人工智能代数学基础大作业：线性代数在概率统计中的运用

112401520 王郑涵 152401213 邱镇海

2026 年 1 月 10 日

目录

1 引言

1.1 选题

本报告选取了 Introduction to Linear Algebra 中 Chapter 12 的内容，报告包括三个部分：

- 学习笔记
- 课后习题
- 学习心得

1.2 分工

- 王郑涵：主要负责了学习笔记书写和编程实验实现。
- 邱镇海：主要负责了课后习题解答书写和编程实验书写。

2 学习笔记

2.1 均值，方差和概率

2.1.1 基本定义

我们从基本的几个统计量出发。

- 样本均值：
  对于样本 $x_1, \ldots, x_N$，样本均值 $m = \dfrac{1}{N}\sum_{i=1}^{N} x_i$。
- 方差和样本方差：
  方差描述了数据围绕均值的离散程度。样本方差描述了相对于样本均值的距离，方差描述了相对于期望的平均距离。

方差定义为 $\sigma^2 = \mathbb{E}((x-m)^2) = \sum_{i=1}^{n} p_i(x_i - m)^2$。
样本方差定义为 $\mathbb{S}^2 = \dfrac{1}{N-1}\sum_{i=1}^{N}(x_i - m)^2$。$\mathbb{S}^2$ 是 σ^2 的无偏估计。

- 协方差：$\sigma_{12} = \sum_i \sum_j p_{ij}(x_i - m_1)(y_j - m_2)$。
- 相关系数：$\rho = \dfrac{\sigma_{xy}}{\sigma_x \sigma_y}$，其中 σ_{xy} 为 x，y 的协方差，σ_x，σ_y 分别为 x，y 的标准差。

对于概率密度函数，我们能求出均值和方差：

$$m = \mathbb{E}[x] = \int xp(x)dx$$
$$\sigma^2 = \mathbb{E}[(x-m)^2] = \int (x-m)^2 p(x)dx$$

2.1.2 Monte Carlo Estimetion Methods

科学计算必须处理数据中的误差。金融计算必须处理不确定的数值和不确定的预测。我们需要接受输入中的不确定性，并估计输出中的方差。

通常我们并不知道概率分布 $p(x)$，我们能做的就是通过不同的输入 b 得到若干个输出 $x(b)$，然后通过样本均值来估计期望。这就是蒙特卡洛方法最简单的形式，这样的误差大概是 $\frac{1}{\sqrt{N}}$ 级别的，由于每次得到输出的计算成本可能很大，这样的效率是不够好的。

多层蒙特卡洛（Multilevel Monte Carlo）是更加优秀的方法，我们找到另一个函数 $y(b)$ 使得计算它的成本比 $x(b)$ 小得多，并且两者十分接近。那么我们便可以用 N 次关于 $y(b)$ 的计算和 $N^* < N$ 次关于 $x(b)$ 的计算来估计 $\mathbb{E}[X]$。

$$\text{2-level Monte Carlo: } \mathbb{E}[x] \approx \frac{1}{N}\sum_{k=1}^{N} y(b_k) + \frac{1}{N^*}\sum_{k=1}^{N^*}[x(b_k) - y(b_k)]. \tag{1}$$

设 $y(b)$ 计算成本为 C，$x(b)$ 计算成本为 C^*，在总成本固定的前提下，我们可以最小化方差，最佳比例为 $\dfrac{N^*}{N} = \sqrt{\dfrac{C}{C^*}}\dfrac{\sigma^*}{\sigma}$。

2.2 协方差矩阵

2.2.1 定义与性质

定义：已知试验的结果取值向量 $\boldsymbol{X}$ 以及均值 $\bar{\boldsymbol{X}} = \mathbb{E}[\boldsymbol{X}]$，协方差矩阵为

$$\boldsymbol{V} = \mathbb{E}[(\boldsymbol{X} - \bar{\boldsymbol{X}})(\boldsymbol{X} - \bar{\boldsymbol{X}})^{\mathrm{T}}] \tag{2}$$

2.2.2 性质

- 协方差矩阵**至少**是半正定的（当变量相互独立时，它是正定的，且是对角矩阵)。协方差矩阵同时也是对称矩阵。
- 由于协方差矩阵 V 是对称矩阵，我们可以把它分解成 $\boldsymbol{V} = \boldsymbol{Q}\Lambda\boldsymbol{Q}^{\mathrm{T}}$。这样我们得到了若干个特征值 $\lambda_i \geq 0$ 和若干个正交的单位特征向量 q_i，原来的若干个试验相当于分解得到的若干个独立试验的线性组合。
- 对原向量 $\boldsymbol{X}$ 做线性变换得到 $\boldsymbol{Z} = \boldsymbol{AX}$，那么 $\boldsymbol{Z}$ 的协方差矩阵为

$$\boldsymbol{V_Z} = \boldsymbol{AV_XA}^{\mathrm{T}} \tag{3}$$

2.3 加权最小二乘与卡尔曼滤波

2.3.1 加权最小二乘

我们知道对于一个无解的线性方程 $\boldsymbol{Ax} = \boldsymbol{b}$，我们选择 $\hat{\boldsymbol{x}}$ 来最小化误差 $||\boldsymbol{b} - \boldsymbol{Ax}||^2$，有方程 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\hat{\boldsymbol{x}} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$。

当 $\boldsymbol{b}$ 的测量误差是**相互独立的随机变量**，均值为 0，方差为 1 且服从正态分布，那么最小二乘法是合适的。当其不独立或方差不同时，需要使用加权最小二乘法。

误差为

$$\boldsymbol{E} = (\boldsymbol{b} - \boldsymbol{Ax})^{\mathrm{T}}\boldsymbol{V}^{-1}(\boldsymbol{b} - \boldsymbol{Ax})$$

方程变为

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{V}^{-1}\boldsymbol{A}\hat{\boldsymbol{x}} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{V}^{-1}\boldsymbol{b} \tag{4}$$

为了衡量整个试验的可靠性，我们同样考虑 $\hat{\boldsymbol{x}}$ 的均方误差，并将其与输入的协方差矩阵建立联系。我们有矩阵 $\boldsymbol{W}$：

$$\boldsymbol{W} = \mathbb{E}[(\hat{\boldsymbol{x}} - \boldsymbol{x})(\hat{\boldsymbol{x}} - \boldsymbol{x})^{\mathrm{T}}] = (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{V}^{-1}\boldsymbol{A})^{-1} \tag{5}$$

假设 $\boldsymbol{b}$ 的协方差矩阵为 $\boldsymbol{V}$，$\hat{\boldsymbol{x}} = \boldsymbol{L}\boldsymbol{b}$。由 (3) 可知，$\hat{\boldsymbol{x}}$ 的协方差矩阵为 $\boldsymbol{L}\boldsymbol{V}\boldsymbol{L}^{\mathrm{T}}$。由 (4) 可知

$$\boldsymbol{L}\boldsymbol{V}\boldsymbol{L}^{\mathrm{T}} = (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{V}^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{V}^{-1}\ \boldsymbol{V}\ \boldsymbol{V}^{-1}\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{V}^{-1}\boldsymbol{A})^{-1} = (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{V}^{-1}\boldsymbol{A})^{-1}$$

2.3.2 卡尔曼滤波

考虑一个动态问题：每次新增一个观测方程，如何根据之前的观测方程，状态转移方程和最优估计 $\hat{x}_{k-1}$ 来更新最优估计 $\hat{x}_k$。

$$A_0x_0 = b_0 \quad x_1 = F_0x_0 \quad A_1x_1 = b_1 \quad x_2 = F_1x_1 \quad A_2x_2 = b_2$$

$$x_{\mathrm{new}} = F_{\mathrm{old}}x_{\mathrm{old}}$$

考虑简化的问题：不考虑矩阵 F。我们初始有 $A_0x_0 = b_0$，有加权最小二乘方程

$$A_0^{\mathrm{T}}V_0^{-1}A_0\hat{x}_0 = A_0^{\mathrm{T}}V_0^{-1}b_0 \tag{6}$$

现在新增一个方程

$$\begin{bmatrix} A_0 \\ A_1 \end{bmatrix} \hat{x}_1 = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}$$

有加权方程

$$\begin{bmatrix} A_0^T & A_1^T \end{bmatrix} \begin{bmatrix} V_0^{-1} & \\ & V_1^{-1} \end{bmatrix} \begin{bmatrix} A_0 \\ A_1 \end{bmatrix} \hat{x}_1 = \begin{bmatrix} A_0^T & A_1^T \end{bmatrix} \begin{bmatrix} V_0^{-1} & \\ & V_1^{-1} \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}. \tag{7}$$

有卡尔曼更新方程

$$\hat{x}_1 = \hat{x}_0 + K_1(b_1 - A_1\hat{x}_0) \tag{8}$$

其中 K_1 是卡尔曼增益矩阵，K_1 又可以从方差-协方差矩阵 W_1 的逆得到：

$$W_1^{-1} = W_0^{-1} + A_1^{\mathrm{T}}V_1^{-1}A_1 \tag{9}$$

$$K_1 = W_1A_1^{\mathrm{T}}V_1^{-1} \tag{10}$$

3 课后习题解答

3.1 习题 1

已知样本方差 S^2 有如下两种等价形式（其中 m 为样本均值）:

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - m)^2 = \frac{1}{N-1}\left[\left(\sum_{i=1}^{N} x_i^2\right) - Nm^2\right].$$

请验证：期望方差（加权方差）σ^2 也有对应的等价恒等式。设

$$m = \sum_i p_i x_i,$$

证明

$$\sigma^2 = \sum_i p_i (x_i - m)^2 = \left(\sum_i p_i x_i^2\right) - m^2.$$

解答:

从期望方差（加权方差）的定义出发:

$$\sigma^2 = \sum_i p_i (x_i - m)^2.$$

先展开平方项:

$$\begin{aligned}\sigma^2 &= \sum_i p_i \big(x_i^2 - 2x_i m + m^2\big) \\ &= \sum_i p_i x_i^2 - 2m\sum_i p_i x_i + m^2 \sum_i p_i.\end{aligned}$$

注意到给定条件

$$\sum_i p_i x_i = m \quad 且 \quad \sum_i p_i = 1,$$

代入可得

$$\begin{aligned}\sigma^2 &= \sum_i p_i x_i^2 - 2m(m) + m^2(1)\\ &= \sum_i p_i x_i^2 - 2m^2 + m^2\\ &= \sum_i p_i x_i^2 - m^2.\end{aligned}$$

因此恒等式成立：

$$\sigma^2 = \left(\sum_i p_i x_i^2\right) - m^2.$$

3.2 习题 2

题目：

掷一枚公平硬币 N 次，得到 i 次正面朝上的概率为

$$p_i = \frac{b_i}{2^N}, \qquad b_i = \binom{N}{i} = \frac{N!}{i!(N-i)!}.$$

由于

$$\sum_{i=0}^{N} b_i = (1+1)^N = 2^N,$$

所以

$$\sum_{i=0}^{N} p_i = 1.$$

并且注意到 $b_i = b_{N-i}$。

问题： 证明该分布的均值

$$m = 0 \cdot p_0 + 1 \cdot p_1 + \cdots + N \cdot p_N = \sum_{i=0}^{N} i\, p_i$$

等于 $\dfrac{N}{2}$。
解答：

我们计算

$$m = \sum_{i=0}^{N} i\, p_i = \frac{1}{2^N} \sum_{i=0}^{N} i \binom{N}{i}.$$

利用组合恒等式

$$i \binom{N}{i} = N \binom{N-1}{i-1},$$

于是

$$\sum_{i=0}^{N} i \binom{N}{i} = \sum_{i=1}^{N} N \binom{N-1}{i-1} = N \sum_{j=0}^{N-1} \binom{N-1}{j} = N \cdot 2^{N-1},$$

其中令 $j = i - 1$，并用到了

$$\sum_{j=0}^{N-1} \binom{N-1}{j} = (1+1)^{N-1} = 2^{N-1}.$$

因此

$$m = \frac{1}{2^N} \cdot N \cdot 2^{N-1} = \frac{N}{2}.$$

所以均值确实为 $\dfrac{N}{2}$。

3.3 习题 3

题目：

对任意函数 $f(x)$，其期望为

$$\mathbb{E}[f] = \sum_i p_i f(x_i) \quad 或 \quad \mathbb{E}[f] = \int p(x) f(x)\, dx$$

(分别对应离散概率与连续概率)。设均值为 $\mathbb{E}[x]=m$,方差为

$$\mathbb{E}\big[(x-m)^2\big]=\sigma^2.$$

求 $\mathbb{E}[x^2]$。

解答:

由方差定义

$$\sigma^2=\mathbb{E}\big[(x-m)^2\big]=\mathbb{E}\big[x^2-2mx+m^2\big]=\mathbb{E}[x^2]-2m\mathbb{E}[x]+m^2.$$

又因为 $\mathbb{E}[x]=m$,代入得

$$\sigma^2=\mathbb{E}[x^2]-2m^2+m^2=\mathbb{E}[x^2]-m^2.$$

因此

$$mathbbE[x^2]=\sigma^2+m^2.$$

3.4 习题 4

题目:

对于 $M=3$ 个相互独立的实验,其均值分别为 m_1,m_2,m_3,方差分别为 $\sigma_1^2,\sigma_2^2,\sigma_3^2$。求它们的协方差矩阵 V。

解答:

由于三个实验相互独立,因此不同实验之间协方差为 0:

$$\mathrm{Cov}(X_i,X_j)=0,\quad i\neq j$$

每个实验的方差为:

$$\mathrm{Var}(X_i)=\sigma_i^2$$

因此协方差矩阵为对角矩阵：

$$V = \begin{pmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{pmatrix}$$

3.5 习题 5

题目：

对同一个变量 x 进行两次测量，得到两条观测方程

$$x = b_1, \qquad x = b_2.$$

假设两次测量误差均值为零，方差分别为 σ_1^2 与 σ_2^2，且两次误差相互独立。

因此误差协方差矩阵

$$V = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}.$$

将两条方程写成 $Ax = b$ 的形式（其中 A 是 2×1 矩阵）。

参照教材 Example 1，求基于 b_1, b_2 的最优估计 $\hat{x}$，以及估计误差的方差

$$\mathbb{E}\left[\hat{x}\hat{x}^T\right].$$

解答：

(1) 写成矩阵形式

两条观测方程为

$$x = b_1, \qquad x = b_2.$$

写成 $Ax = b$ 形式：

$$A = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \qquad b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}.$$

误差协方差矩阵为

$$V=\begin{pmatrix}\sigma_1^2 & 0\\ 0 & \sigma_2^2\end{pmatrix},\qquad V^{-1}=\begin{pmatrix}\frac{1}{\sigma_1^2} & 0\\ 0 & \frac{1}{\sigma_2^2}\end{pmatrix}.$$

(2) 最小二乘（加权）最优估计

加权最小二乘解为

$$\hat{x}=(A^TV^{-1}A)^{-1}A^TV^{-1}b.$$

先计算：

$$A^TV^{-1}A=\begin{pmatrix}1 & 1\end{pmatrix}\begin{pmatrix}\frac{1}{\sigma_1^2} & 0\\ 0 & \frac{1}{\sigma_2^2}\end{pmatrix}\begin{pmatrix}1\\ 1\end{pmatrix}=\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}.$$

再计算：

$$A^TV^{-1}b=\begin{pmatrix}1 & 1\end{pmatrix}\begin{pmatrix}\frac{1}{\sigma_1^2} & 0\\ 0 & \frac{1}{\sigma_2^2}\end{pmatrix}\begin{pmatrix}b_1\\ b_2\end{pmatrix}=\frac{b_1}{\sigma_1^2}+\frac{b_2}{\sigma_2^2}.$$

因此

$$\hat{x}=\frac{\frac{b_1}{\sigma_1^2}+\frac{b_2}{\sigma_2^2}}{\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}}.$$

(3) 估计误差方差

加权最小二乘估计的协方差为

$$\mathrm{Var}(\hat{x})=(A^TV^{-1}A)^{-1}=\left(\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}\right)^{-1}.$$

因此题目所求为

$$\mathbb{E}\big[\hat{x}\hat{x}^T\big]=\left(\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}\right)^{-1}.$$

3.6 习题 6

题目：

若随机变量 x 与 y 相互独立，其边缘概率密度分别为 $p_1(x)$ 与 $p_2(y)$，则联合概率密度

$$p(x,y) = p_1(x)p_2(y).$$

通过将二维积分（积分区间均为 $(-\infty,\infty)$）拆分为两个一维积分的乘积，证明：

$$\iint_{-\infty}^{\infty} p(x,y)\,dx\,dy = 1, \qquad \iint_{-\infty}^{\infty} (x+y)\,p(x,y)\,dx\,dy = m_1 + m_2.$$

其中

$$m_1 = \int_{-\infty}^{\infty} x\,p_1(x)\,dx, \qquad m_2 = \int_{-\infty}^{\infty} y\,p_2(y)\,dy.$$

解答：

(1) 证明归一化： 由于 $p(x,y) = p_1(x)p_2(y)$，

$$\iint p(x,y)\,dx\,dy = \iint p_1(x)p_2(y)\,dx\,dy.$$

先对 x 积分、再对 y 积分（或反之），可将其分离为乘积：

$$\iint p_1(x)p_2(y)\,dx\,dy = \left(\int_{-\infty}^{\infty} p_1(x)\,dx\right)\left(\int_{-\infty}^{\infty} p_2(y)\,dy\right).$$

而 p_1, p_2 都是概率密度，因此各自积分为 1：

$$\int_{-\infty}^{\infty} p_1(x)\,dx = 1, \qquad \int_{-\infty}^{\infty} p_2(y)\,dy = 1.$$

故

$$\iint p(x,y)\,dx\,dy = 1 \cdot 1 = 1.$$

(2) 证明期望可加：

$$\iint (x+y)p(x,y)\,dx\,dy = \iint (x+y)\,p_1(x)p_2(y)\,dx\,dy.$$

利用线性性拆分：

$$\iint (x+y)\,p_1(x)p_2(y)\,dx\,dy = \iint x\,p_1(x)p_2(y)\,dx\,dy + \iint y\,p_1(x)p_2(y)\,dx\,dy.$$

对第一项分离变量：

$$\iint x\,p_1(x)p_2(y)\,dx\,dy = \left(\int_{-\infty}^{\infty} x\,p_1(x)\,dx\right)\left(\int_{-\infty}^{\infty} p_2(y)\,dy\right) = m_1\cdot 1 = m_1.$$

对第二项同理：

$$\iint y\,p_1(x)p_2(y)\,dx\,dy = \left(\int_{-\infty}^{\infty} p_1(x)\,dx\right)\left(\int_{-\infty}^{\infty} y\,p_2(y)\,dy\right) = 1\cdot m_2 = m_2.$$

相加得

$$\iint (x+y)p(x,y)\,dx\,dy = m_1 + m_2.$$

证毕。

3.7 习题 7

题目：

继续第 6 题。设随机变量 x, y 相互独立，因此联合概率密度

$$p(x,y) = p_1(x)\,p_2(y).$$

证明

$$\iint (x-m_1)^2\,p(x,y)\,dx\,dy = \sigma_1^2, \qquad \iint (x-m_1)(y-m_2)\,p(x,y)\,dx\,dy = 0.$$

因此 2×2 的协方差矩阵 V 为对角矩阵，其元素为 ________。

解答：

已知独立性给出 $p(x,y)=p_1(x)p_2(y)$，并记

$$m_1=\mathbb{E}[X],\quad m_2=\mathbb{E}[Y],\quad \sigma_1^2=\mathrm{Var}(X),\quad \sigma_2^2=\mathrm{Var}(Y).$$

(1) 证明 $\iint(x-m_1)^2p(x,y)\,dx\,dy=\sigma_1^2$**：**

$$\begin{aligned}\iint(x-m_1)^2p(x,y)\,dx\,dy&=\iint(x-m_1)^2p_1(x)p_2(y)\,dx\,dy\\&=\int(x-m_1)^2p_1(x)\,dx\quad\int p_2(y)\,dy\\&=\int(x-m_1)^2p_1(x)\,dx\\&=\mathrm{Var}(X)=\sigma_1^2,\end{aligned}$$

其中用到 $\int p_2(y)\,dy=1$。
(2) 证明 $\iint(x-m_1)(y-m_2)p(x,y)\,dx\,dy=0$**：**

$$\begin{aligned}\iint(x-m_1)(y-m_2)p(x,y)\,dx\,dy&=\iint(x-m_1)(y-m_2)p_1(x)p_2(y)\,dx\,dy\\&=\left(\int(x-m_1)p_1(x)\,dx\right)\left(\int(y-m_2)p_2(y)\,dy\right)\\&=\big(\mathbb{E}[X-m_1]\big)\big(\mathbb{E}[Y-m_2]\big)\\&=(\mathbb{E}[X]-m_1)(\mathbb{E}[Y]-m_2)=0.\end{aligned}$$

因此协方差矩阵

$$V=\begin{pmatrix}\mathrm{Var}(X)&\mathrm{Cov}(X,Y)\\\mathrm{Cov}(Y,X)&\mathrm{Var}(Y)\end{pmatrix}=\begin{pmatrix}\sigma_1^2&0\\0&\sigma_2^2\end{pmatrix}.$$

3.8 习题 8

题目：

设二维协方差矩阵

$$V=\begin{pmatrix}\sigma_1^2 & \sigma_{12}\\ \sigma_{12} & \sigma_2^2\end{pmatrix},$$

证明其逆矩阵可写为

$$V^{-1}=\begin{pmatrix}\sigma_1^2 & \sigma_{12}\\ \sigma_{12} & \sigma_2^2\end{pmatrix}^{-1}=\frac{1}{1-\rho^2}\begin{pmatrix}\frac{1}{\sigma_1^2} & -\frac{\rho}{\sigma_1\sigma_2}\\ -\frac{\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2}\end{pmatrix},\quad \rho=\frac{\sigma_{12}}{\sigma_1\sigma_2}.$$

并指出该形式会出现在二维高斯分布指数项

$$-(x-m)^\top V^{-1}(x-m)$$

中。

解答：

对任意 2×2 矩阵

$$\begin{pmatrix}a & b\\ b & d\end{pmatrix},$$

若行列式 $ad-b^2\neq 0$，则其逆为

$$\begin{pmatrix}a & b\\ b & d\end{pmatrix}^{-1}=\frac{1}{ad-b^2}\begin{pmatrix}d & -b\\ -b & a\end{pmatrix}.$$

令 $a=\sigma_1^2$, $b=\sigma_{12}$, $d=\sigma_2^2$，得到

$$V^{-1}=\frac{1}{\sigma_1^2\sigma_2^2-\sigma_{12}^2}\begin{pmatrix}\sigma_2^2 & -\sigma_{12}\\ -\sigma_{12} & \sigma_1^2\end{pmatrix}.$$

又由相关系数定义 $\rho=\dfrac{\sigma_{12}}{\sigma_1\sigma_2}$，可写

$$\sigma_{12}=\rho\,\sigma_1\sigma_2,\qquad \sigma_1^2\sigma_2^2-\sigma_{12}^2=\sigma_1^2\sigma_2^2\big(1-\rho^2\big).$$

代回并将分母因子拆开:

$$V^{-1} = \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)} \begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix}.$$

将 $\sigma_1^2\sigma_2^2$ 分配进矩阵元素（逐项相除）:

$$V^{-1} = \frac{1}{1-\rho^2} \begin{pmatrix} \frac{\sigma_2^2}{\sigma_1^2\sigma_2^2} & -\frac{\rho\sigma_1\sigma_2}{\sigma_1^2\sigma_2^2} \\ -\frac{\rho\sigma_1\sigma_2}{\sigma_1^2\sigma_2^2} & \frac{\sigma_1^2}{\sigma_1^2\sigma_2^2} \end{pmatrix} = \frac{1}{1-\rho^2} \begin{pmatrix} \frac{1}{\sigma_1^2} & -\frac{\rho}{\sigma_1\sigma_2} \\ -\frac{\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{pmatrix}.$$

证毕。

4 编程实验

4.1 实验：加权最小二乘

4.1.1 实验目的

本实验为加权最小二乘法和普通最小二乘法的对比:

1. 构造带异方差噪声的线性回归模型，噪声方差随样本变化;
2. 对比普通最小二乘与加权最小二乘的参数估计差异;

4.1.2 理论模型

假设带异方差噪声的一元线性模型:

$$y = ax + b + \varepsilon, \qquad \mathbb{E}[\varepsilon_i] = 0, \qquad \mathrm{Var}(\varepsilon_i) = \sigma_i^2.$$

将其写成矩阵形式。令

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}, \quad A = \begin{bmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_N & 1 \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} a \\ b \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{bmatrix},$$

则

$$\boldsymbol{y} = A\boldsymbol{\beta} + \boldsymbol{\varepsilon}.$$

普通最小二乘： 通过最小化平方残差 $\|\boldsymbol{y} - A\boldsymbol{\beta}\|_2^2$ 得到解

$$\hat{\boldsymbol{\beta}}_1 = (A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}\boldsymbol{y}.$$

加权最小二乘： 若噪声协方差矩阵为

$$V = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_N^2),$$

则权重矩阵取

$$W = V^{-1} = \mathrm{diag}\left(\frac{1}{\sigma_1^2}, \ldots, \frac{1}{\sigma_N^2}\right),$$

最小化加权残差二次型 $(\boldsymbol{y} - A\boldsymbol{\beta})^{\mathrm{T}}W(\boldsymbol{y} - A\boldsymbol{\beta})$，其解为

$$\hat{\boldsymbol{\beta}}_2 = (A^{\mathrm{T}}WA)^{-1}A^{\mathrm{T}}W\boldsymbol{y}.$$

4.1.3 数据构造

1. 生成自变量 x_i：在区间 $[x_{\min}, x_{\max}]$ 内均匀采样；
2. 设定真实参数 a, b，构造真实直线 $y_i^\star = ax_i + b$；
3. 令噪声标准差随 x 变化：

   $$\sigma(x) = \sigma_0\big(1 + k|x|\big),$$

   从而 $\sigma_i = \sigma(x_i)$；
4. 对每个样本生成噪声 $\varepsilon_i \sim \mathcal{N}(0, \sigma_i^2)$，并得到观测值

   $$y_i = y_i^\star + \varepsilon_i = ax_i + b + \varepsilon_i.$$

5. 参数：
   - $N = 80$，$a = 2.0$，$b = -1.0$；
   - $x_{\min} = -3.0$，$x_{\max} = 3.0$；
   - $\sigma_0 = 0.06$，$k = 20$。

4.1.4 实验步骤

1. 生成带异方差噪声的数据集 $\{(x_i, y_i, \sigma_i)\}_{i=1}^{N}$；
2. 分别用两种方法估计参数，得到

$$\hat{\boldsymbol{\beta}}_1 = \begin{bmatrix} \hat{a}_1 \\ \hat{b}_1 \end{bmatrix}, \qquad \hat{\boldsymbol{\beta}}_2 = \begin{bmatrix} \hat{a}_2 \\ \hat{b}_2 \end{bmatrix};$$

3. 计算并对比参数误差：

$$|\hat{a}_1 - a|,\ |\hat{b}_1 - b|, \qquad |\hat{a}_2 - a|,\ |\hat{b}_2 - b|;$$

4. 计算预测误差（相对于无噪声真值直线 $y^\star = ax + b$）的均方误差：

$$\mathrm{MSE}(\hat{\boldsymbol{\beta}}) = \frac{1}{N} \sum_{i=1}^{N} \big(\hat{a}x_i + \hat{b} - (ax_i + b)\big)^2. \tag{11}$$

4.1.5 可视化结果与解释

图 1 为异方差数据散点图。可以观察到，当 $|x|$ 较大时，样本点围绕真实直线的离散程度明显增大；而在 $|x|$ 较小的区域，数据点更加集中，对应较小的噪声方差。

图 2 展示了权重 $w_i = 1/\sigma_i^2$ 随 x_i 的变化关系。可以看到，在噪声方差较小的区域，权重更大；在噪声方差较大的区域，权重显著减小。这表明加权最小二乘会在拟合过程中更信任噪声较小的样本点，从而降低高噪声样本对参数估计的干扰。

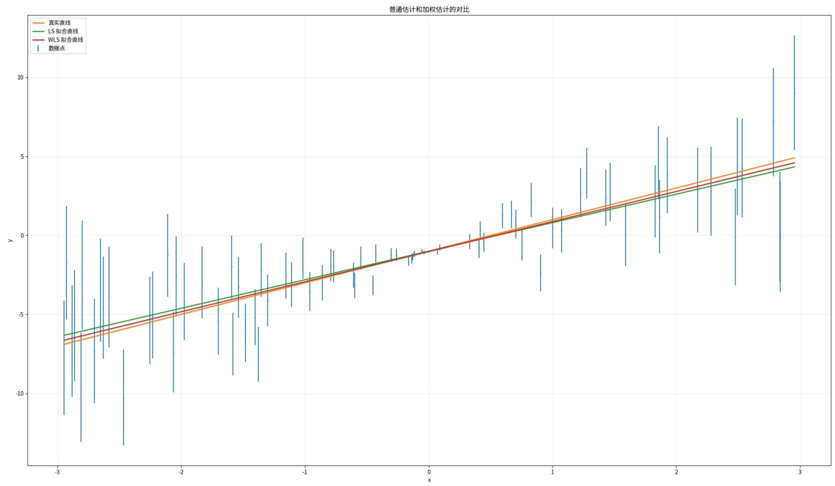

图 1: 数据分布

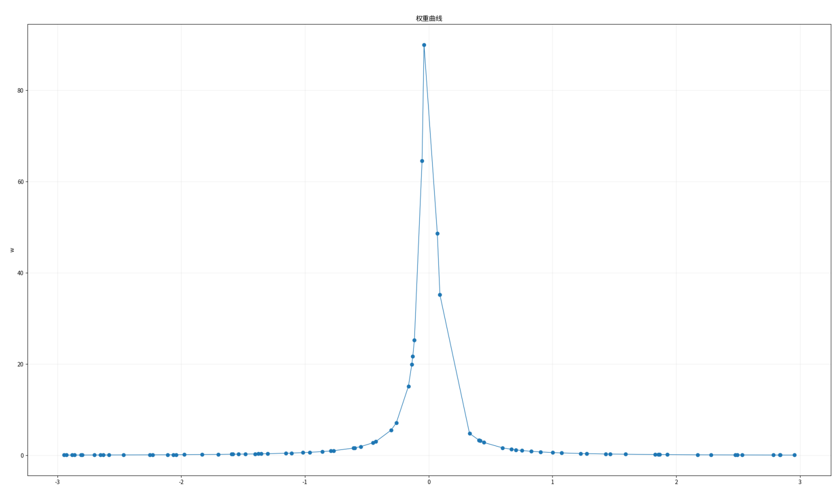

图 2: 权重曲线

代码实现 实验中数据生成与可视化的完整实现见 Listing 1。

Listing 1: main.py

```
import numpy as np
import matplotlib
import matplotlib.pyplot as plt

plt.rcParams['font.sans-serif'] = ['Noto Sans CJK JP']
plt.rcParams['axes.unicode_minus'] = False

def gen_data(N = 200, a = 2.0, b = -1.0, lf = -3.0, ri =
    3.0, sigma_0 = 0.2, k = 0.25):
    rng = np.random.default_rng(114514)
    x = rng.uniform(lf, ri, size = N)
    x = np.sort(x)
    # print(x)
    sigma = sigma_0 * (1.0 + k * abs(x))
    eps = rng.normal(loc = 0.0, scale = sigma, size = N)
    y = a * x + b + eps
    return x, y, sigma

def least_squares(x, y):
    A = np.column_stack([x, np.ones(len(x))])
    # print(A)
    sol = np.linalg.solve(A.T @ A, A.T @ y)
    return sol

def weighted_least_squares(x, y, sigma):
    A = np.column_stack([x, np.ones(len(x))])
    inv_v = 1.0 / (sigma ** 2)
    W = np.diag(inv_v)
```

```
    # print(W)
    L = A.T @ W @ A
    R = A.T @ W @ y
    sol = np.linalg.solve(L, R)
    return sol

def calc_MSE(x, xx, a, b):
    y = a * x + b
    y_hat = xx[0] * x + xx[1]
    return np.mean((y_hat - y) ** 2)

if __name__ == "__main__":
    a_true = 2.0; b_true = -1.0
    # print(a_true, b_true)
    x, y, sigma = gen_data()
    sol1 = least_squares(x, y)
    sol2 = weighted_least_squares(x, y, sigma)
    # print(sol1)
    # print(sol2)
    print("真实参数 (a, b) = ", a_true, b_true)
    print("普通最小二乘估计 (a_1, b_1) = ", sol1[0], sol1[1])
    print("加权最小二乘估计 (a_2, b_2) = ", sol2[0], sol2[1])

    print("\na 误差: 普通:", abs(sol1[0] - a_true), " 加权:", abs(sol2[0] - a_true))
    print("b 误差: 普通:", abs(sol1[1] - b_true), " 加权:", abs(sol2[1] - b_true))

    print("\n普通 MSE = ", calc_MSE(x, sol1, a_true, b_true
```

```
))
    print("加权 MSE = ", calc_MSE(x, sol2, a_true, b_true))

    # pic1
    plt.figure()
    plt.errorbar(x, y, yerr = sigma, fmt = 'o', markersize = 3, capsize = 2, label = "数据点")
    xx = np.linspace(x.min(), x.max(), 200)
    y_true = a_true * xx + b_true
    y_1 = sol1[0] * xx + sol1[1]
    y_2 = sol2[0] * xx + sol2[1]
    plt.plot(xx, y_true, linewidth = 2, label = "真实直线")
    plt.plot(xx, y_1, linewidth = 2, label = "LS 拟合直线")
    plt.plot(xx, y_2, linewidth = 2, label = "WLS 拟合直线")
    plt.title("普通估计和加权估计的对比")
    plt.xlabel("x")
    plt.ylabel("y")
    plt.legend()
    plt.grid(True, alpha = 0.2)
    # plt.show()

    # pic 2
    plt.figure()
    w = 1.0 / (sigma ** 2)
    plt.plot(x, w, marker = 'o', linewidth = 1)
    plt.title("权重曲线")
    plt.xlabel("x")
    plt.ylabel("w")
    plt.grid(True, alpha = 0.2)
    plt.show()
```

5 学习心得

本次学习感受最深的是协方差矩阵的分解，它把概率统计中的“独立”和线性代数中的正交对角化关联起来，是个非常有用的方法。

其次蒙特卡洛方法让我初步认识到了在科学计算中，可以从计算结构上改进，用便宜的近似模型提供主要贡献，再用少量昂贵模型修正偏差。

最后就是加权最小二乘法，让我学到了通过噪声方差的大小来加权估计，方差越小的数据越可靠，权重应该越大，这样可以得到更好的估计。