# 基本数学基础

陈飞宇

fchen@cqu.edu.cn

办公室：软件学院529室

# A. 矩阵论

# 迹函数

迹函数（Trace）：对于一个 n 阶方阵 A 来说，它的迹是主对角线上的元素之和。

$$\operatorname{tr}(A)=\sum_{i=1}^{n}A_{ii}$$

迹函数性质：

$$\operatorname{tr}(A)=\operatorname{tr}\left(A^T\right)$$

$$\operatorname{tr}(A+B)=\operatorname{tr}(A)+\operatorname{tr}(B)$$

$$\operatorname{tr}(AB)=\operatorname{tr}(BA)$$

$$\operatorname{tr}(ABC)=\operatorname{tr}(BCA)=\operatorname{tr}(CAB)$$

# F-范数

Frobenius 范数：对于矩阵 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$

$$\|\boldsymbol{A}\|_F = \left(\mathbf{tr}\left(\boldsymbol{A}^T\boldsymbol{A}\right)\right)^{1/2} = \left(\sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}^2\right)^{1/2}$$

或者

$$\|\boldsymbol{A}\|_F^2 = \mathbf{tr}\left(\boldsymbol{A}^T\boldsymbol{A}\right) = \sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}^2$$

因此 $\|\boldsymbol{A}\|_F = \|\mathrm{vec}(\boldsymbol{A})\|_2$

# 向量、矩阵对标量的求导

假设标量 $x \in \mathbb{R}$，向量 $\boldsymbol{a} \in \mathbb{R}^n$， 那么 $\frac{\partial \boldsymbol{a}}{\partial x} \in \mathbb{R}^n$

$$\left(\frac{\partial \boldsymbol{a}}{\partial x}\right)_i = \frac{\partial \boldsymbol{a}_i}{\partial x}$$

假设标量 $x \in \mathbb{R}$，矩阵 $\boldsymbol{A} \in \mathbb{R}^{m \times n}$， 那么 $\frac{\partial \boldsymbol{A}}{\partial x} \in \mathbb{R}^{m \times n}$

$$\left(\frac{\partial \boldsymbol{A}}{\partial x}\right)_{ij} = \frac{\partial \boldsymbol{A}_{ij}}{\partial x}$$

# 向量、标量函数对向量求导

假设向量 $\boldsymbol{x} \in \mathbb{R}^{\boldsymbol{n}}$，$\boldsymbol{y} \in \mathbb{R}^{\boldsymbol{m}}$， 那么 $\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}} \in \mathbb{R}^{\boldsymbol{n} \times \boldsymbol{m}}$ $(\mathbb{R}^{\boldsymbol{m} \times \boldsymbol{n}})$

$$\left(\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}}\right)_{\boldsymbol{ij}} = \frac{\partial \boldsymbol{y}_{\boldsymbol{j}}}{\partial \boldsymbol{x}_{\boldsymbol{i}}}$$

假设函数：$\boldsymbol{f}$ : $\mathbb{R}^{\boldsymbol{m} \times \boldsymbol{n}} \longrightarrow \mathbb{R}$ 是变量为矩阵的函数，那么

$$\left(\frac{\partial \boldsymbol{f}(\boldsymbol{A})}{\partial \boldsymbol{A}}\right)_{\boldsymbol{ij}} = \frac{\partial \boldsymbol{f}(\boldsymbol{A})}{\partial \boldsymbol{A}_{\boldsymbol{ij}}}$$

# 求导法则

乘法法则：

$$\frac{\partial AB}{\partial x} = \frac{\partial A}{\partial x}B + A\frac{\partial B}{\partial x}$$

链式法则：

假设函数 $f$ 是 $f$ 和 $f$ 的复合，即 $f(x) = g(h(x))$, 则有

$$\frac{\partial f(x)}{\partial x} = \frac{\partial g(h(x))}{\partial h(x)} \cdot \frac{\partial h(x)}{\partial x}$$

# 例子

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

$$\frac{\partial \mathrm{tr}(A)}{\partial A} = ? \qquad \frac{\partial \mathrm{tr}(AB)}{\partial A} = ?$$

$$\frac{\partial x^T A x}{\partial x} = ? \qquad \frac{\partial \|Ax - b\|_F^2}{\partial x} = ?$$

# 例子

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

$$\frac{\partial \operatorname{tr}(A)}{\partial A} = I \qquad \frac{\partial \operatorname{tr}(AB)}{\partial A} = B^T$$

$$\frac{\partial x^T A x}{\partial x} = (A + A^T)x \qquad \frac{\partial \|Ax - b\|_F^2}{\partial x} = 2A^T(Ax - b)$$

# 例子1

例子： $\frac{\partial Ax}{\partial x} = A^T$

$$\left(\frac{\partial y}{\partial x}\right)_{ij} = \frac{\partial y_j}{\partial x_i}$$

$$(Ax)_j = \sum_{k=1}^{n} A_{jk} x_k$$

$$\frac{\partial (Ax)_j}{\partial x_i} = \frac{\partial \sum_{k=1}^{n} A_{jk} x_k}{\partial x_i} = \frac{\partial A_{ji} x_i}{\partial x_i} = A_{ji}$$

$$\frac{\partial Ax}{\partial x} = A^T$$

# 例子4

例子： $\dfrac{\partial \mathrm{tr}(AB)}{\partial A} = B^T$

$$\left(\frac{\partial f(A)}{\partial A}\right)_{ij} = \frac{\partial f(A)}{\partial A_{ij}}$$

$$(AB)_{ij} = \sum_{k=1}^{p} A_{ik} B_{kj}$$

$$\mathrm{tr}(AB) = \sum_{i=1}^{n} (AB)_{ii} = \sum_{i=1}^{n} \sum_{k=1}^{p} A_{ik} B_{ki}$$

$$\frac{\partial \mathrm{tr}(AB)}{\partial A_{ij}} = \frac{\partial \sum_{i=1}^{n} \sum_{k=1}^{p} A_{ik} B_{ki}}{\partial A_{ij}} = \frac{\partial A_{ij} B_{ji}}{\partial A_{ij}} = B_{ji}$$

# 例子6

例子： $\frac{\partial\|Ax-b\|_F^2}{\partial x} = 2A^T(Ax-b)$

$$\frac{\partial f(x)}{\partial x} = \frac{\partial g(h(x))}{\partial h(x)} \cdot \frac{\partial h(x)}{\partial x}$$

$$\frac{\partial\|Ax-b\|_F^2}{\partial x} = \frac{\partial(Ax-b)}{\partial x} \cdot \frac{\partial\|Ax-b\|_F^2}{\partial(Ax-b)} = 2(Ax-b)A^T$$

# 多元线性回归

线性回归：寻求最优参数$\boldsymbol{w_0}, \boldsymbol{w_1}$，使得

$$\boldsymbol{y_i} \approx \boldsymbol{w_0} + \boldsymbol{w_1 x_i} = \widehat{\boldsymbol{w}}^{\mathrm{T}} \widehat{\boldsymbol{x}}_i$$

其中$\widehat{\boldsymbol{w}} = (\boldsymbol{w_0}, \boldsymbol{w_1})^{\mathrm{T}}$，$\widehat{\boldsymbol{x}}_i = (\mathbf{1}, \boldsymbol{x_i})^{\mathrm{T}}$.

考虑d维数据点 $\boldsymbol{x_i} = (\boldsymbol{x_{i1}}, \boldsymbol{x_{i2}}, \dots, \boldsymbol{x_{id}})$，

$$\boldsymbol{y_i} \approx \boldsymbol{w_0} + (\boldsymbol{w_1}, \boldsymbol{w_2}, \dots, \boldsymbol{w_d}) \begin{pmatrix} \boldsymbol{x_{i1}} \\ \boldsymbol{x_{i2}} \\ \vdots \\ \boldsymbol{x_{id}} \end{pmatrix} = \widehat{\boldsymbol{w}}^{\mathrm{T}} \widehat{\boldsymbol{x}}_i$$

其中$\widehat{\boldsymbol{w}} = (\boldsymbol{w_0}, \boldsymbol{w_1}, \dots, \boldsymbol{w_d})^{\mathrm{T}}, \widehat{\boldsymbol{x}}_i = (\mathbf{1}, \boldsymbol{x_{i1}}, \dots, \boldsymbol{x_{id}})^{\mathrm{T}}$.

# 多元线性回归

令 $\boldsymbol{y} = (y_1, y_2, \dots y_n)^{\mathrm{T}}$,

$$\boldsymbol{X} = \begin{pmatrix} \widehat{\boldsymbol{x}_1}^{\mathrm{T}} \\ \widehat{\boldsymbol{x}_2}^{\mathrm{T}} \\ \vdots \\ \widehat{\boldsymbol{x}_n}^{\mathrm{T}} \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1d} \\ 1 & x_{21} & x_{22} & \cdots & x_{2d} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nd} \end{pmatrix} \in \mathbb{R}^{n \times (d+1)}$$

$$\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{w}} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} - \begin{pmatrix} \widehat{\boldsymbol{x}_1}^{\mathrm{T}}\widehat{\boldsymbol{w}} \\ \widehat{\boldsymbol{x}_2}^{\mathrm{T}}\widehat{\boldsymbol{w}} \\ \vdots \\ \widehat{\boldsymbol{x}_n}^{\mathrm{T}}\widehat{\boldsymbol{w}} \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} - \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1d} \\ 1 & x_{21} & x_{22} & \cdots & x_{2d} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nd} \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \\ \vdots \\ w_d \end{pmatrix}$$

# 多元线性回归

因此，单变量线性回归模型

$$\underset{\boldsymbol{w_0},\boldsymbol{w_1}}{\operatorname{argmin}}\ \frac{\mathbf{1}}{\boldsymbol{n}}\sum_{\boldsymbol{i}}(\boldsymbol{y_i}-(\boldsymbol{w_0}+\boldsymbol{w_1}\boldsymbol{x_i}))^{\mathbf{2}}$$

可以推广到多变量回归模型

$$\underset{\widehat{\boldsymbol{w}}}{\operatorname{argmin}}\ \mathcal{L}(\widehat{\boldsymbol{w}}) = \|\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{w}}\|_{\mathbf{2}}^{\mathbf{2}}$$

该问题的最优解:

$$\boxed{\widehat{\boldsymbol{w}}^{*} = (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{y}}$$

# 练习题

对于线性回归问题，给定

$$w_0^* = \left(\frac{1}{n}\sum_i y_i\right) - w_1^*\left(\frac{1}{n}\sum_i x_i\right)$$

$$w_1^* = -\frac{1}{n}\sum_i x_i(w_0^* - y_i)/\frac{1}{n}\sum_i x_i^2$$

试推导：

$$w_1^* = \frac{\sum_i y_i(x_i - \frac{1}{n}\sum_i x_i)}{\sum_i x_i^2 - \frac{1}{n}(\sum_i x_i)^2}$$

# 练习题

对于一维数据，试证明下列两种解法的等价性。

公式1. $\boxed{w_1^* = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - (\bar{x})^2}, \quad w_0^* = \bar{y} - w_1^* \bar{x}}$

公式2. $\boxed{\hat{w}^* = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y}$