

人工智能与机器学习基础

HW0: 数学基础诊断

中国科学技术大学

2026 年秋季学期 发布日期与截止日期待定

说明

本次作业用于帮助大家检查学习本课程所需的数学基础，内容限于线性代数、微积分、概率统计和基础优化。预计完成时间为 **2-3 小时**。

请写出必要的计算过程；仅给出最终数值可能无法获得全部分数。可以查阅教材和课程提供的复习资料，但请独立完成。除简单算术核对外，不需要使用编程或符号计算软件。

本作业满分 100 分，最终只记录完成状态，不计入正式书面作业平均分。

如对题意、提交格式或数学符号有疑问，请先搜索并通过课程仓库的 GitHub Issues 提问：

https://github.com/Intelligent114/ALML_public/issues/new/choose

提问时注明 HW0 和具体题号；不要在公开 Issue 中发布本人学号、个人邮箱或完整作业答案。

提交一个 PDF 文件，命名为：

学号-姓名-HW0.pdf.

1 向量与矩阵

[20 分]

设数据矩阵 $X \in \mathbb{R}^{n \times d}$ ，参数向量 $w \in \mathbb{R}^d$ ，标签向量 $y \in \mathbb{R}^n$ ，并定义预测向量

$$\hat{y} = Xw.$$

(a) 分别写出 $X^\top X$ 、 $X^\top y$ 、 $XX^\top$ 和 $\hat{y} - y$ 的维度。

[6 分]

(b) 取

$$X = \begin{bmatrix} 1 & 2 \\ 0 & -1 \\ 2 & 1 \end{bmatrix}, \quad w = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad y = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}.$$

计算 Xw 、 $Xw - y$ 以及 $X^\top(Xw - y)$ 。

[8 分]

(c) 对向量 $a = (1, 2, -1)^\top$ 和 $b = (2, 0, 2)^\top$ ，计算它们的内积、 ℓ_2 范数以及余弦相似度

$$\frac{a^\top b}{\|a\|_2 \|b\|_2}.$$

根据结果说明两向量是否正交。

[6 分]

2 特征值与二次型

[15 分]

给定对称矩阵

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}.$$

(a) 求 A 的两个特征值，并为每个特征值写出一个对应的单位特征向量。 [8 分]

(b) 判断 A 是否为正定矩阵，并说明理由。你可以使用特征值，也可以直接整理二次型

$$\mathbf{x}^\top A \mathbf{x}, \quad \mathbf{x} = (x_1, x_2)^\top.$$

[5 分]

(c) 不显式计算矩阵逆，利用特征向量或直接求解线性方程，求满足

$$A\mathbf{x} = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$$

的 $\mathbf{x}$ 。

[2 分]

3 多元微积分与链式法则

[25 分]

设 $W \in \mathbb{R}^{m \times d}$ 、 $\mathbf{x} \in \mathbb{R}^d$ 、 $\mathbf{b}, \mathbf{y} \in \mathbb{R}^m$ ，定义

$$\mathbf{z} = W\mathbf{x} + \mathbf{b}, \quad L = \frac{1}{2} \|\mathbf{z} - \mathbf{y}\|_2^2.$$

(a) 将 L 写成求和形式，即写成关于 z_i 和 y_i 的表达式，并求 $\partial L / \partial z_i$ 。 [5 分]

(b) 使用链式法则求 $\nabla_{\mathbf{b}} L$ 和 $\nabla_{\mathbf{x}} L$ 。请同时写出它们的维度。 [8 分]

(c) 求 $\nabla_W L$ 。答案应为一个 $m \times d$ 矩阵，并尽量写成向量外积的形式。 [8 分]

(d) Sigmoid 函数定义为

$$\sigma(t) = \frac{1}{1 + e^{-t}}.$$

证明

$$\sigma'(t) = \sigma(t)(1 - \sigma(t)).$$

[4 分]

4 概率与统计

[25 分]

(a) 某事件 D 的先验概率为 $P(D) = 0.1$ 。一个检测方法满足

$$P(+ | D) = 0.9, \quad P(- | D^c) = 0.95.$$

计算 $P(+)$ 和 $P(D | +)$ 。请写出使用的全概率公式和贝叶斯公式。

[10 分]

(b) 离散随机变量 X 的分布如下：

x	-1	0	2
$P(X = x)$	0.2	0.5	0.3

计算 $\mathbb{E}[X]$ 、 $\mathbb{E}[X^2]$ 和 $\text{Var}(X)$ 。 [9 分]

(c) 判断下面的说法是否正确，并用一句话说明理由：

若 $\text{Cov}(X, Y) = 0$ ，则随机变量 X 与 Y 一定相互独立。

如果你认为不正确，请给出一个简单反例。 [6 分]

5 梯度下降 [15 分]

考虑函数

$$f(w_1, w_2) = (w_1 - 1)^2 + 2(w_2 + 1)^2.$$

(a) 求梯度 $\nabla f(w_1, w_2)$ ，并写出函数的全局最小点。 [5 分]

(b) 从 $\mathbf{w}^{(0)} = (0, 0)^\top$ 出发，使用学习率 $\eta = 0.25$ 做一步梯度下降：

$$\mathbf{w}^{(1)} = \mathbf{w}^{(0)} - \eta \nabla f(\mathbf{w}^{(0)}).$$

求 $\mathbf{w}^{(1)}$ 以及 $f(\mathbf{w}^{(0)})$ 、 $f(\mathbf{w}^{(1)})$ 。 [6 分]

(c) 若改用学习率 $\eta = 1$ ，同样从 $\mathbf{w}^{(0)}$ 做一步更新，目标函数会下降还是上升？这个例子说明学习率过大可能产生什么问题？ [4 分]

完成情况反馈（不计分）

请在提交末尾简要填写：

1. 完成本次作业大约花费了多长时间？
2. 哪一部分最不熟悉：线性代数、微积分、概率统计，还是优化？
3. 是否希望课程提供某一部分的补充讲义或习题课？