ 阶方阵并且可逆, \mathbf{F} 和 \mathbf{H} 分别是 $n_{\mathbf{E}} \times n_{\mathbf{G}}$ 和 $n_{\mathbf{G}} \times n_{\mathbf{E}}$ 阶矩阵, 则如下等式成立:

$$(\mathbf{E} + \mathbf{FGH})^{-1} = \mathbf{E}^{-1} - \mathbf{E}^{-1}\mathbf{F}(\mathbf{G}^{-1} + \mathbf{HE}^{-1}\mathbf{F})^{-1}\mathbf{HE}^{-1}$$

9.1.8 秩 1 扰动

Sherman-Morrison 公式

设矩阵 $\mathbf{A} \in \mathbb{R}^{n \times n}$, $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ 如果我们令矩阵求逆引理公式(1)中

$$\mathbf{A}_{11} = \mathbf{A}, \mathbf{A}_{12} = \mathbf{u}, \mathbf{A}_{22} = -1, \mathbf{A}_{21} = \mathbf{u}^T$$

则我们可以得到如下等式:

$$(\mathbf{A} + \mathbf{u}\mathbf{v}^T)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{u}\mathbf{v}^T\mathbf{A}^{-1}}{1 + \mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}}$$

这个式子让我们能够计算矩阵 \mathbf{A} 的秩 1 扰动的逆, 并且计算仅仅依赖于 \mathbf{A} 的逆。

Sherman-Morrison 公式可用于拟牛顿迭代法的 BFGS 公式的推导。

秩 1 扰动

一个更有趣的性质是矩阵 \mathbf{A} 的秩 1 扰动和原矩阵的秩的变化不超过 1。这个事实不仅仅对方阵成立，对一般的矩阵也成立。

定理 1

令 $\mathbf{A} \in \mathbb{R}^{n \times m}$, $\mathbf{q} \in \mathbb{R}^m$, $\mathbf{p} \in \mathbb{R}^n$ 则有

$$|\text{rank}(\mathbf{A}) - \text{rank}(\mathbf{A} + \mathbf{q}\mathbf{p}^T)| \leq 1$$

证明可利用线性代数基本定理来实现。在 Matrix Completion 类问题中有重要应用。

① 9.1 数学中常见的具有特殊结构的矩阵

② 9.2 数据科学中常见的矩阵

9.2.1 图和矩阵引入：从网络到图

网络：现实世界各种对象之间关系和相互作用过程的表达

- 社交网络、通信网络
- 蛋白质相互作用网络
- 病毒传播网络
- 单词共现网络、图像网络
- ...

图（点和边的集合）：网络表达的结构化和抽象化

- 无向图
- 有向图
- 加权图
- 完全图：二分图、子图
- ...

现实世界的对象可以看成网络和图中的“点”，关系或相互作用可以通过点与点相连的“边”以及给边赋予“权重”和“方向”来进一步表达关系的“远近亲疏、重要程度和因果关系”。

图和矩阵引入：网络和图分析的主要任务

从数据科学的角度看，图分析任务包括：

- 节点分类
- 链接预测
- 聚类
- 降维或可视化
- ...

实现任务的相应模型：

- 随机游走
- 相似性方法、最大似然和概率模型
- 属性基方法
- 嵌入方法
- ...

图和矩阵引入：谱聚类

例 1

设有数据集

$$\mathbb{X} = \{(1, 3), (1, 4), (2, 4), (3, 2), (2, 1), (3, 1)\}$$

如右图 (a) 所示。我们希望能够通过某一种方式将这 6 个点自动地分成两类，如右图 (b) 所示。我们可以将每一个顶点和他距离最近的 3 个顶点进行连接得到图 (c)。从而将问题转化为研究图上顶点聚类。

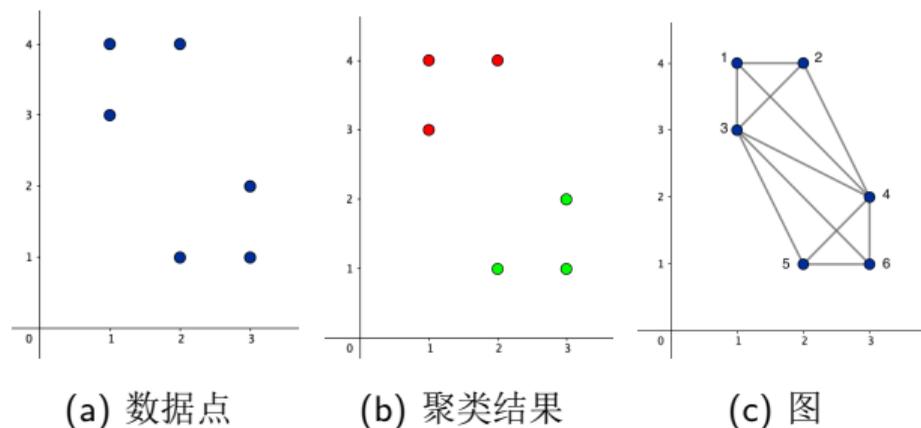


图 1: 谱聚类

图和矩阵引入：谱聚类的基本思想

谱聚类的基本思想：

- 把所有的数据看做空间中的点，这些点之间可以用边连接起来，形成一个图
- 距离较远的两个点之间的边权重值较低，而距离较近的两个点之间的边权重值较高
- 通过对所有数据点组成的图进行切图，让切图后不同的子图间边权重和尽可能的低，而子图内的边权重和尽可能的高，从而达到聚类的目的

基本问题：点与点之间有什么关系、如何表示？点和边之间又有什么关系，如何表示？涉及图的数据结构。

图和矩阵引入：图的数据结构

单个点、边之间的关系描述：

- 关联 (incidence)：点与边的关系
- 邻接 (adjacent)：点与点的关系
- 度 (degree)：相邻节点的数量
- 路径、回路、连通
- ...

很多点、边的关系表示：

- 关联矩阵
- 邻接矩阵
- 度矩阵
- 拉普拉斯矩阵
- ...

大部分图分析任务中的模型可以直接定义在原始图的邻接矩阵或由邻接矩阵和度矩阵导出的拉普拉斯矩阵上。

9.2.2 图的基本概念回顾

图是由一些节点和连接这些节点的边组成的离散结构。

定义 4

一张图 G 是一个二元组, $G = (\mathbb{V}, \mathbb{E})$ 是由节点集合 \mathbb{V} 和边集 \mathbb{E} 组成的。其中 \mathbb{E} 中的元素是一个二元对 $\{x, y\}$, 其中 $x, y \in \mathbb{V}$ 。

- 对于无向图而言, $\{x, y\}$ 是无序对, $\{x, y\}$ 和 $\{y, x\}$ 是 \mathbb{E} 中的同一个元素, 表示点 x 和 y 有一条边相连。
- 对于有向图, $\{x, y\}$ 表示有一条由 x 指向 y 的有向边, 和 $\{y, x\}$ 是 \mathbb{E} 中不同的元素。
- 如果图 $G = (\mathbb{V}, \mathbb{E})$ 中每一条边 $\{v_i, v_j\}$ 都被赋予一个权重 w_{ij} , 则称这样的图为加权图或赋权图。

在实际问题中, 权重 w_{ij} 通常是具有某种含义的数值, 比如在聚类中是衡量节点远近关系的距离度量数值。本节讨论的图都是简单图。

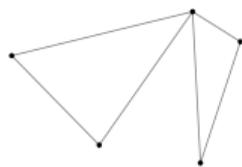
图的基本概念

定义 5

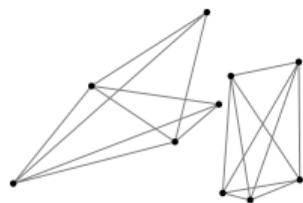
设 n 为正整数, $G = (\mathbb{V}, \mathbb{E})$ 为一简单图。我们称图中的一条长度为 n 的通路为 n 条边 e_1, e_2, \dots, e_n 的序列。其中 $e_1 = \{v_0, v_1\}$, $e_2 = \{v_1, v_2\}$, \dots , $e_n = \{v_{n-1}, v_n\}$ 我们可以用顶点序列 v_0, v_1, \dots, v_n 来表示这条通路。如果 $v_0 = v_n$ 我们则称这条通路为一条回路。如果通路 v_0, v_1, \dots, v_n 中 v_1, v_2, \dots, v_n 是互异的, 那么我们称这条通路为简单通路。

定义 6

设 $G = (\mathbb{V}, \mathbb{E})$ 为一简单图。如果 $\forall u_1, u_2 \in \mathbb{V}$ 都存在一条通路 v_0, v_1, \dots, v_n 使得 $v_0 = u_1, v_n = u_2$ 。我们则称图 G 是连通的。



(a) 连通图



(b) 非连通图

图 2: 连通图和非连通图

图的基本概念

定义 7

设图 $G = (\mathbb{V}_G, \mathbb{E}_G)$, $H = (\mathbb{V}_H, \mathbb{E}_H)$ 如果 $\mathbb{V}_H \subseteq \mathbb{V}_G$ 且 $\mathbb{E}_H \subseteq \mathbb{E}_G$, 那么我们称图 H 为 G 的子图。

定义 8

设图 $H = (\mathbb{V}_H, \mathbb{E}_H)$ 是图 $G = (\mathbb{V}_G, \mathbb{E}_G)$ 的子图。如果 $\forall v \in \mathbb{V}_H$, $u \in \mathbb{V}_G / \mathbb{V}_H$ 都满足 $\{v, u\} \notin \mathbb{E}_G$ 则称 H 是图 G 的一个连通分量。

定理 2

如果图 $G = (\mathbb{V}, \mathbb{E})$ 是一连通图。那么图 G 有唯一的连通分量为自身。

本节如无特殊说明, 一般讨论的是连通图, 只有一个连通分量。



图 3: 子图

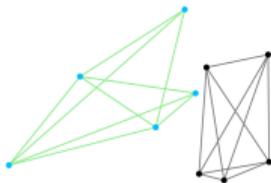


图 4: 连通分量

9.2.3 有向图相关的矩阵：关联矩阵

定义 9

设有向图 $G = \langle \mathbb{V}, \mathbb{E} \rangle$ ，所有顶点的排列为 v_1, v_2, \dots, v_m ，其中 $m = |\mathbb{V}|$ ，所有边的排列为 e_1, e_2, \dots, e_n ，其中 $n = |\mathbb{E}|$ ，用 b_{ij} 表示顶点 v_i 与边 e_j 关联的次数，其中 b_{ij} 定义为

$$b_{ij} = \begin{cases} 1 & v_i \text{ 是边 } e_j \text{ 的起点} \\ -1 & v_i \text{ 是边 } e_j \text{ 的终点} \\ 0 & \text{其他} \end{cases}$$

则称所得的矩阵 $\mathbf{B} = (b_{ij})_{m \times n}$ 为有向图 G 的关联矩阵。

网络流和关联矩阵

例 2

物品、交通、电荷和信息等网络可以表示成一个由 m 个顶点和 n 条有向边构成的有向图。我们可以通过顶点-边的 $m \times n$ 关联矩阵来描述这样的网络。图5是一个具有 4 个顶点和 4 条边的网络例子，其顶点-边的关联矩阵是：

$$B = \begin{bmatrix} -1 & -1 & 0 & 0 \\ 1 & 0 & -1 & -1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

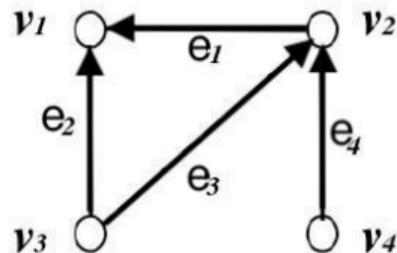


图 5: 一个有 4 个顶点的图

关联矩阵的四个基本子空间

性质 1

设图 $G = \langle \mathbb{V}, \mathbb{E} \rangle$ 是一个具有 m 个顶点和 n 条边的连通图，其对应的关联矩阵为 B ，则关于 B 的四个基本子空间具有以下性质：

- B 的左零空间维数为 1，即 $\text{Null}(B^T) = 1$ ，且 $\text{Null}(B^T) = \text{span}\{\mathbf{1}\}$ 。
- B 的行空间维数为 $m-1$ ，即 $\text{Col}(B^T) = m-1$ ，且 $\text{Col}(B^T)$ 可由 B^T 任意 $m-1$ 个列向量生成。
- B 的列空间维数为 $m-1$ ，即 $\text{Col}(B) = m-1$ ，且若 T 是图 G 的一棵生成树，那么 $\text{Col}(B)$ 可由 T 的关联矩阵的 $m-1$ 个列向量生成。
- B 的零空间维数为 $n-m+1$ ，即 $\text{Null}(B) = n-m+1$ ，这个数等于图 G 中小圈的个数。

9.2.4 无向图相关的矩阵：邻接矩阵和度矩阵

定义 10

设图 $G = \langle \mathbb{V}, \mathbb{E} \rangle$ ，我们把图 G 的顶点排列成 $v_1, v_2, \dots, v_n, n = |\mathbb{V}|$ 。用 a_{ij} 表示顶点 v_i 与顶点 v_j 之间的边数，其中 a_{ij} 定义为

$$a_{ij} = \begin{cases} 1 & \{v_i, v_j\} \in \mathbb{E} \\ 0 & \{v_i, v_j\} \notin \mathbb{E} \end{cases}$$

则称所得的矩阵 $\mathbf{A} = (a_{ij})_{n \times n}$ 为无向图 G 的邻接矩阵。我们把节点 v_i 相邻节点的数量称为 v_i 的度，记为 $d(v_i)$ ，则 $d(v_i) = \sum_j a_{ij}$ ，图 G 的度矩阵 $\mathbf{D} = (d_{ij})_{n \times n}$ 定义为

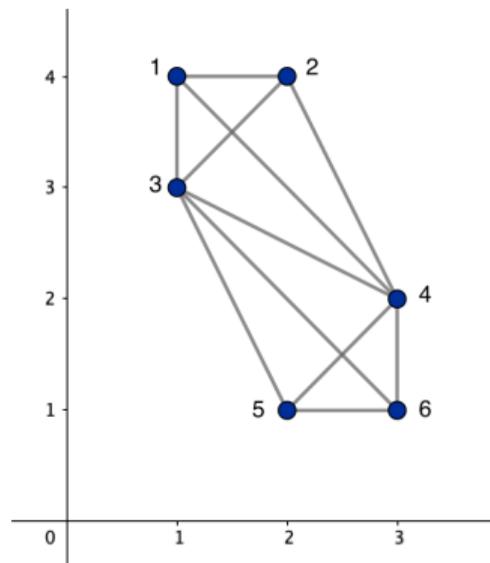
$$d_{ij} = \begin{cases} d(v_i) & i = j \\ 0 & i \neq j \end{cases}$$

谱聚类例子中的邻接矩阵和度矩阵

例 3

例1中图1(c)所对应的邻接矩阵和度矩阵分别为

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{pmatrix}, \quad \mathbf{D} = \begin{pmatrix} 3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3 \end{pmatrix}$$



邻接矩阵的性质

性质 2

设无向图 $G = (\mathbb{V}, \mathbb{E})$ 对应于顶点排列 v_1, v_2, \dots, v_n 的邻接矩阵为 A , 其中 $n = |\mathbb{V}|$, 则 A 有以下性质:

- A 是对称矩阵, 即 $A = A^T$ 。
- A 有 n 个实特征值, 其中一定有最大特征值 λ_1 是单重特征值, 且满足 $\lambda_1 \leq \max_{v \in \mathbb{V}} d(v)$ 。
- 设 v'_1, v'_2, \dots, v'_n 图 G 节点另一种排列, 其对应的邻接矩阵为 A' , 则 A 与 A' 具有相同的特征值。

定义 11

设图 $G = (\mathbb{V}, \mathbb{E})$ 的邻接矩阵为 \mathbf{A} 。我们则称

- 矩阵 \mathbf{A} 的特征值为图 G 的特征值。
- 矩阵 \mathbf{A} 的谱为图 G 的谱。

例 4

在例3中的邻接矩阵的特征值从小到大分别为

$$\lambda_1 = -1.82842712$$

$$\lambda_2 = \lambda_3 = \lambda_4 = -1$$

$$\lambda_5 = 1$$

$$\lambda_6 = 3.82842712$$

无向图的拉普拉斯矩阵

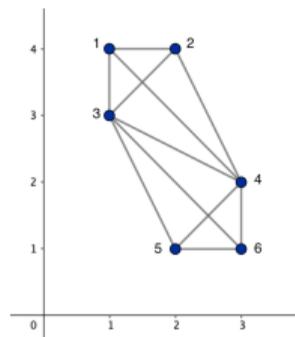
定义 12

设无向图 $G = \langle \mathbb{V}, \mathbb{E} \rangle$ 的邻接矩阵和度矩阵分别为 A 和 D ，我们称矩阵 $L = D - A$ 为图 G 的拉普拉斯矩阵。

例 5

例1中图1(c) 对应的拉普拉斯矩阵为

$$L = \begin{pmatrix} 3 & -1 & -1 & -1 & 0 & 0 \\ -1 & 3 & -1 & -1 & 0 & 0 \\ -1 & -1 & 5 & -1 & -1 & -1 \\ -1 & -1 & -1 & 5 & -1 & -1 \\ 0 & 0 & -1 & -1 & 3 & -1 \\ 0 & 0 & -1 & -1 & -1 & 3 \end{pmatrix}$$



正规化的拉普拉斯矩阵

定义 13

设无向图 $G = \langle \mathbb{V}, \mathbb{E} \rangle$ 的邻接矩阵和度矩阵分别为 A 和 D 。我们称矩阵

$$\tilde{L} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}} = I - D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$$

为图 G 的正规化的拉普拉斯矩阵。

9.2.5 加权图相关的矩阵

对于例 1 中的聚类例子，我们的目的是要把它聚成两类。我们从例 5 中其拉普拉斯矩阵

$$\mathbf{L} = \begin{pmatrix} 3 & -1 & -1 & -1 & 0 & 0 \\ -1 & 3 & -1 & -1 & 0 & 0 \\ -1 & -1 & 5 & -1 & -1 & -1 \\ -1 & -1 & -1 & 5 & -1 & -1 \\ 0 & 0 & -1 & -1 & 3 & -1 \\ 0 & 0 & -1 & -1 & -1 & 3 \end{pmatrix}$$

中可以发现，第 3 个节点和第 4 个节点是对称的。也就是说，如果我们仅仅根据这个无向图的拉普拉斯矩阵，我们如果可以把数据聚成 $\{1, 2, 3\}, \{4, 5, 6\}$ 这两类，也就可以聚成 $\{1, 2, 4\}, \{3, 5, 6\}$ 这样两类。但是后者显然不是很合理。这主要因为我们前面定义的邻接矩阵并没有对连接两个顶点的边的长度进行区别考虑。所以，在实际的聚类中，我们要考虑对边进行赋权，构建权重相关的邻接矩阵和拉普拉斯矩阵。

加权图的邻接矩阵

定义 14

设加权图 $G = \langle \mathbb{V}, \mathbb{E} \rangle$, 我们把图 G 的顶点排列成 $v_1, v_2, \dots, v_n, n = |\mathbb{V}|$, 我们将图 G 的邻接矩阵 $\mathbf{A} = (a_{ij})_{n \times n}$ 定义为

$$a_{ij} = \begin{cases} w_{i,j} & \{v_i, v_j\} \in \mathbb{E} \\ 0 & \{v_i, v_j\} \notin \mathbb{E} \end{cases}$$

其中 $w_{i,j}$ 是边 $\{v_i, v_j\}, i, j = 1, \dots, n$ 上的权重。

在实际问题中, 权重的定义方式多种多样。在聚类中, 一种较为常用的权重定义方式是使用高斯核

$$w_{ij} = e^{-\frac{\|v_i - v_j\|_2^2}{2\sigma^2}}$$

其中 $\|v_i - v_j\|_2$ 表示顶点 v_i 和 v_j 的欧氏距离, σ 是一参数, 用于调节顶点间距离到权重的映射值。这样定义的好处是, 权重介于 0 和 1 之间。

加权图的度矩阵和拉普拉斯矩阵

定义 15

设加权图 $G = \langle \mathbb{V}, \mathbb{E} \rangle$, 我们把图 G 的顶点排列成 $v_1, v_2, \dots, v_n, n = |\mathbb{V}|$, 则顶点 v_i 的带权度数定义为 $d(v_i) = \sum_j w_{ij}$, 其中 w_{ij} 是边 $\{v_i, v_j\}, i, j = 1, \dots, n$ 上的权重。图 G 的度矩阵 $D = (d_{ij})_{n \times n}$ 定义为

$$d_{ij} = \begin{cases} d(v_i) & i = j \\ 0 & i \neq j \end{cases}$$

定义 16

设加权图 $G = \langle \mathbb{V}, \mathbb{E} \rangle$ 的邻接矩阵和度矩阵分别为 A 和 D , 我们称 $L = D - A$ 为图 G 的拉普拉斯矩阵。

加权图的正规化拉普拉斯矩阵

定义 17

设加权图 $G = \langle \mathbb{V}, \mathbb{E} \rangle$ 的邻接矩阵和度矩阵分别为 A 和 D ，我们称矩阵

$$\tilde{L} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}} = I - D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$$

为图 G 的正规化拉普拉斯矩阵。

对于一个加权图，如果我们令图上所有边的权重变为原来的 k 倍， $k \neq 0$ 。那么显然对于未正规化的拉普拉斯矩阵 L 将变为 kL 。而正规化的拉普拉斯矩阵 \tilde{L} 则不会发生变化。这是因为

$$\tilde{L}_{ij} = \frac{L_{ij}}{\sqrt{d(v_i)d(v_j)}} = \frac{kL_{ij}}{\sqrt{kd(v_i)kd(v_j)}}$$

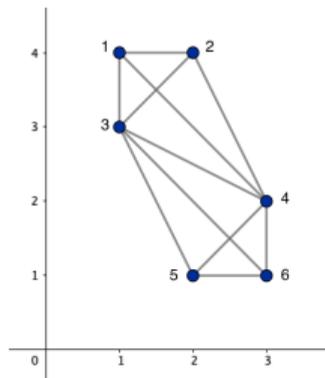
因此正规化的拉普拉斯矩阵更为常用，它能够避免权重绝对值大小的影响。

谱聚类例子中的拉普拉斯矩阵

例 6

如果使用高斯核来给例1中图1(c)上的边进行赋权, 则可得到如下拉普拉斯矩阵:

$$L = \begin{pmatrix} 1.231 & -0.607 & -0.607 & -0.018 & 0 & 0 \\ -0.607 & 1.056 & -0.368 & -0.082 & 0 & 0 \\ -0.607 & -0.368 & 1.157 & -0.082 & -0.082 & -0.018 \\ -0.018 & -0.082 & -0.082 & 1.157 & -0.368 & -0.607 \\ 0 & 0 & -0.082 & -0.368 & 1.056 & -0.607 \\ 0 & 0 & -0.018 & -0.607 & -0.607 & 1.231 \end{pmatrix}$$



拉普拉斯矩阵和关联矩阵的关系

性质 3

设有向无权图 G 的关联矩阵为 B ，其对应的无向图的拉普拉斯矩阵为 L ，则 L 和 B 满足以下关系：

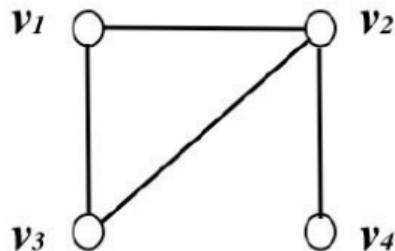
$$L = B^T B$$

例 7

在例 2 中图 5 对应的拉普拉斯矩阵和关联矩阵满足

$$L = BB^T$$

$$L = \begin{bmatrix} 2 & -1 & -1 & 0 \\ -1 & 3 & -1 & -1 \\ -1 & -1 & 2 & 0 \\ 0 & -1 & 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} -1 & -1 & 0 & 0 \\ 1 & 0 & -1 & -1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$



拉普拉斯矩阵的性质

性质 4

设正权图 $G = (\mathbb{V}, \mathbb{E})$ 对应于顶点排列 v_1, v_2, \dots, v_n 的拉普拉斯矩阵和正规化拉普拉斯矩阵分别为 L 和 \tilde{L} , 其中 $n = |\mathbb{V}|$, 则 L 和 \tilde{L} 有以下性质:

1. L 和 \tilde{L} 是对称矩阵, 即有 $L = L^T$ 和 $\tilde{L} = \tilde{L}^T$ 。
2. 对任意的 n 维向量 x , 有 $x^T L x \geq 0$ 和 $x^T \tilde{L} x \geq 0$, 因而 L 和 \tilde{L} 是正半定矩阵。
3. L 和 \tilde{L} 的最小特征值为 0, 且对应的特征向量分别为 $\mathbf{1}$ 和 $D^{-\frac{1}{2}} \mathbf{1}$ 。

证明:

为了证明第 2 条性质, 我们首先证明等式

$$\sum_{\{v_i, v_j\} \in \mathbb{E}} w_{ij}(x_i - x_j)^2 = \mathbf{x}^T \mathbf{L} \mathbf{x}$$

利用拉普拉斯矩阵的定义有

$$\begin{aligned} \mathbf{x}^T \mathbf{L} \mathbf{x} &= \mathbf{x}^T \mathbf{D} \mathbf{x} - \mathbf{x}^T \mathbf{A} \mathbf{x} = \sum_{i=1}^n d_i x_i^2 - \sum_{i,j=1}^n w_{ij} x_i x_j \\ &= \frac{1}{2} \left(\sum_{i,j=1}^n d_i x_i^2 - 2 \sum_{i,j=1}^n w_{ij} x_i x_j + \sum_{j=1}^n d_j x_j^2 \right) \\ &= \frac{1}{2} \sum_{i,j=1}^n w_{ij} (x_i - x_j)^2 \\ &= \sum_{\{v_i, v_j\} \in \mathbb{E}} w_{ij} (x_i - x_j)^2 \end{aligned}$$

证明. 续:

那么对于一个正权图来说, 无论 \mathbf{x} 取什么, $\mathbf{x}^T \mathbf{L} \mathbf{x}$ 都是非负的。所以 \mathbf{L} 是半正定矩阵。
对于 $\tilde{\mathbf{L}}$ 和任意的 \mathbf{x} 有

$$\begin{aligned} & \mathbf{x}^T \tilde{\mathbf{L}} \mathbf{x} \\ &= \mathbf{x}^T \mathbf{D}^{-\frac{1}{2}} \mathbf{L} \mathbf{D}^{-\frac{1}{2}} \mathbf{x} \\ &= (\mathbf{D}^{-\frac{1}{2}} \mathbf{x})^T \mathbf{L} (\mathbf{D}^{-\frac{1}{2}} \mathbf{x}) \geq 0 \end{aligned}$$

所以 $\tilde{\mathbf{L}}$ 也是半正定矩阵。

对于第 3 条性质. 我们只需要分别计算

$$\mathbf{L} \mathbf{1} = 0, \tilde{\mathbf{L}} \mathbf{D}^{-\frac{1}{2}} \mathbf{1} = 0$$

并且综合第 2 条性质 $\mathbf{L}, \tilde{\mathbf{L}}$ 是正半定的, 我们可以知道 0 是 $\mathbf{L}, \tilde{\mathbf{L}}$ 最小的特征值, 并且对应的特征向量分别为 $\mathbf{1}$ 和 $\mathbf{D}^{-\frac{1}{2}} \mathbf{1}$

对于谱聚类，我们最终可以将问题转化为求该图对应的拉普拉斯矩阵或正规化拉普拉斯矩阵次小特征值对应的特征向量问题。在得到特征向量后，对其分量进行聚类，聚类结果即为谱聚类的结果。

例 8

在例 6 中，我们已经得到例 1 中图 1(c) 对应的拉普拉斯矩阵 L 。可以计算得到它的次小特征值对应的特征向量为

$$\mathbf{x} = (-0.442, -0.421, -0.358, 0.358, 0.421, 0.442)^T$$

我们很容易就可以得到前 3 个节点作为一类，后 3 个节点作为一类。

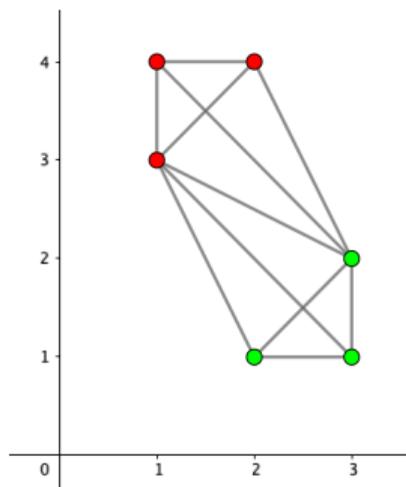


图 6: 谱聚类

9.2.6 稀疏矩阵

定义 18

一个矩阵中，若数值为零的元素的数目远远多于非零元素的数目称这样的矩阵为稀疏矩阵。

稀疏矩阵零元素分布常常是没有规律的。

定义 19

当一个矩阵的非零元素数目远远多于零元素数目时，称这样的矩阵为稠密矩阵。

例 9

如下几个矩阵我们都可以认为是稀疏矩阵

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 1 \\ 0 & -2 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}; \mathbf{B} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 2 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}; \mathbf{C} = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix}$$

常见的稀疏矩阵：邻接矩阵

例 10

矩阵

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

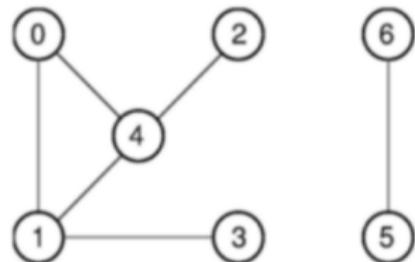


图 7: 图与邻接矩阵

是图7对应的邻接矩阵。

常见的稀疏矩阵：二分图相关的矩阵

有时我们会碰到一些特殊的图，如二分图，此时我们可以给出另外一种将图转换为矩阵的方式。比如图8对应的矩阵为

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

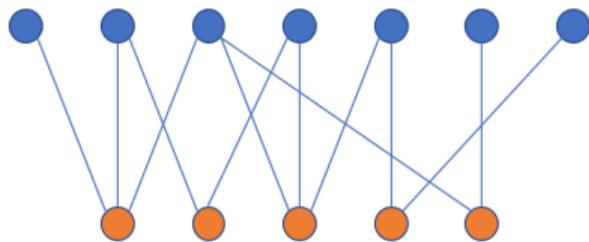


图 8: 二分图

二分图的常见应用场景

- 电子商务；当我们在处理用户 - 网页、用户 - 服务、用户 - 产品等问题时，会遇到这样一个二分图，将这个图转换为矩阵也常常是一个稀疏的矩阵。
- 此外，在深度学习中，一个多层感知机两层之间也是这样一个二分图，
- 但对于大多数一般的深度神经网络，各层之间的连接矩阵不是一个稀疏矩阵而是一个稠密矩阵，
- 在模型压缩领域，现在有一些剪枝方法，可以修剪掉一些不需要的边，这时便有可能得到一个稀疏矩阵。

其它常见应用场景中的稀疏矩阵

在实际应用中，我们会接触大量的稀疏矩阵，尤其是超大型的稀疏矩阵，例如：

- 推荐系统：用户只可能对有限商品进行过评价，对于大量的其他商品是没有过评价信息的，因此在用户 - 商品评价矩阵中有大量的零元素。
- 记数编码：当我们用词汇出现的频率表示文档时，大量的在词汇表中的词汇没有在文档中出现过，使得文档矩阵出现很多零元素。
- 图像矩阵：以手写识别数据集为例，只有图像中间区域出现数字，表示该位置有像素点，其他背景都被标记为 0，是表示图像的矩阵有大量的零元素。

9.2.7 低秩矩阵：引例

例 11

考虑矩阵

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 & -1 \\ 1 & 3 & 1 & 3 & 1 \\ -1 & -2 & -1 & -2 & -1 \\ 2 & 1 & 2 & 1 & 2 \end{pmatrix}$$

尽管 \mathbf{A} 中没有一个零元素，但是矩阵 \mathbf{A} 的秩只有 2，考虑矩阵 \mathbf{A} 的行空间

$$\text{Col}(\mathbf{A}^T) = \text{span}\{(1, 1, 1, 1, 1)^T, (-1, 1, -1, 1, -1)^T\}.$$

在数据科学中，我们会碰到很多大规模矩阵，也是一个稠密矩阵，但是他的秩却很低的矩阵，这样的矩阵，我们称为低秩矩阵。

定义 20

设矩阵 $\mathbf{A} \in \mathbb{R}^{n \times m}$, 如果矩阵 \mathbf{A} 的秩 $\text{rank}(\mathbf{A})$ 远小于 $\max\{n, m\}$, 那么我们称这样的矩阵为低秩矩阵。

之所以考虑的是 $\text{rank}(\mathbf{A})$ 和 $\max\{n, m\}$ 的关系, 是因为在很多时候, 我们考虑低秩矩阵时都是在非方阵的情况下, 并且常常有 n 远小于 m , 或者 m 远小于 n 的情况发生。

例 12

考虑矩阵

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 1 & 3 & 1 & 2 & 3 & 4 & 1 & 2 \\ 2 & 1 & 3 & 2 & 1 & 2 & 1 & 2 & 1 & 1 \end{pmatrix}$$

也是一个低秩矩阵。

例 13

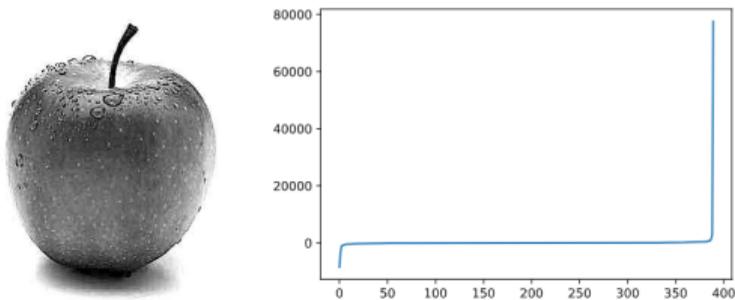


图 9: 图像以及图像的特征值

左图是一个 390×390 的图像对应矩阵的特征值从小到大展示在右图上。可以注意到很多特征值都集中在 0 附近。记左图图像的矩阵为 \mathbf{A} 。

前面我们提到若一个方阵 A 可对角化，那么存在可逆矩阵 P 有

$$A = P\Sigma P^{-1}$$

其中 Σ 是对角矩阵，且主对角线上元素是 A 从小到大的特征值。那么上面图像所对应的矩阵 A 就可以写出

$$A = P\Sigma P^{-1}$$

此时，我们如果令那些绝对值小于 200 的特征值（这些特征值的绝对值大小不到最大特征值绝对值大小的 0.3%）都设为 0，即求

$$A' = P\Sigma' P^{-1}$$

其中 $\Sigma'_{ii} = 0$ ，若 $|\Sigma_{ii}| < 200$ 。此时 $\text{rank}(A') = 112$ ，这说明我们实际上是可以把图像的矩阵看做是低秩矩阵。矩阵的特征值和特征向量存储着图像的信息，尤其是特征值大的那些特征向量存储了更多的信息。

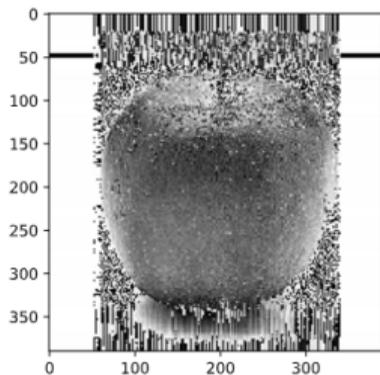


图 10: 新图像

低秩矩阵在很多领域都有用处。

- 图像恢复
- 图像校正
- 图像去噪
- 图像分割
- 图形化建模
- 组合系统辨识
- 视频监控
- 人脸识别
- 潜在语义检索
- 评分与协同筛选
- 矩阵填充
- 背景建模
- 变换不变低秩纹理 (TILT)
- ...

但这些总体上可以分为三大问题

- 低秩矩阵恢复
- 低秩矩阵补全
- 低秩矩阵表示

低秩矩阵恢复

低秩矩阵恢复

- 当低秩矩阵 \mathbf{A} 的观测或样本矩阵 $\mathbf{D} = \mathbf{A} + \mathbf{E}$ 的某些元素被严重损坏时。我们希望能够自动识别被损坏的元素，精确地恢复原低秩矩阵 \mathbf{A} 。
- 在工程和应用科学的许多领域（例如机器学习、控制、系统工程、信号处理、模式识别和计算机视觉）中，将一个数据矩阵分解为一个低秩矩阵与一个误差（或扰动）矩阵之和，旨在恢复低秩矩阵是远远不够的，而是需要将一个数据矩阵 \mathbf{D} 分解为一个低秩矩阵 \mathbf{A} 与一个稀疏矩阵 \mathbf{E} 之和 $\mathbf{D} = \mathbf{A} + \mathbf{E}$ ，并且希望同时恢复低秩矩阵与稀疏矩阵。矩阵的这类分解称为低秩与稀疏矩阵分解。
- 通常这种问题，我们使用鲁棒 PCA 来求解。

鲁棒 PCA

$$\min_{\mathbf{A}, \mathbf{E}} \|\mathbf{A}\|_* + \lambda \|\mathbf{E}\|_1, s.t. \mathbf{X} = \mathbf{A} + \mathbf{E}$$

其中 \mathbf{A} 代表了低秩结构信息， \mathbf{E} 是稀疏噪声

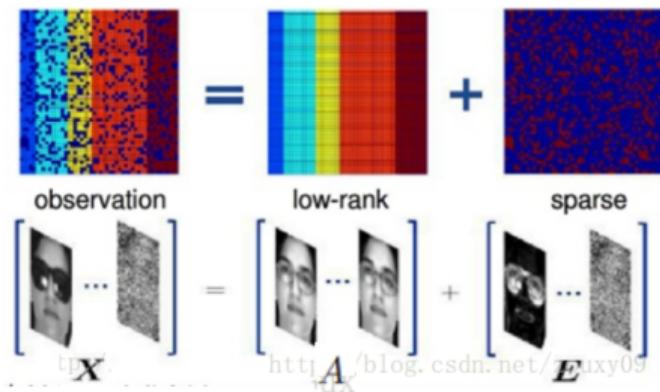


图 11: 低秩矩阵恢复

低秩矩阵补全

低秩矩阵补全

- 当数据矩阵 \mathbf{D} 含丢失元素时，可根据矩阵的低秩结构来恢复的所有元素，称此恢复过程为矩阵补全。
- 记 Ω 为集合 $\{1, 2, \dots, m\} \times \{1, 2, \dots, n\}$ 的子集，矩阵补全的原始模型可描述为如下的优化问题

$$\min_{\mathbf{A}} \text{rank}(\mathbf{A}) \text{ s.t. } P_{\Omega}(\mathbf{A}) = P_{\Omega}(\mathbf{D})$$

其中 $P_{\Omega} : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^{m \times n}$ 为一线性投影算子，即

$$P_{\Omega}(\mathbf{D}_{ij}) = \begin{cases} D_{ij} & (i, j) \in \Omega \\ 0 & (i, j) \notin \Omega \end{cases}$$

- 为便于优化，凸优化后转化为

$$\min_{\mathbf{A}} \|\mathbf{A}\|_* \text{ s.t. } P_{\Omega}(\mathbf{A}) = P_{\Omega}(\mathbf{D})$$

低秩矩阵表示

低秩矩阵表示

- 低秩矩阵表示是将数据集矩阵 D 表示成字典矩阵 B （也称为基矩阵）下的线性组合，即 $D = BZ$ ，并希望线性组合系数矩阵 Z 是低秩的。为此，需要求解下列优化问题

$$\min_Z \text{rank}(Z) \text{ s.t. } D = BZ$$

- 为便于优化，凸松弛后转化为

$$\min_Z \|Z\|_* \text{ s.t. } D = BZ$$

- 若选取数据集 D 本身作为字典，则有

$$\min_Z \|Z\|_* \text{ s.t. } D = DZ$$

那么我们可以直接写出他的解，但是需要用到 SVD 分解的知识。

- 为了对噪声和离群点更加鲁棒，一个更合理的模型为

$$\min_{Z, E} \|Z\|_* + \lambda \|E\|_{2,1} \text{ s.t. } D = DZ + E$$

本讲小结

分解的目标矩阵

- 长方形矩阵
- 方阵
- 对称矩阵
- 正半定矩阵
- 稀疏矩阵和低秩矩阵

具有特殊性质的表示矩阵

- 三角矩阵
- 对角矩阵
- 正交矩阵
- Dyad (秩 1 矩阵)

矩阵分解包括：LU 分解、QR 分解、谱分解和 Cholosky 分解、奇异值分解；将用于线性方程组求解、最小二乘问题和矩阵特征值的计算等，一起形成矩阵的四大基本计算问题。