线性代数基础

七月在线 唐博士

2018年3月10日

课程纲要

- 理论回顾
  - 基本概念
  - 矩阵计算及性质
  - 矩阵微积分
- 实践讲解——最小二乘法
  - 基本引导
  - 统计意义
  - 几何意义

基本概念

- 线性代数提供了一种紧凑地表示和操作线性方程组的方法

$$4x_1 - 5x_2 = -13$$
$$-2x_1 + 3x_2 = 9$$

$$A\boldsymbol{x} = b$$

$$A = \begin{bmatrix} 4 & -5 \\ -2 & 3 \end{bmatrix} \qquad \boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad b = \begin{bmatrix} -13 \\ 9 \end{bmatrix}$$

基本概念

- 由m个方程和n个未知变量构成的线性系统可表示为如下形式

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \cdots + a_{2n}x_n = b_2 \\ \cdots \\ a_{m1}x_1 + a_{m2}x_2 + a_{m3}x_3 + \cdots + a_{mn}x_n = b_m \end{cases}$$

a_{ij} 和 b_i 是给定的系数和常数项

x_j 是待求解的变量

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \in R^{m\times n} \qquad \boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \in R^n \qquad \boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} \in R^n$$

$$A\boldsymbol{x} = \boldsymbol{b}$$

矩阵计算及性质

加法

给定矩阵 $A \in R^{m\times n}$ 和 $B \in R^{m\times n}$ 则

$$A+B=\left(a_{ij}+b_{ij}\right)_{mn}$$

矩阵A与B相加就是两个矩阵对应元素相加

矩阵A与B能够相加的前提是这两个矩阵行列数(或者叫size)相同

数乘

给定实数 $\beta \in R$ 和矩阵 $A \in R^{m\times n}$ 则数乘：

$$\beta A=\left(\beta a_{ij}\right)_{mn}$$ 矩阵A中每个元素均与 β 相乘

矩阵计算及性质

向量-向量乘积

- 内积(inner product)

给定向量 $\boldsymbol{x},\boldsymbol{y}\in R^n$ 则

$$\langle \boldsymbol{x},\boldsymbol{y}\rangle=\boldsymbol{x}\cdot\boldsymbol{y}=\boldsymbol{x}^T\boldsymbol{y}=[x_1,x_2,\ldots,x_n]\begin{bmatrix}y_1\\y_2\\\vdots\\y_n\end{bmatrix}=\sum_{j=1}^{n}x_jy_j$$

- 外积(outer product)

给定向量 $\boldsymbol{x}\in R^m,\boldsymbol{y}\in R^n$ 则

$$\boldsymbol{x}\boldsymbol{y}^T=\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}[y_1,y_2,\ldots,y_n]=\begin{bmatrix}x_1y_1 & x_1y_2 & \cdots & x_1y_n\\x_2y_1 & x_2y_2 & \cdots & x_2y_n\\\vdots & \vdots & \ddots & \vdots\\x_ny_1 & x_ny_2 & \cdots & x_ny_n\end{bmatrix}$$

矩阵计算及性质

内积及正交性

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{y}, \boldsymbol{x} \rangle = \boldsymbol{x} \cdot \boldsymbol{y} = \boldsymbol{x} \cdot \boldsymbol{y} = \boldsymbol{x}^T \boldsymbol{y} = \boldsymbol{y}^T \boldsymbol{x}$$

$$= [x_1, x_2, \ldots, x_n] \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = [y_1, y_2, \ldots, y_n] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \sum_{j=1}^{n} x_j y_j$$

$$\langle \lambda \boldsymbol{x}, \boldsymbol{y} \rangle = \lambda \langle \boldsymbol{x}, \boldsymbol{y} \rangle$$

$$\langle \boldsymbol{x} + \boldsymbol{y}, \boldsymbol{z} \rangle = \langle \boldsymbol{x}, \boldsymbol{z} \rangle + \langle \boldsymbol{y}, \boldsymbol{z} \rangle$$

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle^2 \le \langle \boldsymbol{x}, \boldsymbol{x} \rangle \langle \boldsymbol{y}, \boldsymbol{y} \rangle$$

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle \le \|\boldsymbol{x}\| \|\boldsymbol{y}\|$$

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \|\boldsymbol{x}\| \|\boldsymbol{y}\| \cos\theta, \quad \theta = \arccos \frac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}{\|\boldsymbol{x}\| \|\boldsymbol{y}\|}$$

向量$\boldsymbol{x}$与向量$\boldsymbol{y}$垂直 $\Leftrightarrow \langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0$

如果n阶方阵A满足 $A^T A = I$ 或 $A^T = A^{-1}$ 方阵A称为正交矩阵

矩阵计算及性质

矩阵-向量乘积

给定 $A \in R^{m\times n}$ 和 $\boldsymbol{x} \in R^{n}$ 则

$$A\boldsymbol{x} = x_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix} + \cdots + x_n \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix}$$

$\boldsymbol{a}_{1\cdot}$

$\boldsymbol{a}_{\cdot 1}$

mn次乘法,
$m(n\text{-}1)$次加法
$O(mn)$

$A\boldsymbol{x}$ 可看作矩阵A各列的线性组合

<u>矩阵A的列数必须与$\boldsymbol{x}$的行数相同</u>

令 $\boldsymbol{b} = A\boldsymbol{x}$ 则列向量b的第i个元素为 $b_i = \langle \boldsymbol{a}_{i\cdot}, \boldsymbol{x} \rangle = \sum_{j=1}^{n} a_{ij} x_j$

向量$\boldsymbol{b}$的第i个元素为矩阵A第i行与列向量$\boldsymbol{x}$的内积

矩阵计算及性质

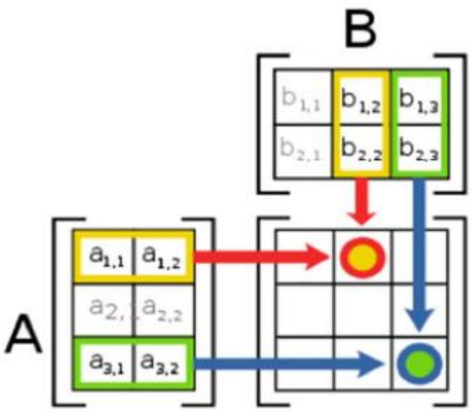
矩阵-矩阵乘积

给定 $A \in R^{m\times n}$ 和 $B \in R^{n\times p}$

$$AB = C = \left(c_{ij}\right) \in R^{m\times p}$$

$$c_{ij} = \sum_{k}^{n} a_{ik} b_{kj}$$

$$= \left\langle \boldsymbol{a}_{i\cdot}, \boldsymbol{b}_{\cdot j} \right\rangle, i = 1, \ldots, n \quad j = 1, \ldots, p$$



mnp次乘法,
$m(n\text{-}1)p$次加法
$O(mnp)$

矩阵A的列数等于矩阵B的行数

$AB \in R^{m\times p}$ 矩阵乘积C的行数为A的行数,C的列数为B的列数

矩阵计算及性质

矩阵乘法运算律

乘法结合律 $A \in R^{m\times n}, B \in R^{n\times p}, C \in R^{p\times q}$ 则 $A(BC)=(AB)C$

左分配律: $A, B \in R^{m\times n}, C \in R^{n\times p}$ 则 $(A+B)C = AC + BC$

右分配律: $A \in R^{m\times n}, B, C \in R^{n\times p}$ 则 $A(B+C) = AB + AC$

一般情况下 $AB \neq BA$

矩阵计算及性质

转置

$$A^T = \begin{bmatrix} a_{11} & a_{21} & a_{31} & \cdots & a_{m1} \\ a_{12} & a_{22} & a_{32} & \cdots & a_{m2} \\ a_{13} & a_{23} & a_{33} & \cdots & a_{m3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & a_{3n} & \cdots & a_{mn} \end{bmatrix} \in R^{n\times m} \qquad \left(A^T\right)^T = A$$

$$\boldsymbol{x}^T = \begin{bmatrix} x_1 & x_2 & x_3 & \cdots & x_m \end{bmatrix} \in R^{1\times m} \qquad \left(\boldsymbol{x}^T\right)^T = \boldsymbol{x}$$

$$(AB)^T = B^T A^T \qquad (A+B)^T = A^T + B^T \qquad \begin{bmatrix} A & B \end{bmatrix}^T = \begin{bmatrix} A^T \\ B^T \end{bmatrix}$$

矩阵计算及性质

特殊类型矩阵

下矩阵

$$L=\begin{bmatrix} \ell_{11} & 0 & 0 & 0 \\ \ell_{21} & \ell_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \ell_{n1} & \ell_{n2} & \cdots & \ell_{nn} \end{bmatrix} \in R^{n\times n}$$

上矩阵

$$U=\begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ 0 & u_{22} & \cdots & u_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & u_{nn} \end{bmatrix} \in R^{n\times n}$$

矩阵计算及性质

特殊类型矩阵

对角阵

$$D=\begin{bmatrix} d_{11} & 0 & 0 & 0 \\ 0 & d_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_{nn} \end{bmatrix} \in R^{n\times n}$$

单位阵

$$I=\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \in R^{n\times n}$$

对于任意对称矩阵 $A \in R^{n\times n}$

$$AI = IA = A$$

对称阵　　当 $A^T = A$ 时，称矩阵A为对称阵,对称阵是方阵

矩阵计算及性质

矩阵(方阵)的迹

$$A=\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{bmatrix} \in R^{n\times n}$$

方阵A的迹为对角线元素之和

$$trA=\sum_{i=1}^{n} a_{ii}$$

方阵迹的性质： $trA=trA^T$

$A,B\in R^{n\times n}$ $\quad tr(A+B)=trA+trB$

$A\in R^{n\times n},\tau\in R$ $\quad tr(\tau A)=\tau trA$

$A\in R^{m\times n},B\in R^{n\times m},AB\in R^{m\times m},BA\in R^{n\times n}$ $\quad trAB=trBA$

矩阵计算及性质

向量的线性无关

为表示方便起见，将矩阵A表示为列向量形式

$A=[\boldsymbol{a}_1,\boldsymbol{a}_2,\ldots,\boldsymbol{a}_n]$,其中 $\boldsymbol{a}_j\in R^m$ 表示列向量, 则前面提及的 $A\boldsymbol{x}=\boldsymbol{b}$

可写成标量与向量的乘积之和

$$\boldsymbol{a}_1x_1+\boldsymbol{a}_2x_2+\cdots+\boldsymbol{a}_nx_n=\boldsymbol{b}$$

上式称为列向量 $\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_n$ 的线性组合

一组m维向量 $\{\boldsymbol{a}_1,\boldsymbol{a}_2,\ldots,\boldsymbol{a}_n\}$ 构成的方程 $x_1\boldsymbol{a}_1+x_2\boldsymbol{a}_2+\ldots+x_n\boldsymbol{a}_n=\boldsymbol{0}$

只有零解 $x_1=x_2=\ldots=x_n=0$ 则称向量组**线性无关**

否则称向量组**线性相关**

矩阵计算及性质

空间概念

前面第18页讲了矩阵A的列空间,下面介绍线性代数中涉及到的其他空间相关概念:

以向量为元素的集合V称为向量空间

若$S=\{\boldsymbol{u}_1,\boldsymbol{u}_2,\ldots,\boldsymbol{u}_m\}$是向量空间$V$的向量子集合,则$\boldsymbol{u}_1,\boldsymbol{u}_2,\ldots,\boldsymbol{u}_m$的所有线性组合的集合$W$称为由$\boldsymbol{u}_1,\boldsymbol{u}_2,\ldots,\boldsymbol{u}_m$张成的子空间,定义为

$$W=Span\{\boldsymbol{u}_1,\boldsymbol{u}_2,\ldots,\boldsymbol{u}_m\}=\{\boldsymbol{u}\mid\boldsymbol{u}=a_1\boldsymbol{u}_1+a_2\boldsymbol{u}_2+\ldots+a_m\boldsymbol{u}_m\}$$

矩阵计算及性质

空间概念

在向量子空间W中,如果存在p个($p \le m$)向量$\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_p$ 满足:

(1) $\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_p$ **线性无关**

注意此处与上页比较$p \le m$

(2) W中任一元素$\boldsymbol{u}$总可由 $\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_p$ 线性表示

则 $\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_p$ 就称为向量子空间W的一个基

子空间W的任何一组基的向量个数称为W的维数, $\dim(W)$

矩阵计算及性质

矩阵的秩

设向量组 $A_0: \boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_r$ 是向量组 $A: \boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$ 的一部分$(r \le n)$

且满足:

(1) 向量组A_0线性无关

(2) 向量组A的任一向量都能由向量组A_0线性表示

则向量组A_0就是向量组A的一个最大线性无关组，最大线性无关组所含向量的<u>**个数**</u>r称为向量组A的**秩**

矩阵计算及性质

矩阵的秩

给定 $A \in R^{m\times n}, \boldsymbol{x}, \boldsymbol{b} \in \boldsymbol{R}^n$ 矩阵方程 $A\boldsymbol{x} = \boldsymbol{b}$ 可分为三种类型:

(1) 适定方程: 若m=n,且rank(A)=n, 即矩阵A非奇异,则称矩阵方程 $A\boldsymbol{x} = \boldsymbol{b}$ 为适定(well-determined)方程. 方程解是唯一的$\boldsymbol{x} = A^{-1}\boldsymbol{b}$

(2) 欠定方程: 若m<rank(A),独立方程个数小于独立未知数个数 则称矩阵方程 $A\boldsymbol{x} = \boldsymbol{b}$ 为欠定(under-determined)方程. 方程无穷多解

(3) 超定方程: 若m>rank(A),独立方程个数大于独立未知数个数 则称矩阵方程 $A\boldsymbol{x} = \boldsymbol{b}$ 为超定(over-determined)方程. 方程无精确解

矩阵计算及性质

矩阵的秩

矩阵的秩定义为该矩阵中线性无关的列向量的数目

若 $rank(A)=r$ 则矩阵A有r个线性无关的列向量,这r个线性无关列向量的所有线性组合,便构成了一个向量空间,称为矩阵A的列空间Col(A)

矩阵A的列空间Col(A) 的维数定义为该矩阵的秩, 即

$$r=rank(A)=\dim\left[\mathrm{Col}(A)\right]$$

若矩阵A的秩为r,则该矩阵的列空间Col(A)是一个r维子空间

矩阵计算及性质

矩阵的秩

令 $r_A = rank(A), r_B = rank(B)$ 则乘积矩阵AB的秩具有下面性质:

$$rank(AB) = r_{AB} \le \min\{r_A, r_B\}$$

矩阵 $A \in R^{m\times n}$ 左乘非奇异方阵 $P \in R^{m\times m}$ 或者右乘非奇异方阵 $Q \in R^{n\times n}$ 将不改变A的秩

$$rank(A+B) \le rank([A,B]) \le rank(A) + rank(B)$$

矩阵计算及性质

可逆与非奇异

给定$n\times n$方阵A,若存在一个$n\times n$方阵B满足：$AB=BA=I$

则称方阵A是可逆的(或非奇异的),方阵B称为A的逆矩阵,记为A^{-1}

n阶方阵A是可逆的 $\Leftrightarrow$ n阶方阵A是非奇异的 $\Leftrightarrow rank(A)=n$

$\Leftrightarrow \det(A)\neq 0$

n阶方阵A的逆矩阵具有以下性质： A^{-1} 是唯一的 $\det\left(A^{-1}\right)=\dfrac{1}{\det(A)}$

$AA^{-1}=A^{-1}A=I$ $\quad$ A为正交矩阵 $\Leftrightarrow A^{-1}=A^{T}$

$\left(A^{-1}\right)^{-1}=A$ $\quad$ $\left(A^{-1}\right)^{T}=\left(A^{T}\right)^{-1}$ $\quad$ $(AB)^{-1}=B^{-1}A^{-1}$

若 $A=diag\left(a_{11},a_{22},\cdots,a_{nn}\right)$ 则$A^{-1}=diag\left(a_{11}^{-1},a_{22}^{-1},\cdots,a_{nn}^{-1}\right)$

矩阵计算及性质

方阵的特征值和特征向量

设n阶方阵A,如果数λ 和n维列向量$\boldsymbol{x}$ 满足

$$A\boldsymbol{x}=\lambda\boldsymbol{x}$$

则这样的数λ 称为方阵A的特征值,非零向量$\boldsymbol{x}$称为方阵A的对应于特征值λ 的特征向量.

$A\boldsymbol{x}=\lambda\boldsymbol{x}\Leftrightarrow(A-\lambda I)\boldsymbol{x}=\boldsymbol{0}$ 存在非零解 $\boldsymbol{x}\neq\boldsymbol{0}$ 的唯一条件是矩阵 $A-\lambda I$ 的行列式等于零,即$\det(A-\lambda I)=\boldsymbol{0}\Leftrightarrow A-\lambda I$ 方阵奇异

设n阶矩阵 $A=(a_{ij})$ 的特征值为 $\lambda_1,\lambda_2\ldots,\lambda_n$,具备下列性质:

(1) $\lambda_1+\lambda_2+\ldots+\lambda_n=a_{11}+a_{22}+\ldots+a_{nn}$

(2) $\lambda_1\lambda_2\cdots\lambda_n=\det(A)$

矩阵计算及性质

方阵的特征值和特征向量

(3) 方阵A奇异,则其特征值至少有一个为0

(4) 方阵A与A^T具有相同的特征值

(5) λ 是方阵A的特征值,则λ也是A^T的特征值

(6) λ 是方阵A的特征值,则 $\lambda+\sigma^2$ 是 $A+\sigma^2 I$ 的特征值

(7) $\lambda(\neq 0)$是方阵A的特征值,则$1/\lambda$是A^{-1}的特征值

矩阵计算及性质

相似矩阵

两个n阶方阵A和B为相似矩阵当且仅当存在一个n阶可逆方阵P使得

$$P^{-1}AP = B$$

方阵P称为A与B之间的相似变换矩阵,方阵A相似B记做:$A \sim B$

相似矩阵的性质:

$A \sim A$

$A \sim B \Leftrightarrow B \sim A$

$A \sim B \Rightarrow rank(A) = rank(B), \det(A) = \det(B), tr(A) = tr(B)$

$A \sim B \Rightarrow$ A与B特征值相同,特征向量未必相同

A与对角阵相似,则称A为可对角化矩阵

n阶矩阵A与对角矩阵相似的充分必要条件为矩阵A有n个线性无关的特征向量

矩阵计算及性质

矩阵的等价、相似与合同

矩阵等价: $A, B \in R^{m\times n}$ 如果两矩阵满足 $Q^{-1}AP = B$

其中, P是n阶可逆方阵,Q是m阶可逆方阵

两个同型矩阵A与B当且仅当他们秩相同的时候,两个矩阵是等价的

矩阵合同: 设A,B均为n阶方阵,若存在n阶可逆矩阵C使得 $C^T AC = B$

则称方阵A与B合同,记做 $A \simeq B$

矩阵计算及性质

二次型

n阶实<u>对称阵</u>A的二次型定义为 $f = \mathbf{x}^T A\mathbf{x}$

<u>对称阵</u>A叫做二次型f的矩阵, f叫做对称阵A的二次型

任给一个二次型就能唯一地确定一个对称阵;反之任给一个对称阵也能唯一确定一个二次型

正定、半正定矩阵

对于n阶对称方阵A

正定矩阵: 二次型 $\mathbf{x}^T A\mathbf{x} > 0, \forall \mathbf{x} \neq \mathbf{0}$ (对称方阵A所有特征值均为正数)

半正定矩阵: 二次型 $\mathbf{x}^T A\mathbf{x} \geq 0, \forall \mathbf{x} \neq \mathbf{0}$ (对称方阵A所有特征值均为非负数)

矩阵计算及性质

向量的范数(vector norm)

$\boldsymbol{x} \in R^n$

ℓ_1 范数 $\|\boldsymbol{x}\|_1 = \sum_{j=1}^{n} |x_j|$

ℓ_2 范数 $\|\boldsymbol{x}\|_2 = \left(x_1^2 + x_2^2 + \cdots + x_n^2\right)^{\frac{1}{2}}$

ℓ_∞ 范数 $\|\boldsymbol{x}\|_\infty = \max\left(|x_1|, \ldots, |x_n|\right)$ 无穷范数

ℓ_p 范数 $\|\boldsymbol{x}\|_p = \left(\sum_{j=1}^{n} |x_j|^p\right)^{1/p}$

矩阵计算及性质

矩阵的范数(matrix norm)

$A=\left(a_{ij}\right)\in R^{m\times n}$

F范数 $\|\boldsymbol{A}\|_F=\left(\sum_{i=1}^{m}\sum_{j=1}^{n}\left|a_{ij}\right|^2\right)^{1/2}$

ℓ_p 范数 $\|A\|_p=\max_{\boldsymbol{x}\neq\boldsymbol{0}}\frac{\|A\boldsymbol{x}\|_p}{\|\boldsymbol{x}\|_p}$

2范数(谱范数)

$\|A\|_2=\sigma_{\max}=\sqrt{\lambda_{\max}}$, $\lambda_{\max}$为A^TA的最大特征值, $\sigma_{\max}$为方阵A最大奇异值

行和范数(无穷范数) $\|A\|_\infty=\max_{1\le i\le m}\left\{\sum_{j=1}^{n}\left|a_{ij}\right|\right\}$

矩阵范数的性质:

$\|cA\|=|c|\|A\|$ $\quad\|A+B\|\le\|A\|+\|B\|$ $\quad\|AB\|\le\|A\|\|B\|$

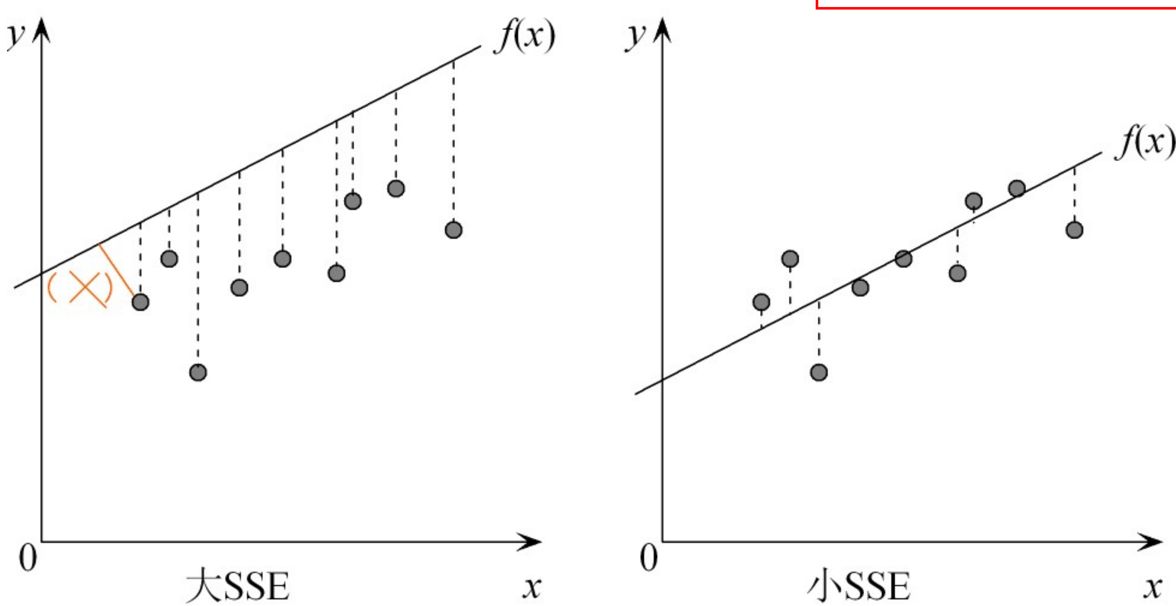
最小二乘法

基本引导

给定数据对序列 $(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)$ 构建**线性回归模型** $\hat{y} = kx + b$

残差(residual): $y_i - \hat{y}_i = y_i - (kx_i + b)$

SSE(sum squares of error): $SSE = \sum_{i=1}^{m} (y_i - \hat{y}_i)^2$



最小二乘法就是要求使SSE最小的直线模型

$$\min \sum_{i=1}^{m} (y_i - \hat{y}_i)^2$$

最小二乘法

$$\frac{\partial SSE}{\partial k}=0$$

$$\frac{\partial SSE}{\partial b}=0$$

$$SSE=\sum_{i=1}^{m}(y_i-\hat{y}_i)^2=\sum_{i=1}^{m}\left(y_i-(kx_i+b)\right)^2=\sum_{i=1}^{m}(y_i-kx_i-b)^2$$

$$\frac{\partial SSE}{\partial b}=\frac{\partial\sum_{i=1}^{m}(y_i-kx_i-b)^2}{\partial b}=\sum_{i=1}^{m}2(y_i-kx_i-b)(-1)$$

$$=\sum_{i=1}^{m}y_i-k\sum_{i=1}^{m}x_i-\sum_{i=1}^{m}b=\sum_{i=1}^{m}y_i-k\sum_{i=1}^{m}x_i-mb=0$$

$$\Rightarrow \boxed{b=\frac{1}{m}\sum_{i=1}^{m}y_i-k\frac{1}{m}\sum_{i=1}^{m}x_i=\bar{y}-k\bar{x}}$$

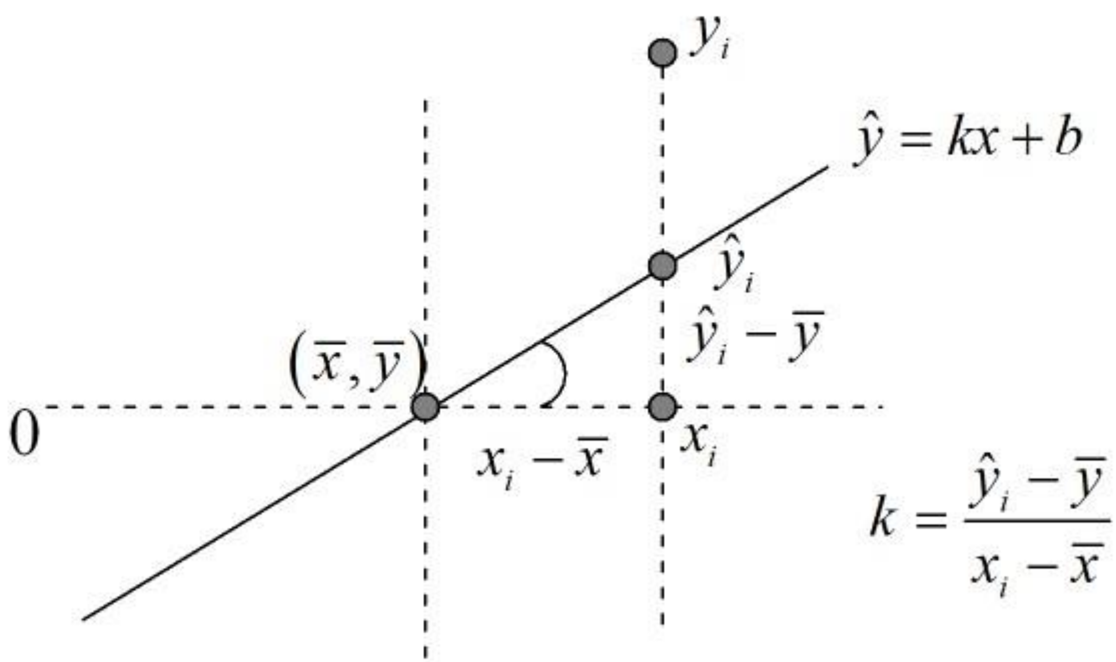
(1) 直线偏移求解公式

$$\Rightarrow \bar{y}=k\bar{x}+b$$

表明 $(\bar{x},\bar{y})$ 为直线 $\hat{y}=kx+b$ 上的点

最小二乘法

$(\overline{x},\overline{y})$ 为直线 $\hat{y}=kx+b$ 上的点 $\Rightarrow k=\dfrac{\hat{y}_i-\overline{y}}{x_i-x}$



最小二乘法

$$k=\frac{\hat{y}_i-\bar{y}}{x_i-\bar{x}}=\frac{(y_i-\bar{y})-(y_i-\hat{y}_i)}{x_i-\bar{x}}\Rightarrow (y_i-\bar{y})-(y_i-\hat{y}_i)=k(x_i-\bar{x})$$

$$\Rightarrow (y_i-\hat{y}_i)=(y_i-\bar{y})-k(x_i-\bar{x})\Rightarrow SSE=\sum_{i=1}^{m}(y_i-\hat{y}_i)^2=\sum_{i=1}^{m}\left[(y_i-\bar{y})-k(x_i-\bar{x})\right]^2$$

$$\frac{\partial SSE}{\partial k}=\frac{\partial\sum_{i=1}^{m}\left[(y_i-\bar{y})-k(x_i-\bar{x})\right]^2}{\partial k}=-2\sum_{i=1}^{m}\left[(y_i-\bar{y})-k(x_i-\bar{x})\right](x_i-\bar{x})=0$$

$$\Rightarrow\sum_{i=1}^{m}(y_i-\bar{y})(x_i-\bar{x})-k\sum_{i=1}^{m}(x_i-\bar{x})^2=0\Rightarrow \boxed{k=\frac{\sum_{i=1}^{m}(y_i-\bar{y})(x_i-\bar{x})}{\sum_{i=1}^{m}(x_i-\bar{x})^2}}$$

(2) 直线斜率求解公式

最小二乘法

如何找到使得SSE最小的直线(线性回归模型)

给定数据对序列 $(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)$ 构建**线性回归模型** $\hat{y} = kx + b$

Step1 根据已有数据求解平均值 $\bar{x}, \bar{y}$

Step2根据已有数据和已计算的平均值求直线斜率 $$k = \frac{\sum_{i=1}^{m}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{m}(x_i - \bar{x})^2}$$

Step3 根据平均值和斜率计算偏移 $b = \bar{y} - k\bar{x}$

最小二乘法

最小二乘法与最大似然法之间的联系

最小二乘法是残差满足正态分布情况下的最大似然估计

$$\varepsilon_i = y_i - \hat{y}_i = y_i - (kx_i + b) \qquad y_i = (kx_i + b) + \varepsilon_i$$

令残差(residual)服从正态分布 $\varepsilon_i \sim N\left(0, \sigma^2\right)$,则残差的概率密度函数如下

$$f\left(\varepsilon_i \middle| k, b\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\varepsilon_i^2}{2\sigma^2}\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y_i - \hat{y}_i)^2}{2\sigma^2}\right)$$

进而定义似然函数

$$L(k, b, \varepsilon_1, \varepsilon_2, \ldots, \varepsilon_m) = \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y_i - \hat{y}_i)^2}{2\sigma^2}\right)$$

最小二乘法

最小二乘法与最大似然法之间的联系

对数似然函数 $$\ell(k,b,\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_m)=\ln L(k,b,\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_m)$$
$$=\underbrace{\sum_{i=1}^{m}\ln\frac{1}{\sqrt{2\pi\sigma}}}_{\text{constant}}-\frac{1}{2\sigma^2}\sum_{i=1}^{m}(y_i-\hat{y}_i)^2$$

最大似然函数 $$\max \ell(k,b,\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_m)\Leftrightarrow \min\sum_{i=1}^{m}(y_i-\hat{y}_i)^2$$

最大似然估计等价于最小二乘

最小二乘法

几何意义前导——矩阵列空间

矩阵 $A \in R^{m\times n}$ 的列空间Col(A)是A阵列向量所有线性组合的集合

如果 $A=[\boldsymbol{a}_1,\boldsymbol{a}_2,\ldots,\boldsymbol{a}_n]$ 则列空间Col(A)可表示为

$$\mathrm{Col}(A)=Span\{\boldsymbol{a}_1,\boldsymbol{a}_2,\ldots,\boldsymbol{a}_n\}$$

在列空间Col(A)中的任一向量可以写成$A\boldsymbol{x}$, 因为$A\boldsymbol{x}$表示矩阵A的列向量的线性组合

$$\mathrm{Col}(A)=\{\boldsymbol{b}\,|\,\boldsymbol{b}=A\boldsymbol{x},\forall \boldsymbol{x}\in R^n\}$$

最小二乘法

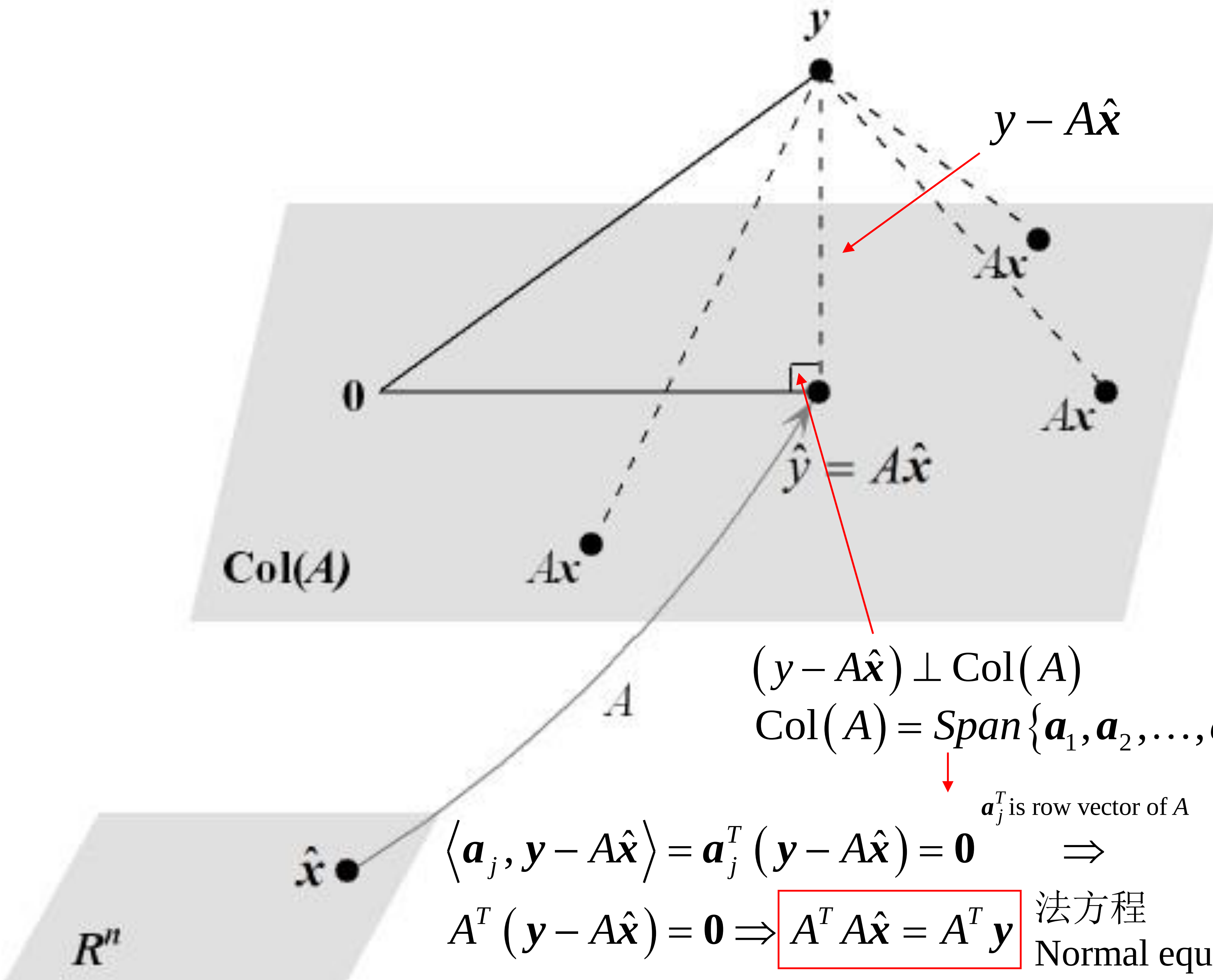
几何意义以及矩阵形式求解

线性方程组的矩阵形式 $A\boldsymbol{x}=\boldsymbol{y}$ 实际应用中,很难找到"极端精确"的方程解 $A\boldsymbol{x}$ 可认为是$\boldsymbol{y}$ 的近似解, 矩阵形式的最小二乘问题定义如下:

给定$A\in R^{m\times n},\boldsymbol{y}\in R^{m}$, 求方程组$A\boldsymbol{x}=\boldsymbol{y}$ 的最小二乘解就是找到满足下面不等式的向量 $\hat{\boldsymbol{x}}\in R^{n}$

$$\|\boldsymbol{y}-A\hat{\boldsymbol{x}}\|\le\|\boldsymbol{y}-A\boldsymbol{x}\|,\forall\boldsymbol{x}\in R^{n}$$

$\forall\boldsymbol{x}\in R^{n},A\boldsymbol{x}$都是列空间$\mathrm{Col}(A)$中的点, 最小二乘法就是找到合适的$\boldsymbol{x}$ 使得 $A\boldsymbol{x}$ 成为列空间$\mathrm{Col}(A)$中距离$\boldsymbol{y}$最近的点



$$\left(y - A\hat{\boldsymbol{x}}\right) \perp \mathrm{Col}\left(A\right)$$

$$\mathrm{Col}\left(A\right) = Span\left\{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n\right\}$$

$\boldsymbol{a}_j^T$ is row vector of A

$$\left\langle \boldsymbol{a}_j, \boldsymbol{y} - A\hat{\boldsymbol{x}} \right\rangle = \boldsymbol{a}_j^T \left(\boldsymbol{y} - A\hat{\boldsymbol{x}}\right) = \boldsymbol{0} \quad \Rightarrow$$

$$A^T\left(\boldsymbol{y} - A\hat{\boldsymbol{x}}\right) = \boldsymbol{0} \Rightarrow \boxed{A^T A\hat{\boldsymbol{x}} = A^T \boldsymbol{y}}$$

法方程

Normal equations

最小二乘法

线性方程$A\boldsymbol{x}=\boldsymbol{y}$ 的最小二乘解与法方程 $A^T A\boldsymbol{x}=A^T\boldsymbol{y}$ 的解一致

<u>当且仅当</u>矩阵A的列线性无关(即列满秩)时, 方阵$A^T A$ 可逆

当方阵$A^T A$ 可逆时,方程 $A\boldsymbol{x}=\boldsymbol{y}$ 有<u>**唯一**</u>的最小二乘解:

$$\hat{\boldsymbol{x}}=\left(A^T A\right)^{-1}A^T\boldsymbol{y}$$

最小二乘解是一种近似解, $A\hat{\boldsymbol{x}}=\hat{\boldsymbol{y}}$ 是对 $\boldsymbol{y}$ 的一种逼近,最小二乘误差 $\left\|\boldsymbol{y}-A\hat{\boldsymbol{x}}\right\|$ 用来衡量逼近效果

作业

- 参考下面链接，了解python (numpy)与今天讲到的线代和矩阵相关概念
- 参考链接：
  - https://github.com/scalanlp/breeze/wiki/Linear-Algebra-Cheat-Sheet
  - https://docs.scipy.org/doc/numpy-1.13.0/reference/routines.linalg.html
  - http://blog.csdn.net/airfish20000/article/details/60880932
  - http://blog.csdn.net/bitcarmanlee/article/details/51945271

感谢大家！

恳请大家批评指正！