以机器学习分析矩阵运算

七月算法 邹博

2015年10月18日

线性代数

- 定义：方阵的行列式
  - 1阶方阵的行列式为该元素本身
  - n阶方阵的行列式等于它的任一行(或列)的各元素与其对应的代数余子式乘积之和。

范德蒙行列式Vandermonde

- 证明范德蒙行列式Vandermonde：

$$D_n=\begin{vmatrix} 1 & 1 & 1 & \cdots & 1 \\ x_1 & x_2 & x_3 & \cdots & x_n \\ x_1^2 & x_2^2 & x_3^2 & \cdots & x_n^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_1^{n-1} & x_2^{n-1} & x_3^{n-1} & \cdots & x_n^{n-1} \end{vmatrix}=\prod_{n\ge i\ge j\ge 1}\left(x_i-x_j\right)$$

  - 提示：数学归纳法
  - 注：参考Lagrange/Newton插值法

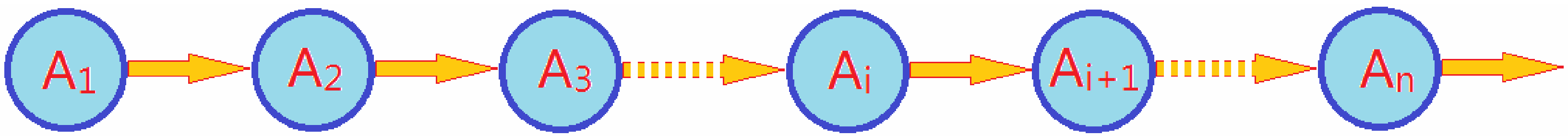
矩阵的乘法

- A为m×s阶的矩阵，B为s×n阶的矩阵，那么，C=A×B是m×n阶的矩阵，其中，

$$c_{ij} = \sum_{k=1}^{s} a_{ik} b_{kj}$$

矩阵模型

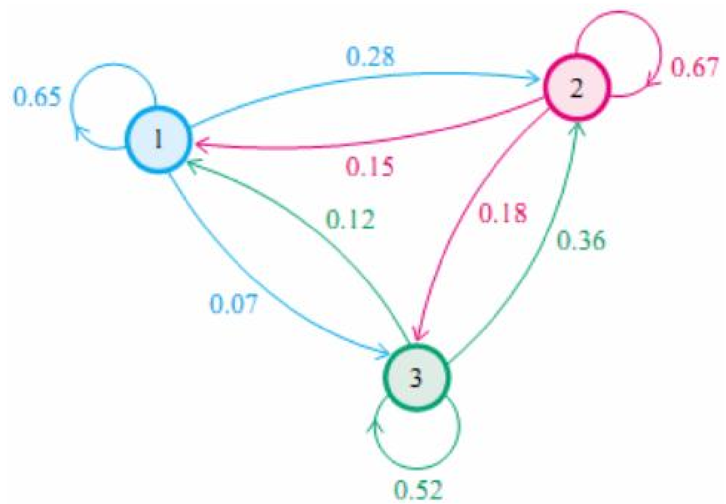
- 考虑某随机过程π，它的状态有n个，用1~n表示。记在当前时刻t时位于i状态，它在t+1时刻位于j状态的概率为P(i,j)=P(j|i)：
  - 即状态转移的概率只依赖于前一个状态。



举例

- 假定按照经济状况将人群分成上、中、下三个阶层，用1、2、3表示。假定当前处于某阶层只和上一代有关，即：考察父代为第i阶层，则子代为第j阶层的概率。假定为如下转移概率矩阵：

$$
\begin{array}{cc} P & \text{子代} \\ \text{父代} & \begin{bmatrix} 0.65 & 0.28 & 0.07 \\ 0.15 & 0.67 & 0.18 \\ 0.12 & 0.36 & 0.52 \end{bmatrix} \end{array}
$$



概率转移矩阵

- 第n+1代中处于第j个阶层的概率为：

$$\pi\left(X_{n+1}=j\right)=\sum_{i=1}^{n}\pi\left(X_{n}=i\right)\cdot P\left(X_{n+1}=j\mid X_{n}=i\right)$$

$$\Rightarrow \pi^{(n+1)}=\pi^{(n)}\cdot P$$

- 因此，矩阵P即为(条件)概率转移矩阵。
  - 第i行元素表示：在上一个状态为i时的分布概率，即：每一行元素的和为1。
- 思考：初始概率分布π对最终分布的影响？

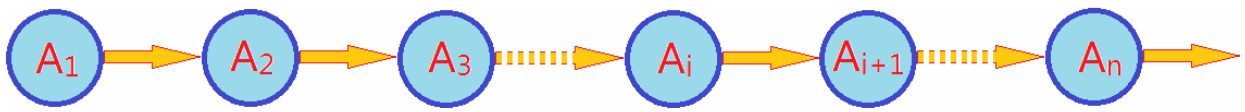
探索：初始概率 $\pi = [0.21, 0.68, 0.1]$迭代

第n代	第1阶层	第2阶层	第3阶层
0	0.21	0.68	0.11
1	0.252	0.554	0.194
2	0.27	0.512	0.218
3	0.278	0.497	0.225
4	0.282	0.49	0.226
5	0.285	0.489	0.225
6	0.286	0.489	0.225
7	0.286	0.489	0.225
8	0.286	0.488	0.225
9	0.286	0.489	0.225
10	0.286	0.489	0.225

初始概率 $\pi = [0.75, 0.15, 0.1]$ 的迭代结果

第n代	第1阶层	第2阶层	第3阶层
0	0.75	0.15	0.1
1	0.522	0.347	0.132
2	0.407	0.426	0.167
3	0.349	0.459	0.192
4	0.318	0.475	0.207
5	0.303	0.482	0.215
6	0.295	0.485	0.22
7	0.291	0.487	0.222
8	0.289	0.488	0.225
9	0.286	0.489	0.225
10	0.286	0.489	0.225

平稳分布



- 初始概率不同，但经过若干次迭代， π 最终稳定收敛在某个分布上。
- 从而，这是转移概率矩阵P的性质，而非初始分布的性质。事实上，上述矩阵P的n次幂，每行都是(0.286,0.489,0.225)，n>20
- 如果一个非周期马尔科夫随机过程具有转移概率矩阵P，且它的任意两个状态都是连通的，则 $\lim_{n\to\infty} P_{ij}^n$ 存在，记做 $\lim_{n\to\infty} P_{ij}^n = \pi(j)$。

平稳分布

- 事实上，下面两种写法等价：

$$\lim_{n\to\infty} P_{ij}^n = \pi(j) \qquad \lim_{n\to\infty} P^n = \begin{bmatrix} \pi(1) & \pi(2) & \dots & \pi(n) \\ \pi(1) & \pi(2) & \dots & \pi(n) \\ \vdots & \vdots & \ddots & \vdots \\ \pi(1) & \pi(2) & \cdots & \pi(n) \end{bmatrix}$$

- 同时，若某概率分布 $\pi P=\pi$，说明
  - 该多项分布 π 是状态转移矩阵P的平稳分布；
  - 线性方程 $xP=x$ 的非负解为 π，而 P^n 唯一，因此 π 是线性方程 $xP=x$ 的唯一非负解。
  - 该问题将在马尔科夫模型中继续探讨。

思考

- 根据定义来计算 $C=A\times B$，需要m*n*s次乘法。
  - 即：若A、B都是n阶方阵，C的计算时间复杂度为$O(n^3)$
  - 问：可否设计更快的算法？
- 三个矩阵A、B、C的阶分别是$a_0\times a_1, a_1\times a_2, a_2\times a_3$，从而$(A\times B)\times C$和$A\times(B\times C)$的乘法次数是$a_0a_1a_2+a_0a_2a_3$、$a_1a_2a_3+a_0a_1a_3$，二者一般情况是不相等的。
  - 问：给定n个矩阵的连乘积：$A_1\times A_2\times A_3\ldots\ldots\times A_n$，如何添加括号来改变计算次序，使得乘法的计算量最小？

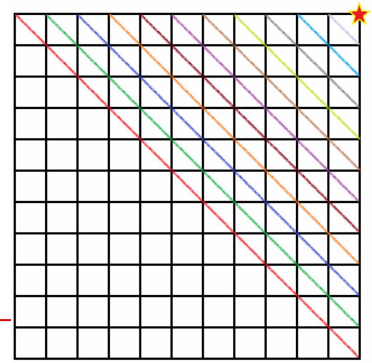
解

- 矩阵乘法C=A×B优化问题
  - 分治法
- 矩阵连乘的加括号最优策略
  - 动态规划
- 属于算法的经典问题，将在姊妹班“算法班”中做进一步探讨。

附：矩阵连乘的提法

- 给定n个矩阵$\{A_1,A_2,\ldots,A_n\}$，其中A_i与A_{i+1}是可乘的，i=1，2…，n-1。考察该n个矩阵的连乘积：$A_1\times A_2\times A_3\ldots\ldots\times A_n$，确定计算矩阵连乘积的计算次序，使得依此次序计算矩阵连乘积需要的乘法次数最少。
  - 即：利用结合律，通过加括号的方式，改变计算过程，使得数乘的次数最少。

矩阵连乘问题从算法到实现

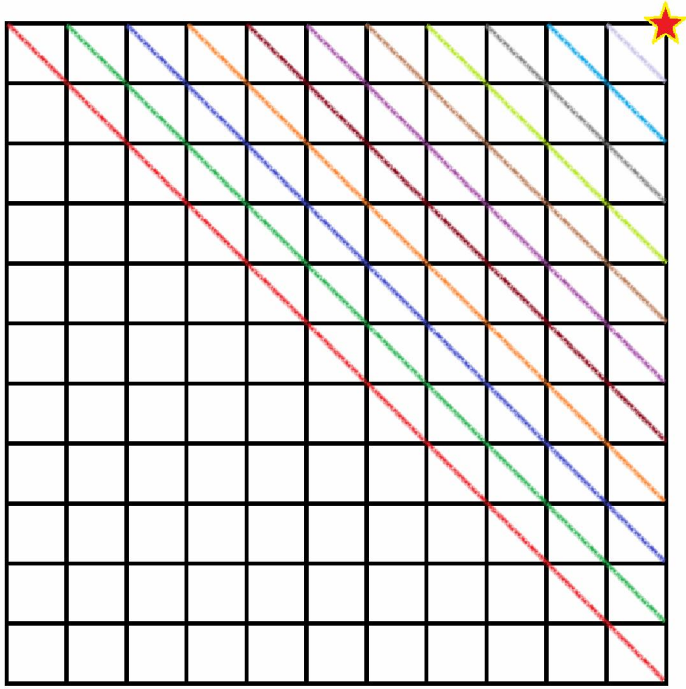


- 由m[i,j]的递推关系式可以看出，在计算m[i,j]时，需要用到m[i+1,j]，m[i+2,j]…m[j-1,j]；

$$m[i,j]=\begin{cases} 0 & i=j \\ \min\limits_{i\le k<j}\{m[i,k]+m[k+1,j]+p_{i-1}p_kp_j\} & i<j \end{cases}$$

- 因此，求m[i,j]的前提，不是m[0…i-1;0…j-1]，而是沿着主对角线开始，依次求取到右上角元素。
- 因为m[i,j]一个元素的计算，最多需要遍历n-1次，共O(n^2)个元素，故算法的时间复杂度是O(n^3)，空间复杂度是O(n^2)。

Code



```
//p[0...n]存储了n+1个数，其中，(p[i-1],p[i])是矩阵i的阶；
//s[i][j]记录A[i...j]从什么位置断开；m[i][j]记录数乘最小值
void MatrixMultiply(int* p, int n, int** m, int** s)
{
    int r,i,j,k,t;
    for(i = 1; i <= n; i++)
        m[i][i] = 0;

    //r个连续矩阵的连乘：上面的初始化，相当于r=1
    for(r = 2; r <= n; r++)
    {
        for(i = 1; i <= n-r+1; i++)
        {
            j=i+r-1;
            m[i][j] = m[i+1][j]+ p[i-1]*p[i]*p[j];
            s[i][j] = i;
            for(k = i+1; k < j; k++)
            {
                t = m[i][k] + m[k+1][j] + p[i-1]*p[k]*p[j];
                if(t < m[i][j])
                {
                    m[i][j] = t;
                    s[i][j] = k;
                }
            }
        }
    }
}
```

矩阵连乘问题的进一步思考

- n个矩阵连乘，可以分解成i个矩阵连乘和(n-i)个矩阵连乘，最后，再将这两个矩阵相乘。故：

$$P(n)=\begin{cases}1 & n=1\\ \sum_{k=1}^{n-1}P(k)P(n-k) & n>1\end{cases}\Rightarrow P(n)=\Omega\left(\frac{4^n}{\sqrt{\pi}*n^{3/2}}\right)$$

$$P(n)=\frac{1}{n}C_{2n-2}^{n-1}$$

- 1, 1, 2, 5, 14, 42, 132, 429, 1430, 4862, 16796, 58786, 208012, 742900, 2674440, 9694845, 35357670, 129644790, 477638700, 1767263190, 6564120420, 24466267020, 91482563640, 343059613650, 1289904147324, 4861946401452……

矩阵的秩

- 在m×n矩阵A中，任取k行k列，不改变这k^2个元素在A中的次序，得到k阶方阵，称为矩阵A的k阶子式。
  - 显然，m×n矩阵A的k阶子式有$C_m^k C_n^k$个。
- 设在矩阵A中有一个不等于0的r阶子式D，且所有r+1阶子式(如果存在的话)全等于0，那么，D称为矩阵A的最高阶非零子式，r称为矩阵A的秩，记做R(A)=r。
  - n×n的可逆矩阵，秩为n
  - 可逆矩阵又称满秩矩阵
  - 矩阵的秩等于它行(列)向量组的秩

秩与线性方程组的解的关系

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

$$Ax = b$$

- 对于n元线性方程组Ax=b,
  - 无解的充要条件是R(A)<R(A,b)
  - 有唯一解的充要条件是R(A)=R(A,b)=n
  - 有无限多解的充要条件是R(A)=R(A,b)<n

推论

- $Ax=0$有非零解的充要条件是$R(A)<n$
- $Ax=b$有解的充要条件是$R(A)=R(A,b)$

向量组等价

- 向量b能由向量组A:$a_1,a_2,\dots,a_m$线性表示的充要条件是矩阵A=$(a_1,a_2,\dots a_m)$的秩等于矩阵B=$(a_1,a_2,\dots a_m,b)$的秩。
- 设有两个向量组A:$a_1,a_2,\dots,a_m$及B:$b_1,b_2,\dots,b_n$，若B组的向量都能由向量组A线性表示，则称向量组B能由向量组A线性表示。若向量组A与向量组B能相互线性表示，则称两个向量组等价。

系数矩阵

- 把向量组A和B所构成的矩阵依次记做 $A=(a_1,a_2,\ldots,a_m)$和$B=(b_1,b_2,\ldots,b_n)$，B组能由A组线性表示，即对每个向量b_j，存在 $k_{1j},k_{2j},\cdots k_{mj}$
- 使得

$$b_j = k_{1j}a_1 + k_{2j}a_2 + \cdots + k_{mj}a_m = \begin{pmatrix} a_1 & a_2 & \cdots & a_m \end{pmatrix} \begin{pmatrix} k_{1j} \\ k_{2j} \\ \vdots \\ k_{mj} \end{pmatrix}$$

- 从而得到系数矩阵K

$$\begin{pmatrix} b_1 & b_2 & \cdots & b_n \end{pmatrix} = \begin{pmatrix} a_1 & a_2 & \cdots & a_m \end{pmatrix} \begin{pmatrix} k_{11} & k_{12} & \cdots & k_{1n} \\ k_{21} & k_{22} & \cdots & k_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ k_{m1} & k_{m2} & \cdots & k_{mn} \end{pmatrix}$$

对C=AB的重认识

- 由此可知，若C=AB，则矩阵C的列向量能由A的列向量线性表示，B即为这一表示的系数矩阵。
- 向量组B: $b_1,b_2,\dots,b_n$能由向量组A: $a_1,a_2,\dots,a_m$线性表示的充要条件是矩阵$A=(a_1,a_2,\dots,a_m)$的秩等于矩阵$(A,B)=(a_1,a_2,\dots,a_m,b_1,b_2,\dots,b_n)$的秩，即：$R(A)=R(A,B)$

正交阵

- 若n阶矩阵A满足$A^TA=I$，成A为正交矩阵，简称正交阵。
  - A是正交阵的充要条件：A的列(行)向量都是单位向量，且两两正交。
- A是正交阵，x为向量，则$A\cdot x$称作正交变换。
  - 正交变换不改变向量长度

思考

- 若A、B都是n阶正交阵，那么，A×B是正交阵吗？
- 正交阵和对称阵，能够通过何种操作获得一定意义下的联系？

特征值和特征向量

- A是n阶矩阵，若数λ和n维非0列向量x满足Ax=λx，那么，数λ称为A的特征值，x称为A的对应于特征值λ的特征向量。
  - 根据定义，立刻得到$(A-\lambda I)x=0$，令关于λ的多项式$|A-\lambda I|$为0，方程$|A-\lambda I|=0$的根为A的特征值；将根λ_0带入方程组$(A-\lambda I)x=0$，求得到的非零解，即λ_0对应的特征向量。

特征值的性质

- 设n阶矩阵$A=(a_{ij})$的特征值为$\lambda_1, \lambda_2, \ldots \lambda_n$,则
- $\lambda_1+\lambda_2+\ldots+\lambda_n=a11+a22+\ldots+ann$
- $\lambda_1\lambda_2\ldots\lambda_n=|A|$
  - 矩阵A主行列式的元素和，称作矩阵A的迹。

思考

- 已知λ是方阵A的特征值，
- 则
  - λ^2是A^2的特征值
  - A可逆时，λ^{-1}是A^{-1}的特征值。

不同特征值对应的特征向量

- 设$\lambda_1,\lambda_2,\ldots,\lambda_m$是方阵A的m个特征值，$p_1,p_2,\ldots,p_m$是依次与之对应的特征向量，若$\lambda_1,\lambda_2,\ldots,\lambda_m$各不相等，则$p_1,p_2,\ldots,p_m$线性无关。

- 总结
  - 不同特征值对应的特征向量，线性无关。
  - 若方阵A是对称阵呢？结论是否会加强？
    - 协方差矩阵、二次型矩阵、无向图的邻接矩阵等都是对称阵
    - 在谱聚类中将会有所涉及

引理

- 实对称阵的特征值是实数
  - 设复数λ为对称阵A的特征值，复向量x为对应的特征向量，即Ax=λx(x≠0)
  - 用 $\overline{\lambda}$ 表示λ的共轭复数，$\overline{x}$ 表示x的共轭复向量，而A是实矩阵，有 $\overline{A}=A$
  - 下面给出证明过程。

证明

- 首先 $A\overline{x} = \overline{A}\overline{x} = \overline{Ax} = \overline{\lambda x} = \overline{\lambda}\overline{x}$
- 因为 $\overline{x}^T(Ax) = \overline{x}^T(Ax) = \overline{x}^T\lambda x = \lambda\overline{x}^T x$

$$\overline{x}^T(Ax) = \left(\overline{x}^T A^T\right)x = \left(A\overline{x}\right)^T x = \left(\overline{\lambda}\overline{x}\right)^T x = \overline{\lambda}\overline{x}^T x$$

- 从而

$$\lambda\overline{x}^T x = \overline{\lambda}\overline{x}^T x \Rightarrow \left(\lambda - \overline{\lambda}\right)\overline{x}^T x = 0$$

- 而

$$\overline{x}^T x = \sum_{i=1}^{n} \overline{x_i} x_i = \sum_{i=1}^{n} \left|x_i\right|^2 \neq 0$$

- 所以

$$\lambda - \overline{\lambda} = 0 \Rightarrow \lambda = \overline{\lambda}$$

实对称阵不同特征值的特征向量正交

- 令实对称矩阵为A，它的两个不同的特征值$\lambda_1 \lambda_2$对应的特征向量分别是$\mu_1 \mu_2$；其中，$\lambda_1 \lambda_2$ $\mu_1 \mu_2$都是实数或是实向量。
- 则有：$A\mu_1=\lambda_1\mu_1$，$A\mu_2=\lambda_2\mu_2$
- $(A\mu_1)^T=(\lambda_1\mu_1)^T$，从而：$\mu_1^T A=\lambda_1\mu_1^T$
- 所以：$\mu_1^T A\mu_2=\lambda_1\mu_1^T\mu_2$
- 同时，$\mu_1^T A\mu_2=\mu_1^T(A\mu_2)=\mu_1^T\lambda_2\mu_2=\lambda_2\mu_1^T\mu_2$
- 所以，$\lambda_1\mu_1^T\mu_2=\lambda_2\mu_1^T\mu_2$
- 故：$(\lambda_1-\lambda_2)\ \mu_1^T\mu_2=0$
- 而$\lambda_1\neq\lambda_2$，所以$\mu_1^T\mu_2=0$，即：μ_1，μ_2正交。

利用上述结论很快得到

- 将实数λ带入方程组(A- λ I)x=0，该方程组为实系数方程组，因此，实对称阵的特征向量可以取实向量。

最终结论

- 设A为n阶对称阵，则必有正交阵P，使得

$$P^{-1}AP = P^{T}AP = \Lambda$$

  - Λ是以A的n个特征值为对角元的对角阵。
  - 该变换称为“合同变换”，A和Λ互为合同矩阵。

二次型

- 含有n个变量的二次齐次函数，称为二次型；
- 一个二次型对应一个对称阵；
- 而对称阵可以由正交阵对角化，
- 从而二次型可以化成只有n个变量平方项的标准型，而这个正交阵，对应着坐标系的旋转变化。

正定阵

- 对于n阶方阵A，若任意n阶向量x，都有$x^TAx>0$，则称A是正定阵。
  - 若条件变成$x^TAx \geqslant 0$，则A称作半正定阵
  - 类似还有负定阵，半负定阵。

正定阵的判定

- 对称阵A为正定阵；
- A的特征值都为正；
- A的顺序主子式大于0；
- 以上三个命题等价。

$$
\begin{pmatrix} a_{11} \end{pmatrix} \quad \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \quad \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \quad \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}
$$

练习题

- 给定凸锥的定义如下：

$$C\text{为凸锥} \Leftrightarrow \forall x_1, x_2 \in C, \theta_1, \theta_1 \geq 0, \text{有}\ \theta_1 x_1 + \theta_2 x_2 \in C$$

- 试证明：n阶半正定方阵的集合为凸锥。
  - 考察半正定阵的定义

利用定义证明

- 若A、B为n阶半正定阵，则

$$\forall \vec{z},\ \vec{z}^T A\vec{z} \ge 0,\ \vec{z}^T B\vec{z} \ge 0$$

- 从而，$\forall \theta_1, \theta_2 \ge 0,$

$$\vec{z}^T \cdot (\theta_1 A + \theta_2 B) \cdot \vec{z} = \vec{z}^T \cdot \theta_1 A \cdot \vec{z} + \vec{z}^T \cdot \theta_2 B \cdot \vec{z}$$

$$= \theta_1 \vec{z}^T A\vec{z} + \theta_2 \vec{z}^T B\vec{z} \ge 0$$

- 即：$\forall \theta_1, \theta_2 \ge 0, \theta_1 A + \theta_2 B$ 为半正定阵。从而，n阶半正定阵的集合为凸锥。

向量的导数

- A为m×n的矩阵，x为n×1的列向量，则Ax为m×1的列向量，记$\vec{y} = A \cdot \vec{x}$
- 思考：$\dfrac{\partial \vec{y}}{\partial \vec{x}} = ?$

推导

- 令

$$A=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \quad \vec{x}=\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad A\cdot\vec{x}=\begin{pmatrix} a_{11}x_1+a_{12}x_2+\cdots+a_{1n}x_n \\ a_{21}x_1+a_{22}x_2+\cdots+a_{2n}x_n \\ \vdots \\ a_{m1}x_1+a_{m2}x_2+\cdots+a_{mn}x_n \end{pmatrix}$$

- 从而，

$$\frac{\partial\vec{y}}{\partial\vec{x}}=\frac{\partial A\vec{x}}{\partial\vec{x}}=\begin{bmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1m} & a_{2m} & \cdots & a_{nm} \end{bmatrix}=A^T$$

结论与直接推广

- 向量偏导公式：
$$\frac{\partial A\vec{x}}{\partial \vec{x}} = A^T$$

$$\frac{\partial A^T\vec{x}}{\partial \vec{x}} = A \qquad \frac{\partial A\vec{x}}{\partial \vec{x}^T} = A$$

- 在线性回归中将直接使用该公式。

思考题

- 给定两个随机变量X和Y，如何度量这两个随机变量的“距离”？
- 如何证明大数定理？
- 仿照指数分布的概率密度函数$f(x)=\lambda e^{-\lambda x}$，猜测相对应的幂分布的概率密度函数，查阅关于幂律分布的相关文献。

$$f(x)=ax^{-r}，a,r\text{为正常数}$$

延伸算法题

- 根据定义来计算 $C=A\times B$，需要m*n*s次乘法。
  - 即：若A、B都是n阶方阵，C的计算时间复杂度为$O(n^3)$
  - 问：可否设计更快的算法？
- 三个矩阵A、B、C的阶分别是$a_0\times a_1$, $a_1\times a_2$, $a_2\times a_3$，从而$(A\times B)\times C$和$A\times(B\times C)$的乘法次数是$a_0a_1a_2+a_0a_2a_3$、$a_1a_2a_3+a_0a_1a_3$，二者一般情况是不相等的。
  - 问：给定n个矩阵的连乘积：$A_1\times A_2\times A_3\ldots\ldots\times A_n$，如何添加括号来改变计算次序，使得乘法的计算量最小？

参考文献

- 王松桂，程维虎，高旅端编，概率论与数理统计，科学出版社，2000

我们在这里

- 7 七月算法 http://www.julyedu.com/
  - 视频/课程/社区
- 七月题库APP：Android/iOS
  - http://www.julyapp.com/
- 微博
  - @研究者July
  - @七月题库
  - @邹博_机器学习
- 微信公众号
  - julyedu

感谢大家！

恳请大家批评指正！