

机器学习常用数学符号与记号规范 (Cheat Sheet)

2025 年 9 月 7 日

目录

1 总览与风格

为便于在全书/各章快速理解公式，本文件约定统一的符号风格：

- 标量 (scalar): 斜体小写, 如 a, b, c, α, β ; 常数可用非斜体 e, π 。
- 向量 (vector): 粗体小写, 如 \mathbf{x}, \mathbf{w} ; 分量 x_i 。
- 矩阵 (matrix): 粗体大写, 如 \mathbf{X}, \mathbf{W} ; 元素 $[\mathbf{X}]_{ij}$ 、第 i 行 $\mathbf{X}_{i,:}$ 、第 j 列 $\mathbf{X}_{:,j}$ 。
- 张量 (tensor): 粗体花体或黑体, 如 \mathcal{X} (本书中常直接以粗体大写表示)。
- 集合/空间: 花体大写, 如 \mathcal{X}, \mathcal{Y} ; 数集用黑板粗体 $\mathbb{R}, \mathbb{N}, \mathbb{Z}$ 。
- 随机变量: 大写 X , 其取值小写 x ; 向量同理 \mathbf{X} 与 \mathbf{x} 。
- 参数/模型: 参数向量 $\boldsymbol{\theta}$ 、权重 \mathbf{w} 、偏置 b ; 模型 $f(\mathbf{x}; \boldsymbol{\theta})$ 。
- 数据维度: 样本数 n 、特征维度 d 、类别数 K 。

2 线性代数与索引

- 单位矩阵 \mathbf{I}_d ; 全零向量 $\mathbf{0}$; 全一向量 $\mathbf{1}$ 。
- 转置 \mathbf{X}^\top ; 逆 \mathbf{X}^{-1} (若存在); 伪逆 \mathbf{X}^\dagger 。
- 迹 $\text{tr}(\mathbf{X})$; 秩 $\text{rank}(\mathbf{X})$; $\text{diag}(\mathbf{v})$ 生成对角矩阵。
- 内积 $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^\top \mathbf{y}$; Hadamard 乘积 $\mathbf{X} \odot \mathbf{Y}$; Kronecker 乘积 $\mathbf{X} \otimes \mathbf{Y}$ 。

- 特征分解 $\mathbf{X} = \mathbf{Q} \text{diag}(\boldsymbol{\lambda}) \mathbf{Q}^\top$; 奇异值分解 $\mathbf{X} = \mathbf{U} \text{diag}(\boldsymbol{\sigma}) \mathbf{V}^\top$ 。
- 切片与索引: 标量 x_i 、列向量 $\mathbf{X}_{:,j}$ 、行向量 $\mathbf{X}_{i,:}$ 。

3 范数、距离与相似度

- 向量 ℓ_p 范数: $\|\mathbf{x}\|_p = (\sum_i |x_i|^p)^{1/p}$; 常见: $\|\mathbf{x}\|_1$, $\|\mathbf{x}\|_2$, $\|\mathbf{x}\|_\infty$ 。
- 矩阵范数: Frobenius $\|\mathbf{X}\|_F = \sqrt{\sum_{ij} X_{ij}^2}$; 谱范数 $\|\mathbf{X}\|_2 = \sigma_{\max}(\mathbf{X})$ 。
- 欧氏距离 $\|\mathbf{x} - \mathbf{y}\|_2$; 曼哈顿距离 $\|\mathbf{x} - \mathbf{y}\|_1$; 马氏距离 $\sqrt{(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})}$ 。
- 余弦相似度 $\frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\|_2 \|\mathbf{y}\|_2}$ 。

4 微积分与优化记号

- 梯度 $\nabla f(\mathbf{x}) \in \mathbb{R}^d$; 雅可比 $\mathbf{J}_f(\mathbf{x})$; Hessian $\nabla^2 f(\mathbf{x})$ 。
- 偏导 $\frac{\partial f}{\partial x_i}$; 全微分 df ; 链式法则与向量化求导遵循矩阵维度一致性。
- 极值算子: $\arg \min_{\boldsymbol{\theta}} f(\boldsymbol{\theta})$ 、 $\arg \max$; 约束优化用拉格朗日函数 $\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda})$ 与 KKT 条件。
- 常见更新: 梯度下降 $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} - \eta \nabla f(\boldsymbol{\theta})$; 动量/Adam 超参 $\beta_1, \beta_2, \epsilon$ 。

5 概率与统计记号

- 概率与密度: $\mathbb{P}(A)$, $p_X(x)$ (密度/质量函数), 条件 $p(y | x)$, 联合 $p(x, y)$ 。
- 期望 $\mathbb{E}[Z]$; 方差 $\text{Var}(Z)$; 协方差 $\text{Cov}(X, Y)$; 相关系数 ρ_{XY} 。
- 独立/条件独立: $X \perp\!\!\!\perp Y$, $X \perp\!\!\!\perp Y | Z$ 。
- 大小符号: \sim 表“服从”: $X \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 。

5.1 常见分布

- 高斯: $\mathcal{N}(\mu, \sigma^2)$ 、多元 $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 。
- 伯努利/二项/多项: $\text{Ber}(p)$ 、 $\text{Bin}(n, p)$ 、 $\text{Mult}(n, \boldsymbol{\pi})$; 类别: $\text{Cat}(\boldsymbol{\pi})$ 。
- 均匀: $\mathcal{U}(a, b)$; 指数: $\text{Exp}(\lambda)$; Gamma: $\text{Ga}(\alpha, \beta)$; Beta: $\text{Be}(\alpha, \beta)$; Dirichlet: $\text{Dir}(\boldsymbol{\alpha})$ 。

5.2 信息论

- 熵: $H(X) = -\sum_x p(x) \log p(x)$; 条件熵 $H(Y | X)$; 互信息 $I(X; Y)$ 。
- KL 散度: $D_{\text{KL}}(p||q) = \sum_x p(x) \log \frac{p(x)}{q(x)}$; 交叉熵 $H(p, q)$ 。

6 数据与模型记号

- 数据集: $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^n$; $\mathbf{x}_i \in \mathbb{R}^d$, $y_i \in \mathcal{Y}$ (回归 $\subseteq \mathbb{R}$, 分类 $\in \{1, \dots, K\}$)。
- 设计矩阵与标签向量: $\mathbf{X} \in \mathbb{R}^{n \times d}$, 第 i 行为 \mathbf{x}_i^\top ; $\mathbf{y} \in \mathbb{R}^n$ 或 one-hot $\mathbf{Y} \in \{0, 1\}^{n \times K}$ 。
- 模型: $f(\mathbf{x}; \boldsymbol{\theta})$; 参数 $\boldsymbol{\theta}$; 预测 \hat{y} 或 $\hat{\mathbf{y}}$ 。
- 损失: 单样本 $\ell(\hat{y}, y)$; 经验风险 $\hat{R}(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^n \ell(f(\mathbf{x}_i; \boldsymbol{\theta}), y_i)$ 。
- 正则与目标: $\Omega(\boldsymbol{\theta})$ (如 $\lambda \|\mathbf{w}\|_2^2$); 总体目标 $J(\boldsymbol{\theta}) = \hat{R} + \Omega$ 。

7 分类与回归常用函数

- Sigmoid: $\sigma(t) = 1/(1 + e^{-t})$; Softmax: $\text{softmax}(\mathbf{z})_k = \frac{e^{z_k}}{\sum_j e^{z_j}}$ 。
- 线性回归: $\hat{y} = \mathbf{w}^\top \mathbf{x} + b$; L2 损失: $\frac{1}{2}(\hat{y} - y)^2$ 。
- 逻辑回归: $p(y = 1 | \mathbf{x}) = \sigma(\mathbf{w}^\top \mathbf{x} + b)$; 对数似然与交叉熵如各章所述。
- SVM 间隔: $\gamma = y(\mathbf{w}^\top \mathbf{x} + b)/\|\mathbf{w}\|_2$; Hinge 损失: $\max(0, 1 - y f(\mathbf{x}))$ 。

8 深度学习记号

- 层与维度: 第 ℓ 层权重 $\mathbf{W}^{(\ell)}$ 、偏置 $\mathbf{b}^{(\ell)}$; 预激活 $\mathbf{z}^{(\ell)}$ 、激活 $\mathbf{a}^{(\ell)}$ 。
- 小批量: 批大小 B ; 样本索引集合 \mathcal{B} 。
- 反向传播: $\frac{\partial \mathcal{L}}{\partial \mathbf{W}^{(\ell)}}$ 、 $\frac{\partial \mathcal{L}}{\partial \mathbf{b}^{(\ell)}}$; 链式法则逐层传递。
- 常见激活: ReLU $\max(0, t)$ 、Leaky-ReLU、Tanh、Sigmoid; 归一化/正则化: Batch-Norm、Dropout (保留率 p)。

9 评估与度量

- 二分类：TP、TN、FP、FN；精准率 $\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$ ，召回率 $\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$ ， $\text{F1} = 2 \frac{\text{PR}}{\text{P} + \text{R}}$ 。
- ROC/PR：TPR、FPR；AUC。
- 回归：MSE、MAE、 R^2 。
- 交叉验证： K 折；训练/验证/测试划分： $\mathcal{D}_{\text{train}}/\mathcal{D}_{\text{val}}/\mathcal{D}_{\text{test}}$ 。

10 集合与逻辑记号

- 基本运算：并 $A \cup B$ 、交 $A \cap B$ 、差 $A \setminus B$ 、补 A^c ；基数 $|A|$ 。
- 指示函数： $\mathbf{1}_A(x)$ ，若 $x \in A$ 则为 1 否则为 0。
- 映射与函数： $f: \mathcal{X} \rightarrow \mathcal{Y}$ ；复合 $f \circ g$ ；恒等映射 id 。

11 速查表（选）

记号	含义
n, d, K	样本数、特征维、类别数
$\mathbf{x} \in \mathbb{R}^d$	特征向量
$\mathbf{X} \in \mathbb{R}^{n \times d}$	设计矩阵（行：样本；列：特征）
$y \in \mathbb{R}$ 或 $\{1, \dots, K\}$	标签（回归/分类）
$f(\mathbf{x}; \boldsymbol{\theta})$	模型
$\ell(\hat{y}, y)$	单样本损失
$\hat{R}(\boldsymbol{\theta})$	经验风险
$\Omega(\boldsymbol{\theta})$	正则项
$J(\boldsymbol{\theta})$	目标函数 $\hat{R} + \Omega$
$\mathbb{E}[\cdot]$, Var , Cov	期望、方差、协方差
$D_{\text{KL}}(p\ q)$	KL 散度
$\text{tr}(\cdot)$, $\text{rank}(\cdot)$	迹、秩
$\ \cdot\ _1$, $\ \cdot\ _2$	常用范数

12 备注

不同书籍/论文偶有记号差异。本文件以“可读性与工程一致性”为先：

- 粗体表示向量/矩阵，尽量避免与随机变量大小写规则冲突；

- 概率分布用 $p(\cdot)$ 或特定记号（如 \mathcal{N} 、Ber）清晰区分；
- 若章节有特别说明，以章节内局部约定为准，但建议与本表保持一致。