[复习]线性代数与微积分 / Linear Algebra and Calculus refresher

翻译&校正 | 韩信子@ShowMeAI 编辑 | 南乔@ShowMeAI 原文作者 | https://stanford.edu/~shervine 本节原文超链

[1]通用符号 / General Notations

1.1 定义/ Definitions

向量 Vector

记 $x \in \mathbb{R}^n$ 为一个 n 维向量，其中 $x_i \in \mathbb{R}$ 表示第 i 维元素：

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$$

备注： 上述向量 x 可以视为一个 n × 1 矩阵，常被称为"列向量"。

矩阵 Matrix

$A \in \mathbb{R}^{m\times n}$ 为一个 m行 n列的矩阵，其中 $A_{i,j} \in \mathbb{R}$ 表示第 i 行第 j 列的元素：

$$A = \begin{pmatrix} A_{1,1} & \cdots & A_{1,n} \\ \vdots & & \vdots \\ A_{m,1} & \cdots & A_{m,n} \end{pmatrix} \in \mathbb{R}^{m\times n}$$

1.2 常用矩阵/ Main Matrices

单位矩阵 Identity matrix

单位矩阵 $I \in \mathbb{R}^{n\times n}$ 是对角线元素为 1、其余元素为 0 的方阵。

$$I = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix}$$

备注： 对任意 n 阶方阵 $A \in \mathbb{R}^{n\times n}$， $A \times I = I \times A = A$。

对角矩阵 Diagonal matrix

对角矩阵 $D \in \mathbb{R}^{n\times n}$ 是对角线元素非 0、其余元素均为 0 的方阵。

$$D = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & d_n \end{pmatrix}$$

备注： D 也可以被记作 $\mathrm{diag}(d_1, \ldots, d_n)$。

[2]矩阵运算 / Matrix Operations

2.1 乘法/ Multiplication

向量-向量 Vector-vector

存在两种类型的向量-向量乘法：

内积：$x, y \in \mathbb{R}^n$：

$$x^T y = \sum_{i=1}^{n} x_i y_i \in \mathbb{R}$$

外积：$x \in \mathbb{R}^m, y \in \mathbb{R}^n$：

$$xy^T = \begin{pmatrix} x_1 y_1 & \cdots & x_1 y_n \\ \vdots & & \vdots \\ x_m y_1 & \cdots & x_m y_n \end{pmatrix} \in \mathbb{R}^{m\times n}$$

矩阵-向量 Matrix-vector

矩阵 $A \in \mathbb{R}^{m\times n}$ 和 向量 $x \in \mathbb{R}^n$ 的乘积是 m 阶向量 $\mathbb{R}^m$， 满足：

$$Ax = \begin{pmatrix} a_{r,1}^T x \\ \vdots \\ a_{r,m}^T x \end{pmatrix} = \sum_{i=1}^{n} a_{c,i} x_i \in \mathbb{R}^m$$

其中，$a_{r,i}^T$ 是矩阵 A 的行向量， $a_{c,j}$ 是矩阵 A 的列向量。x_i 是矩阵 x 的元素。

矩阵-矩阵 Matrix-matrix

矩阵 $A \in \mathbb{R}^{m\times n}$ 和矩阵 $B \in \mathbb{R}^{n\times p}$ 的乘积是 $m\times p$ 矩阵 $\mathbb{R}^{n\times p}$，满足:

$$AB = \begin{pmatrix} a_{r,1}^T b_{c,1} & \cdots & a_{r,1}^T b_{c,p} \\ \vdots & & \vdots \\ a_{r,m}^T b_{c,1} & \cdots & a_{r,m}^T b_{c,p} \end{pmatrix} = \sum_{i=1}^{n} a_{c,i}\, b_{r,i}^T \in \mathbb{R}^{n\times p}$$

备注：$a_{r,i}^T$，$b_{r,i}^T$ 分别是矩阵 A 和 B 的行向量，$a_{c,j}$，$b_{c,j}$ 分别是矩阵 A 和 B 的行向量。

[3]其他矩阵 / Other Operations

转置 Transpose

矩阵 $A \in \mathbb{R}^{m\times n}$ 的转置记作 A^T。A^T 是对 A 的元素进行翻转:

$$\forall i,j,\ A_{i,j}^T = A_{j,i}$$

备注： 对矩阵 A, B， $(AB)^T = B^T A^T$。

逆 Inverse

可逆方阵 A 的逆矩阵记作 A^{-1}。 A^{-1}是唯一满足以下条件的矩阵:

$$AA^{-1} = A^{-1}A = I$$

备注： 并非所有方阵都可逆。同样，对矩阵 A, B， $(AB)^{-1} = B^{-1}A^{-1}$。

迹 Trace

方阵 A 的迹记作 $\mathrm{tr}(A)$。$\mathrm{tr}(A)$ 是对角元素的和。

$$\mathrm{tr}(A) = \sum_{i=1}^{n} A_{i,i}$$

备注： 对矩阵 A, B， $\mathrm{tr}(A^T) = \mathrm{tr}(A)$， $\mathrm{tr}(AB) = \mathrm{tr}(BA)$。

行列式 Determinant

n 阶方阵 $A \in \mathbb{R}^{n\times n}$的行列式，记作 $|A|$ 或者 $\det(A)$。可以用如下形式递归计算得到，$A_{i,j}$表示第 i 行第 j 列元素，$|A_{\backslash i,\backslash j}|$ 表示 去掉第 i 行第 j 列后的矩阵行列式:

$$\det(A) = |A| = \sum_{j=1}^{n} (-1)^{i+j} A_{i,j} |A_{\backslash i,\backslash j}|$$

备注：当且仅当 $|A| \neq 0$ 时， A 是可逆的。同样，$|AB| = |A||B|$，$|A^T| = |A|$。

[4]矩阵的性质 / Matrix Properties

对称分解 Symmetric decomposition

一个给定矩阵 A 可以用其对阵和反对称部分进行表示:

$$A = \underbrace{\frac{A + A^T}{2}}_{\text{Symmetric}} + \underbrace{\frac{A - A^T}{2}}_{\text{Antisymmetric}}$$

范数 Norm

一个范数是一个函数 $N: V \to [0, +\infty]$，其中 V 是一个向量空间。对于所有 $x, y \in V$:

$$N(x + y) \leq N(x) + N(y)$$

对一个标量 a，$N(ax) = |a|N(x)$ 。

$N(x) = 0$ 时， $x = 0$。

对 $x \in V$，下表总结了最常用的范数:

范数	符号	定义	用例
曼哈顿，L^1	$\|\|x\|\|_1$	$\sum_{i=1}^{n} \|x_i\|$	LASSO
欧几里德，L^2	$\|\|x\|\|_2$	$\sqrt{\sum_{i=1}^{n} x_i^2}$	Ridge
p –范数，L^p	$\|\|x\|\|_p$	$\left(\sum_{i=1}^{n} x_i^p\right)^{\frac{1}{p}}$	赫尔德不等式
无穷，L^∞	$\|\|x\|\|_\infty$	$\max_i \|x_i\|$	一致收

线性相关 Linearly dependence

如果向量集中的一个向量可以定义为其他向量的线性组合，则称向量集为线性相关。

备注： 若无向量可以按照此法表示，则这些向量被称为线性无关。

矩阵的秩 Matrix rank

给定矩阵 A 的秩记作 $\text{rank}(A)$ ，是由列向量生成的向量空间的维度。这等价于 A 的线性无关列向量的最大数目。

半正定矩阵 Positive semi-definite matrix

矩阵 $A \in \mathbb{R}^{n\times n}$ 是半正定矩阵（ PSD ），记作 $A \succcurlyeq \boldsymbol{0}$。

$$A = A^T \quad 和 \quad \forall x \in \mathbb{R}^n, \quad x^T A x \geq 0$$

备注： 类似地，矩阵 A 被称作正定，记作 $A \succ 0$，当它是一个 PSD 矩阵且满足所有非零向量 x， $x^T A x > 0$。

特征值，特征向量 Eigenvalue， eigenvector

给定矩阵 $A \in \mathbb{R}^{n\times n}$， λ 被称作 A 的一个特征值。当存在一个向量 $z \in \mathbb{R}^n \setminus \{0\}$，称作特征向量，满足：

$$Az = \lambda z$$

谱定理 Spectral theorem

令 $A \in \mathbb{R}^{n\times n}$。若 A 是对称的，则 A 可以被一个实正交矩阵 $U \in \mathbb{R}^{n\times n}$ 对角化。

记作 $\Lambda = \text{diag}\left(\lambda_1, \ldots, \lambda_n\right)$：

$$\exists \Lambda \text{diagonal}, \quad A = U\Lambda U^T$$

奇异值分解 Singular-value decomposition

给定一个 $m \times n$ 阶的矩阵 A ，奇异值分解（ SVD ）是一个因子分解机巧，能保证存在酉矩阵 U $m \times m$ ， Σ $m \times n$ 和酉矩阵 V $n \times n$ ，满足：

$$A = U\Sigma V^T$$

[5]矩阵的微积分 / Matrix Calculus

梯度 Gradient

函数 $f: \mathbb{R}^{m\times n} \rightarrow \mathbb{R}$ ，矩阵 $A \in \mathbb{R}^{m\times n}$ 。

函数 f 关于 A 的梯度是一个 $m \times n$ 矩阵， 记作 $\nabla_A f(A)$， 如下表示：

$$\left(\nabla_A f(A)\right)_{i,\,j} = \frac{\partial f(A)}{\partial A_{i,j}}$$

备注： 当且仅当 f 是一个返回标量值的函数时，f 的梯度是有定义的。

Hessian

函数 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ ， 向量 $x \in \mathbb{R}^n$ 。

那么，函数 f 关于向量 x 的海森矩阵(Hessian Matrix)是一个 $n \times n$ 的对称阵，记作 $\nabla_x^2 f(x)$， 并满足：

$$\left(\nabla_x^2 f(x)\right)_{i,j} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}$$

备注： 当且仅当 f 是一个返回标量值的函数时，f 的海森矩阵是有定义的。

梯度运算 Gradient operations

对矩阵 A, B, C， 下列梯度性质值得记住：

$$\nabla_A \text{tr}(AB) = B^T$$

$$\nabla_{A^T} f(A) = \left(\nabla_A f(A)\right)^T$$

$$\nabla_A \text{tr}(ABA^T C) = CAB + C^T A B^T$$

$$\nabla_A |A| = |A|(A^{-1})^T$$

Awesome AI Courses Notes Cheat Sheets

Machine Learning CS229 Stanford	Deep Learning CS230 Stanford	Natural Language Processing CS224n Stanford	Computer Vision CS231n Stanford	Deep Reinforcement Learning CS285 UC Berkeley	Neural Networks for NLP CS11-747 CMU	DL for Self-Driving Cars 6.S094 MIT	... ...

是 ShowMeAI 资料库的分支系列，覆盖最具知名度的 TOP20+门 AI 课程，旨在为读者和学习者提供一整套高品质中文速查表，可以点击【这里】查看。

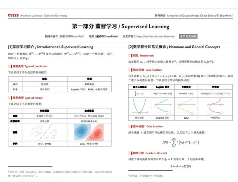
斯坦福大学（Stanford University）的 Machine Learning（CS229）和 Deep Learning（CS230）课程，是本系列的第一批产出。

本批两门课程的速查表由斯坦福大学计算机专业学生 Shervine Amidi 总结整理。原速查表为英文，可点击【这里】查看，ShowMeAI 对内容进行了翻译、校对与编辑排版，整理为当前的中文版本。

有任何建议和反馈，也欢迎通过下方渠道和我们联络 (*￣3￣)

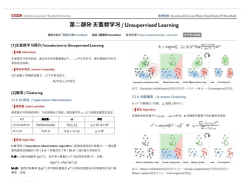
CS229 | Machine Learning @ Stanford University

监督学习
Supervised Learning



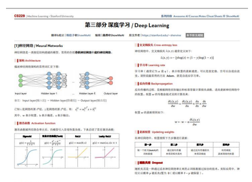
中文速查表链接

无监督学习
Unsupervised Learning



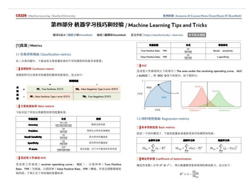
中文速查表链接

深度学习
Deep Learning



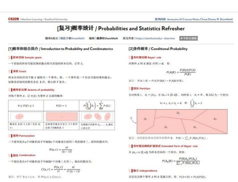
中文速查表链接

机器学习技巧和经验
Tips and Tricks



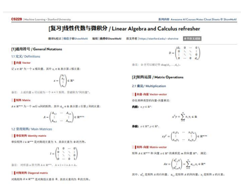
中文速查表链接

概率统计
Probabilities /Statistics



中文速查表链接

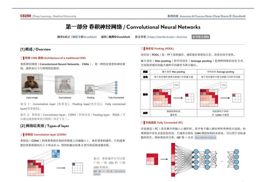
线性代数与微积分
Linear Algebra and Calculus



中文速查表链接

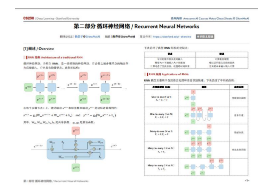
CS230 | Deep Learning @ Stanford University

卷积神经网络
CNN



中文速查表链接

循环神经网络
RNN



中文速查表链接

深度学习技巧与建议
Tips and Tricks

中文速查表链接

GitHub

ShowMeAI

https://github.com/ShowMeAI-Hub/

ShowMeAI 研究中心

扫码回复"**速查表**"
下载**最新**全套资料