

作者华校专，曾任阿里巴巴资深算法工程师，现任智易科技首席算法研究员，《Python 大战机器学习》的作者。

这是作者多年以来学习总结的笔记，经整理之后开源于世。目前还有约一半的内容在陆续整理中，已经整理好的内容放置在此。曾有出版社约稿，但是考虑到出版时间周期较长，而且书本购买成本高不利于技术广泛传播，因此作者采取开源的形式。笔记内容仅供个人学习使用，非本人同意不得应用于商业领域。

笔记内容较多，可能有些总结的不到位的地方，欢迎大家探讨。联系方式:huaxz1986@163.com

数学基础

- [1.线性代数基础](#)
 - 一、基本知识
 - 二、向量操作
 - 三、矩阵运算
- [2.概率论基础](#)
 - 一、概率与分布
 - 二、期望
 - 三、方差
 - 四、大数定律及中心极限定理
 - 五、不确定性来源
 - 六、常见概率分布
 - 七、先验分布与后验分布
 - 八、测度论
 - 九、信息论
- [3.数值计算基础](#)
 - 一、数值稳定性
 - 二、Conditioning
 - 三、梯度下降法
 - 四、海森矩阵
 - 四、牛顿法
 - 五、拟牛顿法
 - 六、约束优化
- [4.常用函数](#)
 - 一、sigmoid
 - 二、softplus
 - 三、Gamma 函数和贝塔函数

统计学习

- [0.机器学习简介](#)
 - 一、基本概念
 - 二、监督学习
 - 三、机器学习三要素
- [1.线性代数基础](#)
 - 一、线性回归
 - 二、广义线性模型
 - 三、对数几率回归
 - 四、线性判别分析
 - 五、感知机
- [2.支持向量机](#)
 - 一、线性可分支持向量机
 - 二、线性支持向量机
 - 三、非线性支持向量机
 - 四、支持向量回归
 - 五、SVDD
 - 六、序列最小最优化方法
 - 七、其它讨论
- [3.朴素贝叶斯](#)
 - 一、贝叶斯定理
 - 二、朴素贝叶斯法
 - 三、半朴素贝叶斯分类器
 - 四、其它讨论
- [4.决策树](#)
 - 一、原理
 - 二、特征选择
 - 三、生成算法
 - 四、剪枝算法
 - 五、CART 树

- 六、连续值、缺失值处理
 - 七、多变量决策树
- [5.knn](#)
 - 一、k 近邻算法
 - 二、kd树
- [6.集成学习](#)
 - 一、集成学习误差
 - 二、Boosting
 - 三、Bagging
 - 四、集成策略
 - 五、多样性分析
- [7.梯度提升树](#)
 - 一、提升树
 - 二、xgboost
 - 三、LightGBM
- [8.特征工程](#)
 - 一、缺失值处理
 - 二、特征编码
 - 三、数据标准化、正则化
 - 四、特征选择
 - 五、稀疏表示和字典学习
 - 六、多类分类问题
 - 七、类别不平衡问题
- [9.模型评估](#)
 - 一、泛化能力
 - 二、过拟合、欠拟合
 - 三、偏差方差分解
 - 四、参数估计准则
 - 五、泛化能力评估
 - 六、训练集、验证集、测试集
 - 七、性能度量
 - 七、超参数调节
 - 八、传统机器学习的挑战
- [10.降维](#)
 - 一、维度灾难
 - 二、主成分分析 PCA
 - 三、核化线性降维 KPCA
 - 四、流形学习
 - 五、度量学习
 - 六、概率PCA
 - 七、独立成分分析
 - 八、t-SNE
 - 九、LargeVis
- [11.聚类](#)
 - 一、性能度量
 - 二、原型聚类
 - 三、密度聚类
 - 四、层次聚类
 - 五、谱聚类
- [12.半监督学习](#)
 - 半监督学习
 - 一、生成式半监督学习方法
 - 二、半监督 SVM
 - 三、图半监督学习
 - 四、基于分歧的方法
 - 五、半监督聚类
 - 六、总结
- [13.EM算法](#)
 - 一、示例
 - 二、EM算法原理
 - 三、EM算法与高斯混合模型
 - 四、EM 算法与 kmeans 模型
 - 五、EM 算法的推广
- [14.最大熵算法](#)
 - 一、最大熵模型MEM
 - 二、分类任务最大熵模型
 - 三、最大熵的学习

深度学习

- [0.深度学习简介](#)
 - 一、介绍
 - 二、历史
- [1.机器学习基础](#)
 - 一、基本概念
 - 二、点估计、偏差方差
 - 三、最大似然估计
 - 四、贝叶斯估计
 - 五、随机梯度下降
 - 七、传统机器学习的挑战
 - 八、低维流形
- [2.深度前馈神经网络](#)
 - 一、基础
 - 二、损失函数
 - 三、输出单元
 - 四、隐单元
 - 五、结构设计
 - 六、历史小记
- [3.反向传播算法](#)
 - 一、链式法则
 - 二、反向传播
 - 三、深度前馈神经网络
 - 四、实现
 - 五、应用
 - 六、自动微分
- [4.正则化](#)
 - 一、基本概念
 - 二、参数范数正则化
 - 三、约束正则化
 - 四、数据集增强
 - 五、噪声鲁棒性
 - 六、早停
 - 七、参数共享
 - 八、**dropout**
 - 九、稀疏表达
 - 十、半监督学习与多任务学习
 - 十一、对抗训练
 - 十二、正切传播算法
 - 十三、正则化和欠定问题
- [5.最优优化](#)
 - 一、代价函数
 - 二、神经网络最优化挑战
 - 三、**mini-batch**
 - 四、基本优化算法
 - 五、自适应学习率算法
 - 六、二阶近似方法
 - 七、共轭梯度
 - 八、优化策略和元算法
 - 九、参数初始化策略
- [6.卷积神经网络](#)
 - 一、卷积运算
 - 二、卷积层、池化层
 - 三、基本卷积的变体
 - 四、算法细节
 - 五、历史和现状
- [7.循环神经网络](#)
 - 一、**RNN**计算图
 - 二、循环神经网络
 - 三、长期依赖
 - 四、序列到序列架构
 - 五、递归神经网络
 - 六、回声状态网络
 - 七、**LSTM** 和其他门控RNN
 - 八、外显记忆
- [8.工程实践指导原则](#)
 - 一、性能度量

- 二、默认的基准模型
- 三、决定是否收集更多数据
- 四、选择超参数
- 五、调试策略
- 六、示例：数字识别系统
- 七、数据预处理
- 八、变量初始化
- 九、结构设计

自然语言处理

- [主题模型](#)
 - 一、Unigram Model
 - 二、pLSA Model
 - 三、LDA Model
 - 四、模型讨论
- [词向量](#)
 - 一、向量空间模型 VSM
 - 二、LSA
 - 三、Word2Vec
 - 四、GloVe

计算机视觉

- [图片分类网络](#)
 - 一、LeNet
 - 二、AlexNet
 - 三、VGG-Net
 - 四、Inception
 - 五、ResNet
 - 六、SENet
 - 七、DenseNet
 - 八、小型网络
 - 九、趋势

工具

CRF

- [CRF++](#)
 - 一、安装
 - 二、使用
 - 三、Python接口
 - 四、常见错误

lightgbm

- [lightgbm使用指南](#)
 - 一、安装
 - 二、调参
 - 三、进阶
 - 四、API
 - 五、Docker

xgboost

- [xgboost使用指南](#)
 - 一、安装
 - 二、调参
 - 三、外存计算
 - 四、GPU计算
 - 五、单调约束
 - 六、DART booster
 - 七、Python API

scikit-learn

- [1.预处理](#)
 - 一、特征处理
 - 二、特征选择
 - 三、字典学习
 - 四、PipeLine
- [2.降维](#)
 - 一、PCA
 - 二、MDS
 - 三、Isomap
 - 四、LocallyLinearEmbedding
 - 五、FA
 - 六、FastICA
 - 七、t-SNE
- [3.监督学习模型](#)
 - 一、线性模型
 - 二、支持向量机
 - 三、贝叶斯模型
 - 四、决策树
 - 五、KNN
 - 六、AdaBoost
 - 七、梯度提升树
 - 八、Random Forest
- [4.模型评估](#)
 - 一、数据集切分
 - 二、性能度量
 - 三、验证曲线 && 学习曲线
 - 四、超参数优化
- [5.聚类模型](#)
 - 一、KMeans
 - 二、DBSCAN
 - 三、MeanShift
 - 四、AgglomerativeClustering
 - 五、BIRCH
 - 六、GaussianMixture
 - 七、SpectralClustering
- [6.半监督学习模型](#)
 - 一、标签传播算法

spark

- [1.基础概念](#)
 - 一、核心概念
 - 二、安装和使用
 - 三、pyspark shell
 - 四、独立应用
- [2.rdd使用](#)
 - 一、概述
 - 二、创建 RDD
 - 三、转换操作
 - 四、行动操作
 - 五、其他方法和属性
 - 六、持久化
 - 七、分区
 - 八、混洗
- [3.dataframe使用](#)
 - 一、概述
 - 二、SparkSession
 - 三、DataFrame 创建
 - 四、DataFrame 保存
 - 五、DataFrame
 - 六、Row
 - 七、Column
 - 八、GroupedData
 - 九、functions
- [4.累加器和广播变量](#)
 - 一、累加器
 - 二、广播变量

numpy

- [numpy 使用指南](#)
 - 一、ndarray
 - 二、ufunc 函数
 - 三、函数库
 - 四、数组的存储和加载

scipy

- [scipy 使用指南](#)
 - 一、常数和特殊函数
 - 二、拟合与优化
 - 三、线性代数
 - 四、统计
 - 五、数值积分
 - 六、稀疏矩阵

matplotlib

- [matplotlib 使用指南](#)
 - 一、matplotlib 配置
 - 二、matplotlib Artist
 - 三、基本概念
 - 四、布局
 - 五、Path
 - 六、path effect
 - 七、坐标变换
 - 八、3D 绘图
 - 九、技巧

pandas

- [pandas 使用指南](#)
 - 一、基本数据结构
 - 二、内部数据结构
 - 三、下标存取
 - 四、运算
 - 五、变换
 - 六、数据清洗
 - 七、字符串操作
 - 八、聚合与分组
 - 九、时间序列
 - 十、DataFrame 绘图
 - 十一、移动窗口函数
 - 十二、数据加载和保存

线性代数

一、基本知识

1. 本书中所有的向量都是列向量的形式：

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

2. 矩阵的 **F** 范数：设 $\mathbf{A} = (a_{ij})_{m \times n}$

$$\|\mathbf{A}\|_F = \sqrt{\sum_{i,j} a_{ij}^2}$$

它是向量的 L_2 范数的推广。

3. 矩阵的迹 $\text{tr}(\mathbf{A}) = \sum_i a_{ii}$ 。其性质有：

- $\|\mathbf{A}\|_F = \sqrt{\text{tr}(\mathbf{A}\mathbf{A}^T)}$
- $\text{tr}(\mathbf{A}) = \text{tr}(\mathbf{A}^T)$
- 假设 $\mathbf{A} \in \mathbb{R}^{m \times n}, \mathbf{B} \in \mathbb{R}^{n \times m}$ ，则有：

$$\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$$

- $\text{tr}(\mathbf{ABC}) = \text{tr}(\mathbf{CAB}) = \text{tr}(\mathbf{BCA})$

二、向量操作

1. 一组向量 $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ 是线性相关的：指存在一组不全为零的实数 a_1, a_2, \dots, a_n ，使得：

$$\sum_{i=1}^n a_i \vec{v}_i = \vec{0}$$

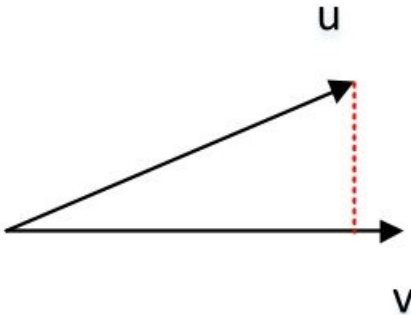
一组向量 $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ 是线性无关的，当且仅当 $a_i = 0, i = 1, 2, \dots, n$ 时，才有

$$\sum_{i=1}^n a_i \vec{v}_i = \vec{0}$$

2. 一个向量空间所包含的最大线性无关向量的数目，称作该向量空间的维数。

3. 三维向量的点积：

$$\vec{u} \cdot \vec{v} = u_x v_x + u_y v_y + u_z v_z = |\vec{u}| |\vec{v}| \cos(\angle(\vec{u}, \vec{v}))$$



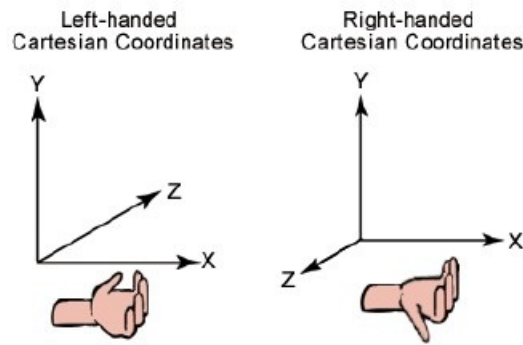
4. 三维向量的叉积：

$$\vec{w} = \vec{u} \times \vec{v} = \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ u_x & u_y & u_z \\ v_x & v_y & v_z \end{vmatrix}$$

其中 $\vec{i}, \vec{j}, \vec{k}$ 分别为 x, y, z 轴的单位向量。

$$\vec{u} = u_x \vec{i} + u_y \vec{j} + u_z \vec{k}, \quad \vec{v} = v_x \vec{i} + v_y \vec{j} + v_z \vec{k}$$

- \vec{u} 和 \vec{v} 的叉积垂直于 \vec{u}, \vec{v} 构成的平面，其方向符合右手规则。
- 叉积的模等于 \vec{u}, \vec{v} 构成的平行四边形的面积
- $\vec{u} \times \vec{v} = -\vec{v} \times \vec{u}$
- $\vec{u} \times (\vec{v} \times \vec{w}) = (\vec{u} \cdot \vec{w})\vec{v} - (\vec{u} \cdot \vec{v})\vec{w}$



5. 三维向量的混合积：

$$\begin{aligned} [\vec{u} \ \vec{v} \ \vec{w}] &= (\vec{u} \times \vec{v}) \cdot \vec{w} = \vec{u} \cdot (\vec{v} \times \vec{w}) \\ &= \begin{vmatrix} u_x & u_y & u_z \\ v_x & v_y & v_z \\ w_x & w_y & w_z \end{vmatrix} = \begin{vmatrix} u_x & v_x & w_x \\ u_y & v_y & w_y \\ u_z & v_z & w_z \end{vmatrix} \end{aligned}$$

- 其物理意义为：以 $\vec{u}, \vec{v}, \vec{w}$ 为三个棱边所围成的平行六面体的体积。当 $\vec{u}, \vec{v}, \vec{w}$ 构成右手系时，该平行六面体的体积为正号。

6. 两个向量的并矢：给定两个向量 $\vec{x} = (x_1, x_2, \dots, x_n)^T, \vec{y} = (y_1, y_2, \dots, y_m)^T$ ，则向量的并矢记作：

$$\vec{x}\vec{y} = \begin{bmatrix} x_1 y_1 & x_1 y_2 & \dots & x_1 y_m \\ x_2 y_1 & x_2 y_2 & \dots & x_2 y_m \\ \vdots & \vdots & \ddots & \vdots \\ x_n y_1 & x_n y_2 & \dots & x_n y_m \end{bmatrix}$$

也记作 $\vec{x} \otimes \vec{y}$ 或者 $\vec{x}\vec{y}^T$

三、矩阵运算

1. 给定两个矩阵 $\mathbf{A} = (a_{ij}) \in \mathbb{R}^{m \times n}, \mathbf{B} = (b_{ij}) \in \mathbb{R}^{m \times n}$ ，定义：

- 阿达马积 **Hadamard product**（又称作逐元素积）：

$$\mathbf{A} \circ \mathbf{B} = \begin{bmatrix} a_{1,1}b_{1,1} & a_{1,2}b_{1,2} & \dots & a_{1,n}b_{1,n} \\ a_{2,1}b_{2,1} & a_{2,2}b_{2,2} & \dots & a_{2,n}b_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1}b_{m,1} & a_{m,2}b_{m,2} & \dots & a_{m,n}b_{m,n} \end{bmatrix}$$

- 克罗内积 **Kronecker product**：

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{1,1}\mathbf{B} & a_{1,2}\mathbf{B} & \dots & a_{1,n}\mathbf{B} \\ a_{2,1}\mathbf{B} & a_{2,2}\mathbf{B} & \dots & a_{2,n}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1}\mathbf{B} & a_{m,2}\mathbf{B} & \dots & a_{m,n}\mathbf{B} \end{bmatrix}$$

2. 设 $\vec{x}, \vec{a}, \vec{b}, \vec{c}$ 为 n 阶向量， $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{X}$ 为 n 阶方阵，则：

$$\frac{\partial(\vec{a}^T \vec{x})}{\partial \vec{x}} = \frac{\partial(\vec{x}^T \vec{a})}{\partial \vec{x}} = \vec{a}$$

$$\frac{\partial(\vec{a}^T \mathbf{X} \vec{b})}{\partial \mathbf{X}} = \vec{a} \vec{b}^T = \vec{a} \otimes \vec{b} \in \mathbb{R}^{n \times n}$$

$$\frac{\partial(\vec{a}^T \mathbf{X}^T \vec{b})}{\partial \mathbf{X}} = \vec{b} \vec{a}^T = \vec{b} \otimes \vec{a} \in \mathbb{R}^{n \times n}$$

$$\frac{\partial(\vec{a}^T \mathbf{X} \vec{a})}{\partial \mathbf{X}} = \frac{\partial(\vec{a}^T \mathbf{X}^T \vec{a})}{\partial \mathbf{X}} = \vec{a} \otimes \vec{a}$$

$$\frac{\partial(\vec{a}^T \mathbf{X}^T \mathbf{X} \vec{b})}{\partial \mathbf{X}} = \mathbf{X}(\vec{a} \otimes \vec{b} + \vec{b} \otimes \vec{a})$$

$$\frac{\partial[(\mathbf{A}\vec{x} + \vec{a})^T \mathbf{C}(\mathbf{B}\vec{x} + \vec{b})]}{\partial \vec{x}} = \mathbf{A}^T \mathbf{C}(\mathbf{B}\vec{x} + \vec{b}) + \mathbf{B}^T \mathbf{C}(\mathbf{A}\vec{x} + \vec{a})$$

$$\frac{\partial(\vec{x}^T \mathbf{A} \vec{x})}{\partial \vec{x}} = (\mathbf{A} + \mathbf{A}^T) \vec{x}$$

$$\frac{\partial[(\mathbf{X}\vec{b} + \vec{c})^T \mathbf{A}(\mathbf{X}\vec{b} + \vec{c})]}{\partial \mathbf{X}} = (\mathbf{A} + \mathbf{A}^T)(\mathbf{X}\vec{b} + \vec{c}) \vec{b}^T$$

$$\frac{\partial(\vec{b}^T \mathbf{X}^T \mathbf{A} \mathbf{X} \vec{c})}{\partial \mathbf{X}} = \mathbf{A}^T \mathbf{X} \vec{b} \vec{c}^T + \mathbf{A} \mathbf{X} \vec{c} \vec{b}^T$$

3. 如果 f 是一元函数，则：

- 其逐元向量函数为：

$$f(\vec{x}) = (f(x_1), f(x_2), \dots, f(x_n))^T$$

- 其逐矩阵函数为：

$$f(\mathbf{X}) = [f(x_{ij})]$$

- 其逐元导数分别为：

$$\begin{aligned} f'(\vec{x}) &= (f'(x_1), f'(x_2), \dots, f'(x_n))^T \\ f'(\mathbf{X}) &= [f'(x_{ij})] \end{aligned}$$

4. 各种类型的偏导数：

- 标量对标量的偏导数

$$\frac{\partial u}{\partial v}$$

- 标量对向量（ n 维向量）的偏导数

$$\frac{\partial u}{\partial \vec{v}} = \left(\frac{\partial u}{\partial v_1}, \frac{\partial u}{\partial v_2}, \dots, \frac{\partial u}{\partial v_n} \right)^T$$

- 标量对矩阵（ $m \times n$ 阶矩阵）的偏导数

$$\frac{\partial u}{\partial \mathbf{V}} = \begin{bmatrix} \frac{\partial u}{\partial v_{1,1}} & \frac{\partial u}{\partial v_{1,2}} & \cdots & \frac{\partial u}{\partial v_{1,n}} \\ \frac{\partial u}{\partial v_{2,1}} & \frac{\partial u}{\partial v_{2,2}} & \cdots & \frac{\partial u}{\partial v_{2,n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial u}{\partial v_{m,1}} & \frac{\partial u}{\partial v_{m,2}} & \cdots & \frac{\partial u}{\partial v_{m,n}} \end{bmatrix}$$

- 向量（ m 维向量）对标量的偏导数

$$\frac{\partial \vec{u}}{\partial v} = \left(\frac{\partial u_1}{\partial v}, \frac{\partial u_2}{\partial v}, \dots, \frac{\partial u_m}{\partial v} \right)^T$$

- 向量（ m 维向量）对向量（ n 维向量）的偏导数（雅可比矩阵，行优先）

$$\frac{\partial \mathbf{u}}{\partial \mathbf{v}} = \begin{bmatrix} \frac{\partial u_1}{\partial v_1} & \frac{\partial u_1}{\partial v_2} & \cdots & \frac{\partial u_1}{\partial v_n} \\ \frac{\partial u_2}{\partial v_1} & \frac{\partial u_2}{\partial v_2} & \cdots & \frac{\partial u_2}{\partial v_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial u_m}{\partial v_1} & \frac{\partial u_m}{\partial v_2} & \cdots & \frac{\partial u_m}{\partial v_n} \end{bmatrix}$$

如果为列优先，则为上面矩阵的转置

- 矩阵($m \times n$ 阶矩阵)对标量的偏导数

$$\frac{\partial \mathbf{U}}{\partial v} = \begin{bmatrix} \frac{\partial U_{1,1}}{\partial v} & \frac{\partial U_{1,2}}{\partial v} & \cdots & \frac{\partial U_{1,n}}{\partial v} \\ \frac{\partial U_{2,1}}{\partial v} & \frac{\partial U_{2,2}}{\partial v} & \cdots & \frac{\partial U_{2,n}}{\partial v} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial U_{m,1}}{\partial v} & \frac{\partial U_{m,2}}{\partial v} & \cdots & \frac{\partial U_{m,n}}{\partial v} \end{bmatrix}$$

- 更复杂的情况依次类推。对于 $\frac{\partial \mathbf{u}}{\partial \mathbf{v}}$ 。根据 `numpy` 的术语：

- 假设 \mathbf{u} 的 `ndim` (维度) 为 d_u

对于标量，`ndim` 为 0；对于向量，`ndim` 为 1；对于矩阵，`ndim` 为 2

- 假设 \mathbf{v} 的 `ndim` 为 d_v

- 则 $\frac{\partial \mathbf{u}}{\partial \mathbf{v}}$ 的 `ndim` 为 $d_u + d_v$

5. 对于矩阵的迹，有下列偏导数成立：

$$\frac{\partial [\text{tr}(f(\mathbf{X}))]}{\partial \mathbf{X}} = (f'(\mathbf{X}))^T$$

$$\frac{\partial [\text{tr}(\mathbf{A}\mathbf{X}\mathbf{B})]}{\partial \mathbf{X}} = \mathbf{A}^T \mathbf{B}^T$$

$$\frac{\partial [\text{tr}(\mathbf{A}\mathbf{X}^T \mathbf{B})]}{\partial \mathbf{X}} = \mathbf{B}\mathbf{A}$$

$$\frac{\partial [\text{tr}(\mathbf{A} \otimes \mathbf{X})]}{\partial \mathbf{X}} = \text{tr}(\mathbf{A})\mathbf{I}$$

$$\frac{\partial [\text{tr}(\mathbf{A}\mathbf{X}\mathbf{B}\mathbf{X})]}{\partial \mathbf{X}} = \mathbf{A}^T \mathbf{X}^T \mathbf{B}^T + \mathbf{B}^T \mathbf{X} \mathbf{A}^T$$

$$\frac{\partial [\text{tr}(\mathbf{X}^T \mathbf{B} \mathbf{X} \mathbf{C})]}{\partial \mathbf{X}} = (\mathbf{B}^T + \mathbf{B}) \mathbf{X} \mathbf{C}^T$$

$$\frac{\partial [\text{tr}(\mathbf{C}^T \mathbf{X}^T \mathbf{B} \mathbf{X} \mathbf{C})]}{\partial \mathbf{X}} = \mathbf{B} \mathbf{X} \mathbf{C} + \mathbf{B}^T \mathbf{X} \mathbf{C}^T$$

$$\frac{\partial [\text{tr}(\mathbf{A}\mathbf{X}\mathbf{B}\mathbf{X}^T \mathbf{C})]}{\partial \mathbf{X}} = \mathbf{A}^T \mathbf{C}^T \mathbf{X} \mathbf{B}^T + \mathbf{C} \mathbf{A} \mathbf{X} \mathbf{B}$$

$$\frac{\partial [\text{tr}((\mathbf{A}\mathbf{X}\mathbf{B} + \mathbf{C})(\mathbf{A}\mathbf{X}\mathbf{B} + \mathbf{C}))]}{\partial \mathbf{X}} = 2\mathbf{A}^T (\mathbf{A}\mathbf{X}\mathbf{B} + \mathbf{C}) \mathbf{B}^T$$

6. 假设 $\mathbf{U} = \mathbf{f}(\mathbf{X})$ 是关于 \mathbf{X} 的矩阵值函数 ($f: \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^{m \times n}$)，且 $g(\mathbf{U})$ 是关于 \mathbf{U} 的实值函数 ($g: \mathbb{R}^{m \times n} \rightarrow \mathbb{R}$)，则下面链式法则成立：

$$\begin{aligned} \frac{\partial g(\mathbf{U})}{\partial \mathbf{X}} &= \left(\frac{\partial g(\mathbf{U})}{\partial x_{i,j}} \right) = \begin{bmatrix} \frac{\partial g(\mathbf{U})}{\partial x_{1,1}} & \frac{\partial g(\mathbf{U})}{\partial x_{1,2}} & \cdots & \frac{\partial g(\mathbf{U})}{\partial x_{1,n}} \\ \frac{\partial g(\mathbf{U})}{\partial x_{2,1}} & \frac{\partial g(\mathbf{U})}{\partial x_{2,2}} & \cdots & \frac{\partial g(\mathbf{U})}{\partial x_{2,n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g(\mathbf{U})}{\partial x_{m,1}} & \frac{\partial g(\mathbf{U})}{\partial x_{m,2}} & \cdots & \frac{\partial g(\mathbf{U})}{\partial x_{m,n}} \end{bmatrix} \\ &= \left(\sum_k \sum_l \frac{\partial g(\mathbf{U})}{\partial u_{k,l}} \frac{\partial u_{k,l}}{\partial x_{i,j}} \right) \\ &= \text{tr} \left[\left(\frac{\partial g(\mathbf{U})}{\partial \mathbf{U}} \right)^T \frac{\partial \mathbf{U}}{\partial x_{i,j}} \right] \end{aligned}$$

