

# 线性代数初步

邓明格

七月在线

*mingge\_deng@brown.edu*

March 31, 2017

# 主要内容

## 1 前言与简介

- 向量与矩阵实例
- 网页搜索排名问题

## 2 向量与向量空间

- 向量，向量空间，线性组合，基
- 向量范数，内积

## 3 矩阵论初步

- 矩阵的意义
- 矩阵的性质

## 4 线性回归

- 线性回归问题
- 线性回归问题的近似

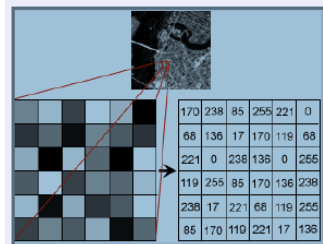
文本:

A (real) vector is just a collection of real numbers, referred to as the components (or, elements) of the vector;  $\mathcal{R}^n$  denotes the set of all vectors with  $n$  elements. If  $x \in \mathcal{R}^n$  denotes a vector, we use subscripts to denote elements, so that  $x_i$  is the  $i$ -th component of  $x$ . Vectors are arranged in a column, or a row. If  $x$  is a column vector,  $x^T$  denotes the corresponding row vector, and vice-versa.

- 行向量  $c = [5; 3; 3; 4]$  表示单词列表  $V = \{\text{vector}, \text{elements}, \text{of}, \text{the}\}$  中单词在文本中出现次数。
- 行向量  $x = [1/3, 1/5, 1/5, 4/15]$  表示单词出现相对频率。
- Frequency-based representation of text documents (bag-of-words).

# 图像表示

灰度图 / RGB 彩图



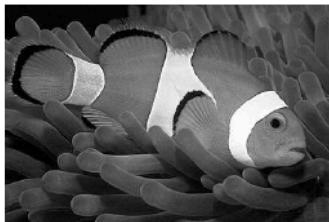
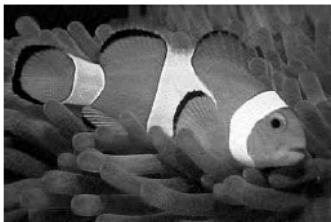
|     |     |    |     |     |   |
|-----|-----|----|-----|-----|---|
| 170 | 238 | 65 | 255 | 221 | 0 |
|-----|-----|----|-----|-----|---|

...

|    |     |     |     |    |     |
|----|-----|-----|-----|----|-----|
| 85 | 170 | 119 | 221 | 17 | 136 |
|----|-----|-----|-----|----|-----|

# 图像压缩

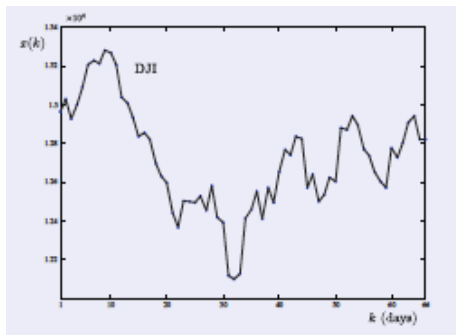
## SVD 压缩



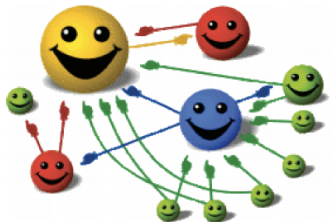
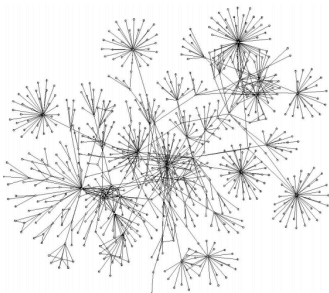
# 时间序列表示

时间序列：物理量  $X(t)$  随离散时间  $t$  变化的序列

$$X(t) = [X(1), X(2), \dots, X(T)]^T \in \mathcal{R}^T \quad (1)$$



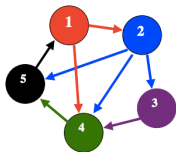
# 网页排名



- 英特网可以表示为连接图
- 更多访问更高排名

# 网页排名

## 英特网的矩阵表示



$$S = \begin{pmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

- **S**: 转移矩阵
- 第  $i$  行表示: 从网页  $i$  到其他网页的概率
- 第  $j$  列表示: 从其他网页到网页  $j$  的概率
- 网页排名: 特征向量问题

$$\pi^T S = \pi^T \quad \pi > 0 \quad \|\pi\|_1 = 1 \quad (2)$$

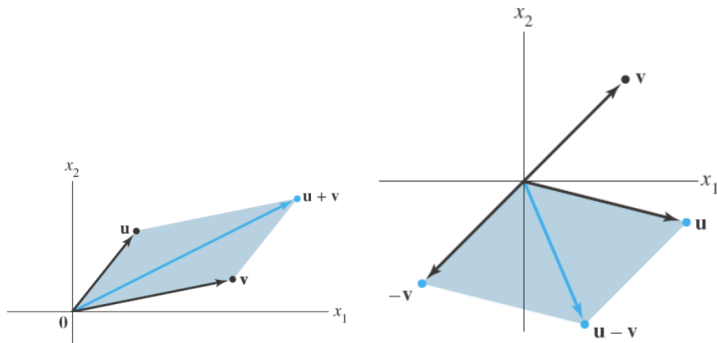


Want  $\pi$  such that  $\pi^T \mathbf{S} = \pi^T$

- 随机选取  $x(0)$
- 迭代直到收敛:  $x(k+1)^T := x(k)^T \mathbf{S}$
- $\pi = x(t \rightarrow \infty)$
  
- 幂法
- 矩阵乘法的 MapReduce 并行

# 向量与向量空间

- 向量加减法, 向量与标量数乘, 满足交换律, 结合律和分配律等
- 向量空间, 零向量, 单位向量
- 实数向量空间  $\mathcal{R}^n$



# 线性组合与线性无关

## 线性组合

向量  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p \in \mathcal{R}^n$ , 标量  $c_1, c_2, \dots, c_p$ , 并且向量  $\mathbf{y} \in \mathcal{R}^n$  可以表达为

$$\mathbf{y} = c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2 + \dots + c_p \mathbf{v}_p \quad (3)$$

称为  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p$  的线性组合, 其中标量  $c_1, c_2, \dots, c_p$  称为线性组合的系数或权.

## 线性无关

向量空间的一组元素称为线性无关, 如果其中没有向量可表示成有限个其他向量的线性组合, 反之称为线性相关.

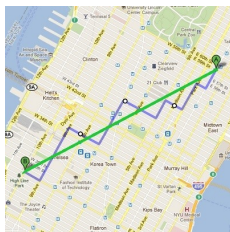
$$c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2 + \dots + c_p \mathbf{v}_p = \mathbf{0} \quad \text{iff} \quad c_1 = c_2 = \dots = c_p = 0 \quad (4)$$

# 线性空间的基

## 基

- 基是线性空间里的一组向量，使得任何一个向量都可以唯一的表示成这组基的线性组合。
- 基向量之间线性无关。
- 基给出了定量描述线性结构的方法 (坐标系)。
- 线性代数的基本问题就是基组的选择问题。

考虑纽约曼哈顿地图，街道用数字编号，非正南正北，如何选取基？



# 线性生成空间与空间维度

## 线性生成空间

向量  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p \in \mathcal{R}^n$ , 他们所有线性组合的集合称为向量  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p$  的线性生成空间, 表示为  $\text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p\}$

$$\text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p\} = \{\mathbf{y} | \exists c_1, \dots, \exists c_p, \mathbf{y} = c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2 + \dots + c_p \mathbf{v}_p\} \quad (5)$$

- $\forall i, \mathbf{v}_i \in \text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p\}$
- $\mathbf{0} \in \text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p\}$

## 空间维度

- 线性空间:  $\text{Span}\{\text{basis}\}$  基组的线性生成空间
- 空间维度: 基的数量

# 向量 $l_p$ 范数

## 向量范数

向量范数是具有“长度”概念的函数，是向量空间到实数的映射  $\|\cdot\|: \mathcal{R}^n \rightarrow \mathcal{R}$ ，并满足

- 正定性:  $\|\mathbf{x}\| \geq 0$      $\|\mathbf{x}\| = 0$     iff    $\mathbf{x} = 0$
- 齐次性:  $\|c\mathbf{x}\| = |c| \|\mathbf{x}\|$ ,
- 三角不等式:  $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$

$l_p$  范数:  $\|\mathbf{x}\|_p = (\sum_{i=1}^N |x_i|^p)^{\frac{1}{p}}$

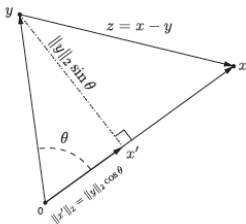
- $l_0$  范数: 向量中非零元素的个数
- $l_1$  范数: 向量元素绝对值之和
- $l_2$  范数: 欧几里德距离 or 欧几里德范数 (通常说的距离或者模)
- $l_\infty$  范数: 所有向量元素绝对值中的最大值
- $l_{-\infty}$  范数: 所有向量元素绝对值中的最小值

向量范数极其重要: 回归问题的正则化

# 向量内积

## 向量内积，投影

- 从内积到范数  $\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$
- 欧式空间  $\mathcal{R}^n$  标准内积  $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y} = \sum_{k=1}^n \mathbf{x}_k \mathbf{y}_k$
- 向量之间的夹角  $\cos \theta = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\|_2 \|\mathbf{y}\|_2}$
- 向量正交（垂直） $\langle \mathbf{x}, \mathbf{y} \rangle = 0$
- 向量平行  $\langle \mathbf{x}, \mathbf{y} \rangle$  极大值



# 向量与向量空间小结

- 向量空间定义及表示（基）
- 向量的度量（范数）



# 矩阵的理解一：线性空间的元素

- 矩阵：线性空间的元素（折叠的向量）
- 由此可以定义加减，乘，数乘，秩（维度），矩阵范数等等

## 矩阵的理解二：线性映射

- 线性方程组

$$\begin{aligned}x_1 + 2x_2 - x_3 &= 4 \\ -5x_2 + 3x_3 &= 1\end{aligned}\tag{6}$$

- 等价矩阵方程

$$\begin{bmatrix} 1 & 2 & -1 \\ 0 & -5 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$$

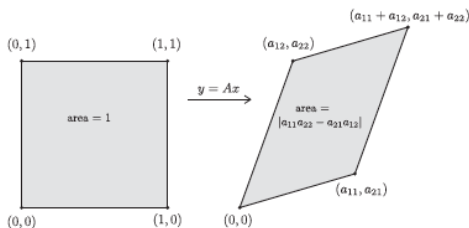
- 线性映射  $\mathbf{V}$  和  $\mathbf{W}$  是两个实线性空间,  $\mathbf{T}: \mathbf{V} \rightarrow \mathbf{W}$  满足如下条件就是线性映射
  - $T(v_1 + v_2) = T(v_1) + T(v_2) \quad \forall v_1, v_2 \in \mathbf{V}$
  - $T(cv) = cT(v) \quad \forall c \in \mathcal{R}, \quad v \in \mathbf{V}$
- 矩阵: 向量 (线性空间) 到向量 (线性空间) 的线性映射
- 矩阵 (线性函数) V.S 函数
- 由此可以定义矩阵的逆, 秩, 特征值, 特征向量等

# 矩阵行列式的值

- 方阵  $\forall \mathbf{A} \in \mathcal{R}^{n,n}$ , 行列式的值定义为

$$\det(\mathbf{A}) = \sum_{j=1}^n (-1)^{i+j} a_{ij} \det \mathbf{A}_{(i,j)} \quad (7)$$

- $\forall \mathbf{A} \in \mathcal{R}^{n,n}$  为奇异方阵  $\Leftrightarrow \det \mathbf{A} = 0$
- $\forall \mathbf{A}, \mathbf{B} \in \mathcal{R}^{n,n}$  以及标量  $\alpha \in \mathcal{R}$ ,
  - $\det \mathbf{A} = \det \mathbf{A}^T$
  - $\det \mathbf{AB} = \det \mathbf{BA} = \det \mathbf{A} \det \mathbf{B}$
  - $\det \alpha \mathbf{A} = \alpha^n \det \mathbf{A}$



# 矩阵的逆

- $\forall \mathbf{A} \in \mathcal{R}^{n,n}$  为非奇异方阵 ( $\det \mathbf{A} \neq 0$ ), 矩阵的逆定义为存在并唯一的  $n \times n$  矩阵满足

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_n \quad (8)$$

- $\forall \mathbf{A}, \mathbf{B} \in \mathcal{R}^{n,n}$  为非奇异方阵

- $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$
- $(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T$
- $\det \mathbf{A} = \det \mathbf{A}^T = \frac{1}{\det \mathbf{A}^{-1}}$

# 矩阵的秩

- $\forall \mathbf{A} \in \mathcal{R}^{m,n}$ , 矩阵  $\mathbf{A}$  的生成域 (线性变换的值域) 定义为

$$\mathcal{R}(\mathbf{A}) = \{\mathbf{Ax} : \mathbf{x} \in \mathcal{R}^n\} \quad (9)$$

- 矩阵  $\mathbf{A}$  的生成域为一个线性子空间, 其维度极为矩阵  $\mathbf{A}$  的秩。
- 矩阵  $\mathbf{A}$  的列秩是  $\mathbf{A}$  的线性独立的纵列的极大数目, 行秩是  $\mathbf{A}$  的线性独立的横行的极大数目。矩阵的列秩和行秩总是相等的, 并等于矩阵  $\mathbf{A}$  的秩  $rank(\mathbf{A})$ 。
- $1 \leq rank(\mathbf{A}) \leq \min(m, n)$
- 矩阵的秩等于其非零特征值的个数

# 矩阵的特征值，特征向量

- $\lambda \in \mathcal{C}$  为方阵  $\mathbf{A} \in \mathcal{R}^{n,n}$  的特征值，满足  $\exists \mathbf{u} \in \mathcal{C}^n, \mathbf{u} \neq \mathbf{0}$ , 并且  $\mathbf{A}\mathbf{u} = \lambda\mathbf{u}$ , 或者  $(\lambda\mathbf{I}_n - \mathbf{A})\mathbf{u} = \mathbf{0}$ ,  $\mathbf{u}$  为对应的特征向量。
- 特征值为矩阵  $\mathbf{A}$  的特征多项式的根  $p(\lambda) = \det(\lambda\mathbf{I}_n - \mathbf{A}) = 0$
- 特征向量构成正交基组
- 对比函数的傅立叶级数（其他正交级数）展开
- Google 网页排名问题

# 矩阵范数

## 矩阵范数

是矩阵所在线性空间到实数的映射  $\|\cdot\|: \mathcal{R}^{n,m} \rightarrow \mathcal{R}$ , 并满足

- 正定性:  $\|\mathbf{A}\| \geq 0$      $\|\mathbf{A}\| = 0$     iff     $\mathbf{A} = \mathbf{0}$
- 齐次性:  $\|c\mathbf{A}\| = |c| \|\mathbf{A}\|$ ,
- 三角不等式:  $\|\mathbf{A} + \mathbf{B}\| \leq \|\mathbf{A}\| + \|\mathbf{B}\|$

诱导范数:

$$\|\mathbf{A}\| = \max\{\|\mathbf{Ax}\|, \|\mathbf{x}\| = 1\} = \max\{\|\mathbf{Ax}\| / \|\mathbf{x}\|, \mathbf{x} \neq \mathbf{0}\}$$

- 1 范数:  $\|\mathbf{A}\|_1 = \max_j \sum_{i=1}^m |a_{i,j}|$  列向量绝对值之和最大值
- 2 范数:  $\|\mathbf{A}\|_2 = \sqrt{\lambda}$ ,  $\lambda$  为  $\mathbf{AA}^T$  的最大特征值, 谱范数。
- $\infty$  范数:  $\|\mathbf{A}\|_\infty = \max_i \sum_{j=1}^n |a_{i,j}|$  行向量绝对值之和最大值
- $F$  范数:  $\|\mathbf{A}\|_F = \left( \sum_{i=1}^m \sum_{j=1}^n |a_{i,j}|^2 \right)^{\frac{1}{2}} = \sqrt{\text{tr}(\mathbf{A}^T \mathbf{A})}$

矩阵范数的应用: 推荐系统 (Netflix million dollar problem) 的正则化

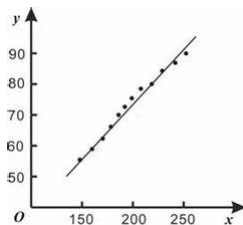
# 矩阵论初步总结

- 矩阵可以理解为线性空间的元素
- 矩阵可以理解为线性映射
- 矩阵的逆（线性映射的反函数）
- 矩阵行列式值（线性变换前后体积比）
- 矩阵的秩（线性映射生成域的维度）
- 矩阵的特征向量（线性映射正交分解的正交基）
- 矩阵的特征值（线性映射正交分解的系数）



# 线性回归问题

- 线性方程组  $\Leftrightarrow$  等价向量方程  $\Leftrightarrow$  等价矩阵方程。
- 未知数个数大于方程个数（无穷多个解），未知数个数等于方程个数（唯一解），未知数个数小于方程个数（无解）。
- 一般来讲，样本个数大于自参数个数。所以方程个数大于这个方程的未知数个数，方程通常是没有解。
- 如果样本存在误差，如何寻找近似解



## 线性回归问题

- 模型:  $y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \epsilon$ , 余项  $\epsilon$  服从正态分布  $N(0, \sigma)$
- 数据: 样本  $(Y_i, X_{i1}, \dots, X_{ik}), 1 \leq i \leq n$
- 目的: 估计参数  $\beta_1, \dots, \beta_k$

# 线性空间投影近似

- 矩阵方程:  $\mathbf{Y} = \mathbf{X} \cdot \beta + \epsilon$
- 忽略误差项  $\mathbf{Y} = \mathbf{X} \cdot \beta \Rightarrow \beta = \mathbf{X}^{-1}\mathbf{Y}$ , 但是  $\mathbf{X}$  非方阵, 且不一定可逆。
- $\mathbf{X}^T\mathbf{X}$  为方阵, 且逆通常存在, 回归问题的近似解为 (正则方程)

$$\mathbf{X}^T\mathbf{X} \cdot \beta = \mathbf{X}^T\mathbf{Y} \Rightarrow \beta = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} \quad (10)$$

- 向量方程:  $\mathbf{Y}$  为矩阵  $\mathbf{X}$  列向量的线性组合, 其系数为  $\beta$ , 问题转化为求  $\mathbf{Y}$  在  $\mathbf{X}$  列向量生成线性空间的坐标。
- $\mathbf{Y}$  通常不在  $\mathbf{X}$  列向量生成线性空间中, 近似解则为其在此空间投影  $\mathbf{Y}^*$  的坐标。
- 投影的定义:  $\mathbf{X}^T \cdot (\mathbf{X}\beta - \mathbf{Y}) = \mathbf{0}$  (垂直)

# 最小二乘

- 极小化误差项  $\mathbf{Y} - \mathbf{X} \cdot \beta$
- $l_2$  范数极小化,  $\min L(\beta) = (\mathbf{Y} - \mathbf{X} \cdot \beta)^T \cdot (\mathbf{Y} - \mathbf{X} \cdot \beta)$
- 梯度法求极值

$$\nabla L(\beta) = 0 \quad \Rightarrow \quad -2\mathbf{X}^T \mathbf{Y} + 2\mathbf{X}^T \mathbf{X} = 0 \quad \Rightarrow \quad \beta = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} \quad (11)$$

# 极大似然估计

- $\epsilon \sim N(0, \sigma)$ , 则  $y \sim N(\sum_{j=0}^k x_j \beta_j, \sigma)$
- 样本的似然函数:

$$L(\beta) = \prod_{i=1}^n \exp\left(-\frac{(\sum_{j=0}^k x_j \beta_j - y)^2}{2\sigma^2}\right) \quad (12)$$

- 样本的对数似然函数:

$$l(\beta) = \sum_{i=1}^n \frac{(\sum_{j=0}^k x_j \beta_j - y)^2}{2\sigma^2} = \frac{1}{2\sigma^2} (\mathbf{Y} - \mathbf{X} \cdot \beta)^T \cdot (\mathbf{Y} - \mathbf{X} \cdot \beta) \quad (13)$$

- 优化函数与最小二乘法一致.

# 梯度下降法

- 正则方程  $\beta = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$  几乎不使用，速度太慢
- 梯度下降法求极值, 如果实值函数  $F(\mathbf{x})$  在点  $\mathbf{x}'$  处可微且有定义, 那么函数  $F(\mathbf{x})$  在  $\mathbf{x}'$  点沿着梯度相反的方向  $-\nabla F(\mathbf{x}')$  下降最快。
- 考虑到这一点, 我们可以从函数  $F(\mathbf{x})$  的局部极小值的初始估计  $\mathbf{x}_0$  出发, 并考虑如下序列  $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \dots$ , 使得

$$\mathbf{x}_{n+1} = \mathbf{x}_n - \gamma \nabla F(\mathbf{x}_n) \quad (14)$$

因此可得到  $F(\mathbf{x}_0) \geq F(\mathbf{x}_1) \geq F(\mathbf{x}_2) \geq \dots$ , 那么  $(\mathbf{x}_n)$  将收敛到期望的极值。

- 线性回归问题

$$\beta_j = \beta_j - \gamma \frac{\partial}{\partial \beta_j} L(\beta) = \beta_j - \gamma \sum_{i=1}^n \left( \sum_{j=0}^k x_j^{(i)} \beta_j - y^{(i)} \right) x_j^{(i)} \quad (15)$$

# 谢谢大家!!

下节课主要内容:

- 线性映射在不同基组下的表示: 矩阵的相似变换
- 矩阵的相合变换
- 对称矩阵的正交标准型 (重要, 协方差矩阵)
- SVD, PCA
- 矩阵分解, 数值算法