矩阵、向量微分计算

2015年11月06 9:25

定义1. 梯度（Gradient） [标量对列向量微分]
设$f(x)$是一个变量为x的标量函数，其中$x=(x_1 \ldots x_N)^T$。那么定义$f(x)$对x的梯度为$\frac{df(x)}{dx}$：

$$\frac{\partial f(x)}{\partial x} = \begin{pmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \vdots \\ \frac{\partial f(x)}{\partial x_N} \end{pmatrix}. \tag{D.1}$$

梯度的转置是一个行向量：

$$\left(\frac{\partial f(x)}{\partial x}\right)^T = \frac{\partial f(x)}{\partial x^T} = \begin{pmatrix} \frac{\partial f(x)}{\partial x_1} & \frac{\partial f(x)}{\partial x_2} & \cdots & \frac{\partial f(x)}{\partial x_N} \end{pmatrix}. \tag{D.2}$$

定义2. 海塞矩阵（Hessian matrix）【海塞矩阵是二阶梯度】
设$f(x)$是一个二阶可微分的标量函数，其中$x=(x_1 \ldots x_N)^T$。那么定义$f(x)$对x的海塞矩阵为$\frac{d^2f(x)}{dxdx^T}$：

$$\frac{\partial^2 f(x)}{\partial x \partial x^T} = \begin{pmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_N} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & & \vdots \\ \vdots & & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_N \partial x_1} & \cdots & \cdots & \frac{\partial^2 f(x)}{\partial x_N^2} \end{pmatrix}. \tag{D.3}$$

海塞矩阵是对称阵。

定义3. 雅可比矩阵（Jacobian matrix）【雅可比矩阵本质上是一阶梯度，向量对向量微分】
设$f(x)$是一个K X 1的列向量函数

$$f(x) = \begin{pmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_K(x) \end{pmatrix} \tag{D.4}$$

其中$x=(x_1 \dots x_L)^T$。那么定义$f(x)$对x的雅可比矩阵为$\frac{df(x)}{dx^T}$：

$$\frac{\partial f(x)}{\partial x^T}=\begin{pmatrix} \frac{\partial f_1(x)}{\partial x_1} & \frac{\partial f_1(x)}{\partial x_2} & \cdots & \frac{\partial f_1(x)}{\partial x_L} \\ \frac{\partial f_2(x)}{\partial x_1} & \frac{\partial f_2(x)}{\partial x_2} & \cdots & \frac{\partial f_2(x)}{\partial x_L} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_K(x)}{\partial x_1} & \frac{\partial f_K(x)}{\partial x_2} & \cdots & \frac{\partial f_K(x)}{\partial x_L} \end{pmatrix}. \qquad \text{(D.5)}$$

定义4. ***[矩阵对标量微分]***

$M \times N$的矩阵 A的元素是一个向量x的元素x_q的函数，定义$\frac{\partial A}{\partial x_q}$为：

$$\frac{\partial A}{\partial x_q}=\begin{pmatrix} \frac{\partial a_{11}}{\partial x_q} & \cdots & \frac{\partial a_{1N}}{\partial x_q} \\ \vdots & \ddots & \vdots \\ \frac{\partial a_{M1}}{\partial x_q} & \cdots & \frac{\partial a_{MN}}{\partial x_q} \end{pmatrix} \qquad \text{(D.7)}$$

矩阵的二阶微分：

$$\frac{\partial^2 A}{\partial x_p \partial x_q}=\begin{pmatrix} \frac{\partial^2 a_{11}}{\partial x_p \partial x_q} & \cdots & \frac{\partial^2 a_{1N}}{\partial x_p \partial x_q} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 a_{M1}}{\partial x_p \partial x_q} & \cdots & \frac{\partial^2 a_{MN}}{\partial x_p \partial x_q} \end{pmatrix}. \qquad \text{(D.8)}$$

定理1. 矩阵的乘积微分

A是$K \times M$的矩阵，B是 $K \times L$的矩阵，$C=AB$。设A和B的元素是向量x的一个元素x_q的函数，那么：

$$\frac{\partial C}{\partial x_p}=\frac{\partial A}{\partial x_p}B+A\frac{\partial B}{\partial x_p}. \qquad \text{(D.9)}$$

证明过程如下：

$$c_{k\ell} = \sum_{m=1}^{M} a_{km} b_{m\ell}. \qquad \text{(D.10)}$$

$$\frac{\partial c_{k\ell}}{\partial x_p} = \sum_{m=1}^{M} \frac{\partial a_{km}}{\partial x_p} b_{m\ell} + a_{km} \frac{\partial b_{m\ell}}{\partial x_p} \qquad \text{(D.11)}$$

$$\frac{\partial C}{\partial x_p} = \frac{\partial A}{\partial x_p} B + A \frac{\partial B}{\partial x_p}\,. \qquad \text{(D.12)}$$

定理2.

A是$M \times N$的非奇异矩阵，设A的元素是向量x的一个元素x_q的函数，那么A^{-1}对x的一阶微分和二阶微分分别为：

$$\frac{\partial A^{-1}}{\partial x_q} = -A^{-1} \frac{\partial A}{\partial x_q} A^{-1} \qquad \text{(D.13)}$$

$$\frac{\partial^2 A^{-1}}{\partial x_p \partial x_q} = A^{-1} \left(\frac{\partial A}{\partial x_p} A^{-1} \frac{\partial A}{\partial x_q} - \frac{\partial^2 A}{\partial x_p \partial x_q} + \frac{\partial A}{\partial x_q} A^{-1} \frac{\partial A}{\partial x_p} \right) A^{-1}. \qquad \text{(D.14)}$$

证明过程如下：

$$A \frac{\partial A^{-1}}{\partial x_q} + \frac{\partial A}{\partial x_q} A^{-1} = \frac{\partial I}{\partial x_q} = O\,, \qquad \text{(D.15)}$$

$$\frac{\partial A^{-1}}{\partial x_q} = -A^{-1} \frac{\partial A}{\partial x_q} A^{-1}. \qquad \text{(D.16)}$$

$$\frac{\partial^2 A^{-1}}{\partial x_p \partial x_q} = -\frac{\partial}{\partial x_p} \left(A^{-1} \frac{\partial A}{\partial x_q} A^{-1} \right). \qquad \text{(D.17)}$$

$$\frac{\partial^2 A^{-1}}{\partial x_p \partial x_q} = -\frac{\partial A^{-1}}{\partial x_p} \frac{\partial A}{\partial x_q} A^{-1} - A^{-1} \frac{\partial^2 A}{\partial x_p \partial x_q} A^{-1} - A^{-1} \frac{\partial A}{\partial x_q} \frac{\partial A^{-1}}{\partial x_p}. \qquad \text{(D.18)}$$

$$\frac{\partial^2 A^{-1}}{\partial x_p \partial x_q} = A^{-1} \left(\frac{\partial A}{\partial x_p} A^{-1} \frac{\partial A}{\partial x_q} - \frac{\partial^2 A}{\partial x_p \partial x_q} + \frac{\partial A}{\partial x_q} A^{-1} \frac{\partial A}{\partial x_p} \right) A^{-1}. \qquad \text{(D.19)}$$

二次函数$f(x) = x^T V x$，其中$x = (x_1 \ldots x_k)^T$，$k \times k$矩阵A。则$f(x)$对$k \times 1$的列向量x的微分为: $\frac{d(x^T V x)}{dx} = (V + V^T)x$

以k=3的情况举例说明：

$$
\begin{aligned}
\beta' V \beta &= \begin{pmatrix} \beta_1 & \beta_2 & \beta_3 \end{pmatrix} \begin{pmatrix} V_{11} & V_{12} & V_{13} \\ V_{21} & V_{22} & V_{23} \\ V_{31} & V_{32} & V_{33} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} \\
&= V_{11}\beta_1^2 + V_{22}\beta_2^2 + V_{33}\beta_3^2 + (V_{12}+V_{21})\beta_1\beta_2 + (V_{13}+V_{31})\beta_1\beta_3 + (V_{23}+V_{32})\beta_2\beta_3.
\end{aligned}
$$

Taking the derivative with respect to β, we get

$$
\begin{aligned}
\frac{\partial(\beta' V \beta)}{\partial \beta} &= \begin{pmatrix} \frac{\partial(\beta' V \beta)}{\partial \beta_1} \\ \frac{\partial(\beta' V \beta)}{\partial \beta_2} \\ \frac{\partial(\beta' V \beta)}{\partial \beta_3} \end{pmatrix} \\
&= \begin{pmatrix} 2V_{11}\beta_1 + (V_{12}+V_{21})\beta_2 + (V_{13}+V_{31})\beta_3 \\ 2V_{22}\beta_2 + (V_{12}+V_{21})\beta_1 + (V_{23}+V_{32})\beta_3 \\ 2V_{33}\beta_3 + (V_{13}+V_{31})\beta_1 + (V_{23}+V_{32})\beta_2 \end{pmatrix} \\
&= \begin{pmatrix} 2V_{11} & V_{12}+V_{21} & V_{13}+V_{31} \\ V_{12}+V_{21} & 2V_{22} & V_{23}+V_{32} \\ V_{13}+V_{31} & V_{23}+V_{32} & 2V_{33} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} \\
&= \left(\begin{pmatrix} V_{11} & V_{12} & V_{13} \\ V_{21} & V_{22} & V_{23} \\ V_{31} & V_{32} & V_{33} \end{pmatrix} + \begin{pmatrix} V_{11} & V_{21} & V_{31} \\ V_{12} & V_{22} & V_{32} \\ V_{13} & V_{23} & V_{33} \end{pmatrix} \right) \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} \\
&= (V + V')\beta.
\end{aligned}
$$

矩阵微分的应用

线性回归模型

$$Y = X\beta + \varepsilon$$

Y是T×1向量，X是T×k的矩阵，β是待估计的k×1的参数向量

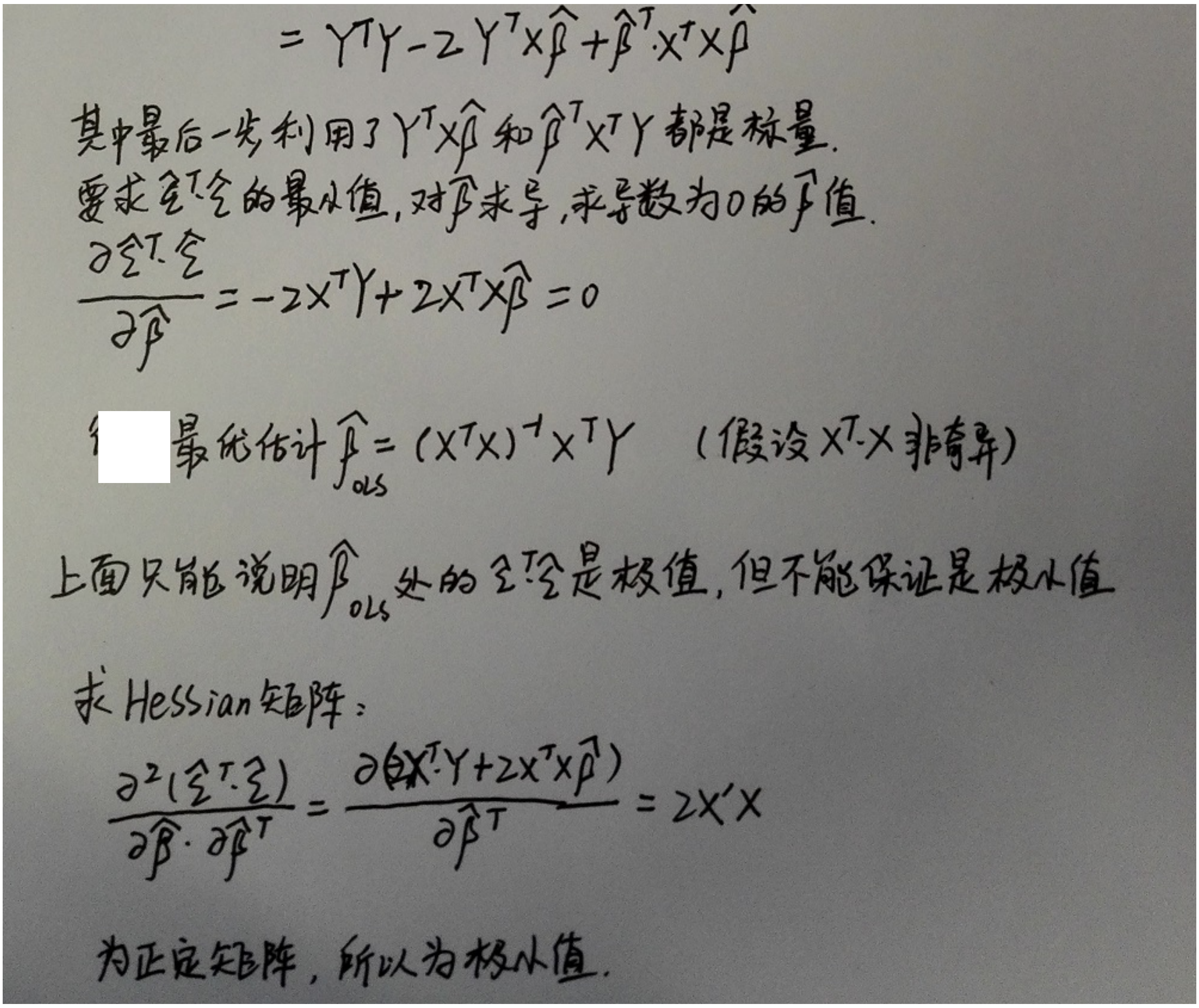
ε是T×1的误差项，(T是总的观测样本数量)

最优化误差的平方和：$\hat{\beta}_{OLS} = \arg\min_{\hat{\beta}} \sum_{t=1}^{T} \hat{\varepsilon}_t^2 = \arg\min_{\hat{\beta}} \hat{\varepsilon}^T \hat{\varepsilon}$

其中 $\hat{\varepsilon}^T \cdot \hat{\varepsilon} = (Y - X\hat{\beta})^T \cdot (Y - X\hat{\beta})$

$$= (Y^T - \hat{\beta}^T X^T) \cdot (Y - X\hat{\beta})$$

$$= Y^T Y - Y^T X\hat{\beta} - \hat{\beta}^T X^T Y + \hat{\beta}^T X^T X\hat{\beta}$$



$= Y^TY - 2Y^TX\hat{\beta} + \hat{\beta}^T X^T X\hat{\beta}$

其中最后一步利用了$Y^TX\hat{\beta}$和$\hat{\beta}^TX^TY$都是标量.

要求$\hat{\varepsilon}^T\hat{\varepsilon}$的最小值，对$\hat{\beta}$求导，求导数为0的$\hat{\beta}$值.

$$\frac{\partial \hat{\varepsilon}^T\hat{\varepsilon}}{\partial\hat{\beta}} = -2X^TY+2X^TX\hat{\beta}=0$$

最优估计 $\hat{\beta}_{OLS} = (X^TX)^{-1}X^TY$ （假设X^TX非奇异）

上面只能说明$\hat{\beta}_{OLS}$处的$\hat{\varepsilon}^T\hat{\varepsilon}$是极值，但不能保证是极小值

求Hessian矩阵：

$$\frac{\partial^2(\hat{\varepsilon}^T\hat{\varepsilon})}{\partial\hat{\beta}\cdot\partial\hat{\beta}^T} = \frac{\partial(-2X^TY+2X^TX\hat{\beta})}{\partial\hat{\beta}^T} = 2X'X$$

为正定矩阵，所以为极小值.

矩阵迹的微分（Derivative of Traces）

在机器学习中，有时候需要对一个矩阵的F模进行微分，而矩阵的F是可以转换为矩阵的迹，矩阵的迹的微分的计算可以帮助我们计算矩阵的F模的微分回归模型中，输出不是0和1，而是一个向量，这时整个输出矩阵就不是向量而是矩阵的。这会在最后的例子中具体说明。

矩阵的F模和迹的关系：

$$\|A\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2} = \sqrt{\operatorname{trace}(A^*A)} = \sqrt{\sum_{i=1}^{\min\{m,n\}}\sigma_i^2}$$

其中A^*是A的共轭转置。

矩阵的迹的性质：

$$\operatorname{tr}(ABCD) = \operatorname{tr}(BCDA) = \operatorname{tr}(CDAB) = \operatorname{tr}(DABC)$$

上面的两点会在后面的例子中用到。

Matrix Cookbook中给出了矩阵迹的微分的一般表达式：
$\frac{\partial}{\partial x}tr(F(x)) = f(x)^T$ 其中，$f()$是$F()$的微分。

一阶：

$$
\begin{aligned}
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}) &= \mathbf{I} \\
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{XA}) &= \mathbf{A}^T \\
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{AXB}) &= \mathbf{A}^T\mathbf{B}^T \\
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{AX}^T\mathbf{B}) &= \mathbf{BA} \\
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}^T\mathbf{A}) &= \mathbf{A} \\
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{AX}^T) &= \mathbf{A} \\
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\qquad) &= \mathrm{Tr}(\mathbf{A})\mathbf{I}
\end{aligned}
$$

二阶：

$$
\begin{aligned}
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}^2) &= 2\mathbf{X}^T && (106)\\
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}^2\mathbf{B}) &= (\mathbf{XB}+\mathbf{BX})^T && (107)\\
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}^T\mathbf{BX}) &= \mathbf{BX}+\mathbf{B}^T\mathbf{X} && (108)\\
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{BXX}^T) &= \mathbf{BX}+\mathbf{B}^T\mathbf{X} && (109)\\
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{XX}^T\mathbf{B}) &= \mathbf{BX}+\mathbf{B}^T\mathbf{X} && (110)\\
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{XBX}^T) &= \mathbf{XB}^T+\mathbf{XB} && (111)\\
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{BX}^T\mathbf{X}) &= \mathbf{XB}^T+\mathbf{XB} && (112)\\
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}^T\mathbf{XB}) &= \mathbf{XB}^T+\mathbf{XB} && (113)\\
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{AXBX}) &= \mathbf{A}^T\mathbf{X}^T\mathbf{B}^T+\mathbf{B}^T\mathbf{X}^T\mathbf{A}^T && (114)\\
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}^T\mathbf{X}) &= \frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{XX}^T) = 2\mathbf{X} && (115)\\
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{B}^T\mathbf{X}^T\mathbf{CXB}) &= \mathbf{C}^T\mathbf{XBB}^T+\mathbf{CXBB}^T && (116)\\
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}\left[\mathbf{X}^T\mathbf{BXC}\right] &= \mathbf{BXC}+\mathbf{B}^T\mathbf{XC}^T && (117)\\
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{AXBX}^T\mathbf{C}) &= \mathbf{A}^T\mathbf{C}^T\mathbf{XB}^T+\mathbf{CAXB} && (118)\\
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}\left[(\mathbf{AXB}+\mathbf{C})(\mathbf{AXB}+\mathbf{C})^T\right] &= 2\mathbf{A}^T(\mathbf{AXB}+\mathbf{C})\mathbf{B}^T && (119)\\
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}\otimes\mathbf{X}) &= \frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X})\mathrm{Tr}(\mathbf{X}) = 2\mathrm{Tr}(\mathbf{X})\mathbf{I} && (120)
\end{aligned}
$$

高阶：

$$
\begin{aligned}
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}^k) &= k(\mathbf{X}^{k-1})^T && (121)\\
\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{AX}^k) &= \sum_{r=0}^{k-1}(\mathbf{X}^r\mathbf{AX}^{k-r-1})^T && (122)\\
\tfrac{\partial}{\partial \mathbf{X}}\mathrm{Tr}\left[\mathbf{B}^T\mathbf{X}^T\mathbf{CXX}^T\mathbf{CXB}\right] &= \mathbf{CXX}^T\mathbf{CXBB}^T \\
&\quad +\mathbf{C}^T\mathbf{XBB}^T\mathbf{X}^T\mathbf{C}^T\mathbf{X} \\
&\quad +\mathbf{CXBB}^T\mathbf{X}^T\mathbf{CX} \\
&\quad +\mathbf{C}^T\mathbf{XX}^T\mathbf{C}^T\mathbf{XBB}^T && (123)
\end{aligned}
$$

看了给的这些例子后，感觉有些情况下计算$F()$的微分$f()$还是有点困难，不明白到底计算的规则是怎么样的。比如高阶中的前
$\frac{\partial}{\partial X}tr(X^k)=k(X^{k-1})^T$和$\frac{\partial}{\partial X}tr(AX^k)=\sum_{r=0}^{k-1}(X^rAX^{k-r-1})^T$，第二个例子仅因为多了一个A结果却大相径庭，但是这里计算$F()$的微分$f()$
的？我感觉不是很明白，资料上也没看具体说明。

我自己就琢磨出来了一套计算规律，感觉挺好用的。
具体说来，首先就是对$tr(T_1XT_2X^TT_3)$中所有的X分别计算，对于tr中出现的X，结果就是X**前面**部分整体的转置乘以X**后面**部分整体的转置，对
X^T，结果就是X^T前面部分整体的转置乘以X后面部分整体的转置，然后再整体来一个转置。或者进一步简化，结果就是X^T**后面**的整体部分乘以
分。

形式化的式子：
$\frac{\partial}{\partial X}tr(T_1XT_2X^TT_3)=T_1^T(T_2X^TT_3)^T+\{(T_1XT_2)^TT_3^T\}^T$
或者进一步化为：
$\frac{\partial}{\partial X}tr(T_1XT_2X^TT_3)=T_1^T(T_2X^TT_3)^T+\{T_3(T_1XT_2)\}$
上面式子中T_1、T_2、T_3为任意长的矩阵连乘表达式，可以包含X。

对上面列出 行计算，发现都符合。举一个例子。
$\frac{\partial}{\partial X}tr(AX^k)=\sum_{r=0}^{k-1}(X^rAX^{k-r-1})^T$
使用上面的计算规则计算，共有k个X，所以结果有k项，对每一个X依次计算：
$\frac{\partial}{\partial X}tr(AX^k)=A^T(X^{k-1})^T+(AX)^T(X^{k-2})^T+(AX^2)^T(X^{k-3})^T+...+(AX^{k-2})^T(X)^T$
$+(AX^{k-1})^T$
相信聪明的读者已经看出规律了，写成求和的形式就是：
$\frac{\partial}{\partial X}tr(AX^k)=\sum_{r=0}^{k-1}(X^rAX^{k-r-1})^T$
和上面给出的式子一模一样的！

举个例子，$Y=XW+E$，这里Y、X、W、E都是矩阵，这可以看作是机器学习中的回归的问题，Y是输出矩阵，X是特征向量，W是待学习的
是误差，$E=Y-XW$，是真实值和预测值的差，机器学习中希望学习到一个参数矩阵W使得误差最小，即E的F模最小（这里E是矩阵，如果是向
取向量的模就可以）
$||E||_F=tr(E^TE)=tr[(Y-XW)^T(Y-XW)]=tr(Y^TY-Y^TXW-W^TX^TY+W^TX^TXW)$

$\frac{\partial}{\partial W}||E||_F=\frac{\partial}{\partial W}tr(E^TE)=\frac{\partial}{\partial W}tr(Y^TY-Y^TXW-W^TX^TY+W^TX^TXW)$

$\frac{\partial}{\partial W}||E||_F=-X^TY-X^TY+X^TXW+X^TXW=-2X^TY+2X^TXW=0$
$X^TY=X^TXW$

最后的结果就是：
$W=(X^TX)^{-1}X^TY$

参考：
APPENDIX D VECTOR AND MATRIX DIFFERENTIATION
INTRODUCTION TO VECTOR AND MATRIX DIFFERENTIATION
Matrix CookBook <这个比较全>

文章标签： 函数 matrix ▼查看关于本篇文章更多信息

想对作者说点什么？ 我来说一句

22世纪_冲刺 2018-01-07 22:28:45 #4楼
写博文不易，值得推荐。同时，也提出个人的不成熟建议，今天我翻书看了矩阵论微分学这一块，对于矩阵的微分，针对微分的分式而言：分母上可以是 标量，行向
阵；同理，分子上，也可以是标量，行向量，列向量和矩阵。

lshmusic 2017-06-26 22:40:41 #3楼
文章非常棒，感谢分享！！！

vacant2012 2017-06-08 15:31:53 #2楼

最后部分应该是对W求偏导才对吧

查看 4 条热评

矩阵求导计算法则 3.8万

转自：http://blog.sina.com.cn/s/blog_4a033b090100pwjq.html 求导公式(撇号为转置）： Y = A * X --> DY/DX = A' Y =...

矩阵向量求导(Matrix calculus) 2373

原文地址：https://en.wikipedia.org/wiki/Matrix_calculus#Other_matrix_derivatives翻译：part1:http://blog.csdn...

矩阵和向量的微分方法

矩阵和向量的微分求导方法,编程的数学工具,十分有用。

常用矩阵微 [illegible] - CSDN博客

矩阵微分(Ma[illegible]ential)也称矩阵求导(Matrix Derivative),在机器学习、图像处理... 雅可比矩阵也可以看做是向量对向量的求导而...

矩阵论：向量求导/微分和矩阵微分 1.1万

http://blog.csdn.net/pipisorry/article/details/68961388复杂的矩阵函数求导。著名的matrix cookbook为广大的研究者们提供了一本大字...

矩阵微分与向量函数Taylor展开 - CSDN博客

第一部分:矩阵微分 计算∂F∂X\frac{\partial F}{\partial X}时,根据F和X的类型有不同的微分公式。F和X可以分别是标量、向量和矩阵。 1....

对向量、方阵的求导 - CSDN博客

接下来针对机器学习公式推导过程中经常用到的矩阵求导,我们做一个详细介绍。矩阵求导(Matrix Derivative)也称作矩阵微分(... 相关热...

矩阵微分（matrix derivatives） 2892

关于矩阵求导，得到的导数则是矩阵形式；关于矢量求导，得到的导数则是矢量形式；关于标量求导，得到的仍是标量形式。 共存在 6...

矩阵论笔记（七）——矩阵的微分和积分 1086

对矩阵求微分和积分，就是对其每个元素求微分和积分。定义 导数：矩阵 A(t)=(aij(t))m×nA(t) = (a_{ij}(t))_{m\times n} 的每个元素可微...

开始学习Matlab_计算功能_微分diff/求导或向量矩阵比较 - CSDN博客

1、微分(注:diff()函数也可用于向量、矩阵之间列的比较)diff(f) 此函数是对函数f中的符号变量x(默认是x),或者定义的符号变量中没有x...

[免费]数学向量实现

矩阵和向量的微分方法 立即下载 上传者: xub09 时间: 2013-10-30 综合评分: 5 积分/C币:10 as3.0鸟飞动画 立即下载 上传者: 澄清寰宇 ...

Matrix calculus(矩阵微积分)（前四节） 1007

原文地址：https://en.wikipedia.org/wiki/Matrix_calculus 注：不要把它和几何运算或者是向量运算混淆前言：在数学中，矩阵微积分是...

向量的求导 - CSDN博客

向量或者矩阵对元素的求导很简单元素对想两矩阵求导依然简单行对列与列对行 注意分母的对应行对行 列对列矩阵对行向量以及列向...

矩阵向量求导(Matrix calculus) - CSDN博客

这大大简化例如找到多元函数的最大值或最小值,以及求解微分方程组的操作。 ...在这里我们把矩阵作为最一般的情况,把向量和标量分...

矩阵微分法 下载 2018年08

所有用矩阵论的学生,都可以看看,里面的推导过程很详细,希望对你有所帮助!... 2012-03-10 上传大小:391KB 矩阵微分...

常用矩阵微分公式.pdf 下载 2018年08

常用矩阵微分公式

矩阵、向量求导法则 - CSDN博客

关于矩阵、向量求导法则,详细公式: http://www.cnblogs.com/huashiyiqike/p/3568922.html

对向量求导 - CSDN博客

上一篇MatConvNet 使用VGG网络模型对图像做分类处理 下一篇MATLAB 视频分帧 向量...矩阵论:向量求导/微分和矩阵微分 pipisorry 0…

常用矩阵微分公式

7674

转自http://www.cnblogs.com/xuxm2007/p/3332035.html 矩阵微分 http://www.iwenchao.com/mathematics/matrix-d...

机器学习中的线性代数

7887

第二章 机器学习中的线性代数知识线性代数作为数学中的一个重要的分支，广发应用在科学与工程中。掌握好线性代数对于理解和从…

相关热词 in矩阵 矩阵的λμ 矩阵∑ 矩阵

一、不懂这些线性代数知识 别说你是搞机器学习的

3903

数学是计算机技术的基础，线性代数是机器学习和深度学习的基础，了解数据知识最好的方法我觉得是理解概念，数学不只是上学时用…

深度学习/机器学习入门基础数学知识整理（一）：线性代数基础，矩阵，范数等

4282

前面大概有2年时间，利用业余时间断断续续写了一个机器学习方法系列，和深度学习方法系列，还有一个三十分钟理解系列（一些趣…

开始学习Matlab_计算功能_微分diff/求导或向量矩阵比较

230

1、微分（注：diff()函数也可用于向量、矩阵之间列的比较）diff(f)　 此函数是对函数f中的符号变量x(默认是x)，或者定义的符号变量…

pytorch自动微分的几个例子

8636

昨天微信'机器之心'发布了开源软件pytorch出场的重磅消息。看内容感觉动态计算图的思路比较新颖。在初步体验其自动微分操作之后…

PDF

矩阵 向量求导法则

2017年04月05日 85KB 下载

标量对矩阵求导

2666

矩阵（微分）求导的相关公式

机器学习中的矩阵求导总结

894

下图为常见的矩阵求导公式及其推导。

矩阵微分

628

矩阵微分(Matrix Differential) 矩阵微分(Matrix Differential)也称矩阵求导(Matrix Derivative)，在机器学习、图像处理、最...

Matrix calculus 矩阵微分

1153

转载地址：https://en.wikipedia.org/wiki/Matrix_calculus 前提： 若 x为向量，则默认 x为列向量， xT 为行向量，后面提到的两种布局都…

矩阵微分与向量函数Taylor展开

1051

第一部分：矩阵微分 计算∂F∂X\frac{\partial F}{\partial X}时，根据F和X的类型有不同的微分公式。F和X可以分别是标量、向量和矩阵…

常用的微分运算法则

8732

机器学习涉及到较多的数学知识，在工程应用领域，这些数学知识不是必要的，其实很多算法都是数值运算专家写好了的。然而知其然…

数学-矩阵计算（4）两种布局 4502
本博文来自维基上的矩阵计算：https://en.wikipedia.org/wiki/Matrix_calculus#Denominator-layout_notation 之前会发现在有的求导上最...

机器学习基础（六十一）—— 范数及范数的微分 2435
ℓ1\ell_1 范数的微分λ‖s‖1 \lambda \|s\|_1 ℓ1\ell_1 范数在 0 点不可微会影响梯度方法的应用。解决方案：（1）非梯度方法（2）"平...

深度学习/机器学习入门基础数学知识整理（二）：梯度与导数，矩阵求导，泰勒展开等 2672
导数与梯度 导数：一个一元函数函数在某一点的导数描述了这个函数在这一点附近的变化率。 f′(a)=limh→0f(a+h)−f(a)h f'(a) = \lim_{h ...

线性代数 阵的微分 2007
向量与矩阵微 简介对于可导实函数在某点x处的导数，有 f′(x)=limh→0f(x+h)−f(x)hf'(x)=\lim_{h\to 0}\frac{f(x+h)-f(x)}{h} 从...

PPT **矩阵微分与向量函数的Taylor展开.ppt**
2016年12月31日 136KB 下载

数学-矩阵计算（2）矩阵函数微积分前奏 3980
来自：http://www4.ncsu.edu/~pfackler/ 下面的《Notes on Matrix Calculus》，这是Paul I. Fackler 在2005年9月27日写的矩阵微积...

关于被忽略的转置矩阵的公式 3157
关于转置的公式常用的有： (AB)T=BTAT,(AT)T=A,(kA)T=kAT (AB)^T = B^TA^T, \\ (A^T)^T = A, \\ (kA)^T = kA^T 有一个非常不同...

矩阵求导简要笔记 1660
本文介绍了矩阵求导和矩阵乘法。

矩阵求导、几种重要的矩阵及常用的矩阵求导公式 2884
一、矩阵求导 一般来讲，我们约定x=(x1,x2,...xN)Tx=(x1,x2,...xN)Tx=(x_1,x_2,...x_N)^T，这是分母布局。常见的矩阵求导方式有...

通过一个例子快速上手矩阵求导 3.8万
前提及说明第一次遇见矩阵求导，大多数人都是一头雾水，而搜了维基百科看也还是云里雾里，一堆的名词和一堆的表格到底都是什么...

运用numpy进行数组、向量、矩阵运算 797
众所周知，python中3个用于数学计算的库分别是:numpy scipy pandas。之前已经介绍过了pandas的用法，这篇笔记主要介绍numpy...

PDF **矩阵和向量的微分方法**
2013年10月30日 115KB 下载

PDF **闲话矩阵求导.pdf**
2018年01月04日 1.68MB 下载

雅克比矩阵&行列式——单纯的矩阵和算子 1.3万
最近接触了一点雅克比的东西，以前学习雅克比矩阵和雅克比行列式是在高数上，就知道个二重积分的时候可以用一下，其他的真没遇...

史上最生成形象的理解矩阵的维度和乘法 565
线性代数的维度如何理解？其实这个很简单，以现在的找对象比喻，要看好几个方面，看脸，看钱，看房子，看车子，看家境，看潜力...

使用tr对矩阵进行求导 167
使用tr对矩阵进行求导：常用：

机器学习中常用的矩阵求导公式 3375
矩阵求导好像读书的时候都没学过，因为讲矩阵的课程上不讲求导，讲求导的课又不提矩阵。如果从事机器学习方面的工作，那就一定...

常用的向量矩阵求导公式 2.8万
总结下数理推导中常用的向量矩阵求导公式，方便以后查询。

矩阵函数X …导问题 4096
A, B, C 是不… 的矩阵，a,b 是不依赖于x 的向量。

矩阵求导公式【转】 8995
原文地址：矩阵求导公式【转】作者：三寅今天推导公式，发现居然有对矩阵的求导，狂汗--完全不会。不过还好网上有人总结了。吼...

主要介绍了有关矩阵、向量求导以及链式求导 209
矩阵、向量求导以及链式求导

Markdown中编写LaTeX数学公式 5969
编写LaTeX数学公式编写LaTeX数学公式 LaTeX基本语法 希腊字母 三角函数与逻辑数学字符 字体转换 花括号用法 多行数学式对齐 矩...

矩阵指数解线性微分方程 2313
矩阵指数 维基百科，自由的百科全书 矩阵指数是方块矩阵的一种矩阵函数，与指数函数类似。矩阵指数给出了矩阵李代数与对应的李...

python numpy模块玩转矩阵与科学计算 1.2万
学生时代玩矩阵最爽的工具自然是matlab了。而且matlab天生就是为科学计算，为矩阵而生。matlab的一切对象皆可看成矩阵，最简单...

求两个矩阵中向量的欧氏距离(python实现) 4870
假设有两个三维向量集，用矩阵表示: 要求A，B两个集合中的元素两两间欧氏距离。先求出ABT： 然后对A和BT分别求其中每个向量的...

向量和矩阵的微分 31
对元素求导 列向量对元素求导 y=[y1,y2,...,yd]T∈Rd,x∈R,∂y∂x=[∂y1x,∂y2x,…,∂ydx]Ty=[y1,y2,...,yd]T∈Rd,x∈R,∂y∂x=[∂y...

矩阵的求导运算 392
首先我们来看看最复杂的情况，也就是矩阵对矩阵的求导，假设我们有如下矩阵： 可以看作是因变量，而下面的矩阵可以看作是自变量...

多元微分学 201
实际应用中的函数普遍包含多个变量，当进入高维时，微积分的普遍法则本质上保持原样，虽然必须引入一些新的记号，但幸运的是并...

（Math）矩阵求导 1.4万
本地址：http://blog.csdn.net/mounty_fsc/article/details/51583809 前言 本文为维基百科上矩阵微积分部分的翻译内容。本文为原文...

最新文章

OpenCV Image Flip 图像

sigmoid_cross_entropy_w
ighted_cross_entropy_wit

Batch Normalization

improved partition in quick

ssh login server without p

个人分类

computer-vision

machine learning

c-c++

algorithm

code-tool

展开

归档

2018年7月

2018年1月

2017年12月

2017年9月

2017年5月

展开

热门文章

Matlab mex -setup 找不到
阅读量：35581

gcc 使用入门教程
阅读量：15052

Jsoncpp 使用方法大全
阅读量：14388

矩阵、向量微分计算
阅读量：10567

cs231n - assignment1 - s
阅读量：10338

最新评论