

矩阵之美——算法篇

2025 年 10 月 9 日

目 录

第 1 章 最小二乘法	1
1.1 问题背景	1
1.2 线性方程组的两个图像	2
1.2.1 线性方程组的行空间图像	3
1.2.2 线性方程组的列空间图像	4
1.3 线性方程组的最小二乘解	5
1.3.1 最小二乘法的行空间方法	5
1.3.2 最小二乘法的列空间方法	6
1.3.3 直线拟合	7
1.4 最小二乘法的几何解释	9
1.5 最小二乘法的概率解释	10
1.6 最小二乘法在应用中的问题	13
1.6.1 变量问题	13
1.6.2 约束问题	17
1.6.3 病态问题	18
1.6.4 异常问题	20
1.6.5 目标函数问题	21
1.7 小 结	23
第 2 章 主成分分析	25
2.1 问题背景	25
2.2 基本统计概念	27
2.2.1 随机变量的数字特征	27
2.2.2 样本统计量	29
2.2.3 样本统计量的向量表示	31
2.3 主成分分析的基本原理	33
2.3.1 任意方向的方差	34
2.3.2 模型与求解	34
2.4 主成分分析的几何解释	38

2.5 主成分分析的子空间逼近解释	43
2.6 主成分分析的概率解释	46
2.7 主成分分析的信息论解释	46
2.8 主成分分析在应用中的问题	47
2.8.1 非高斯问题	48
2.8.2 量纲问题	48
2.8.3 维数问题	49
2.8.4 噪声问题	50
2.9 小 结	50
第 3 章 主偏度分析	51
3.1 问题背景	51
3.2 基本概念	52
3.2.1 偏度的定义	52
3.2.2 数据白化	53
3.2.3 张量基本运算	54
3.2.4 统计量映射图	56
3.3 主偏度分析	58
3.3.1 任意方向的偏度	58
3.3.2 协偏度张量的计算	59
3.3.3 模型与求解	61
3.4 非正交约束主偏度分析	64
3.4.1 克罗内克积	64
3.4.2 非正交约束	65
3.5 主偏度分析与独立成分分析	71
3.5.1 快速独立成分分析	71
3.5.2 FastICA 与主偏度分析	73
3.6 主偏度分析的几何解释	74
3.6.1 单形体的偏度映射图	75
3.6.2 几何解释	77
3.7 主偏度分析在应用中的问题	78
3.7.1 收敛问题	79
3.7.2 噪声问题	81

3.7.3 精确解问题	81
3.8 小 结	82
第 4 章 典型相关分析	83
4.1 问题背景	83
4.2 互相关分析	84
4.2.1 模型与求解	84
4.2.2 存在的问题	88
4.3 典型相关分析	90
4.4 典型相关分析与互相关分析	95
4.5 典型相关分析的几何解释	96
4.5.1 幂法	97
4.5.2 几何解释	98
4.6 典型相关分析的变形	100
4.6.1 多视图典型相关分析	100
4.6.2 二维典型相关分析	103
4.7 典型相关分析在应用中的问题	105
4.7.1 病态问题	105
4.7.2 失配问题	106
4.7.3 目标函数和优化模型问题	106
4.8 小 结	107
第 5 章 非负矩阵分解	109
5.1 问题背景	109
5.2 非负矩阵分解的基本原理	110
5.2.1 问题描述	110
5.2.2 问题求解	111
5.3 非负矩阵分解的概率解释	114
5.3.1 高斯分布情形	114
5.3.2 泊松分布情形	115
5.4 非负矩阵分解的物理解释	116
5.5 非负矩阵分解与奇异值分解	118
5.6 非负矩阵分解与 K-means	120

5.7 非负矩阵分解与 KKT 条件	121
5.8 非负矩阵分解在应用中的问题	122
5.8.1 目标函数的凸凹性	122
5.8.2 局部极值问题	124
5.8.3 分母零值问题	125
5.8.4 观测数据负值问题	125
5.9 小 结	126
第 6 章 局部线性嵌入	129
6.1 问题背景	129
6.2 基本概念	130
6.3 局部线性嵌入	135
6.4 拉普拉斯映射	139
6.5 随机邻域嵌入	141
6.6 多维尺度变换	143
6.7 等距特征映射	144
6.8 局部线性嵌入在应用中的问题	148
6.8.1 病态问题	148
6.8.2 改进局部线性嵌入	149
6.8.3 黑塞局部线性嵌入	150
6.9 小 结	152
第 7 章 傅里叶变换	155
7.1 问题背景	155
7.2 傅里叶级数	156
7.3 连续傅里叶变换	159
7.3.1 从傅里叶级数到傅里叶变换	159
7.3.2 傅里叶变换的性质	165
7.4 离散傅里叶变换	168
7.5 快速傅里叶变换	172
7.6 离散傅里叶变换与循环移位矩阵	176
7.6.1 循环移位矩阵特征分解及频域解释	177
7.6.2 循环移位矩阵的时域解释	179

7.7 离散傅里叶变换与完美差分矩阵	184
7.8 离散傅里叶变换与离散余弦变换	190
7.9 傅里叶变换的物理解释	194
7.10 傅里叶变换在应用中的问题	194
7.10.1 频谱分辨率问题	195
7.10.2 频谱泄漏问题	196
7.10.3 时变信号问题	197
7.10.4 分数傅里叶变换	200
7.11 小 结	201
第 8 章 连通中心演化	203
8.1 问题背景	203
8.2 基于 K-means 的中心确定算法	204
8.3 图论的基本概念	205
8.3.1 图的基本术语	205
8.3.2 图的存储结构	208
8.4 连通中心演化	209
8.4.1 动机与理论依据	209
8.4.2 相关概念	213
8.4.3 算法具体步骤	215
8.5 基于特征分解的快速连通中心演化算法	218
8.5.1 算法的计算复杂度	219
8.5.2 时间复杂度的降低	219
8.5.3 空间复杂度的降低	225
8.6 连通中心演化在应用中的问题	226
8.6.1 相似度矩阵构建问题	226
8.6.2 中心数跳变问题	227
8.6.3 样本量失衡问题	228
8.6.4 相似度矩阵的负值问题	230
8.6.5 中心位置局限问题	232
8.7 小 结	233
第 9 章 瑞利商	235

9.1 问题背景	235
9.2 瑞利商的定义与性质	236
9.3 瑞利商的取值范围	237
9.3.1 特征分析法	238
9.3.2 线性规划法	238
9.3.3 广义瑞利商的取值范围	239
9.4 瑞利商的应用	240
9.4.1 主成分分析	240
9.4.2 最小化噪声分量变换	241
9.4.3 典型相关分析	241
9.4.4 线性判别分析	242
9.4.5 局部线性嵌入	242
9.4.6 法曲率	243
9.4.7 自然频率估计	244
9.4.8 谱聚类	245
9.4.9 约束能量最小化	247
9.4.10 正交子空间投影	247
9.5 小 结	248
参考文献	249
附录 A 向量范数与矩阵范数	253
A.1 向量范数	253
A.2 矩阵范数	256
附录 B 矩阵微积分	259
B.1 实值标量函数相对于实向量的梯度	259
B.2 实值向量函数相对于实向量的梯度	260
B.3 实值函数相对于实矩阵的梯度	262
B.4 矩阵微分	263
B.5 迹函数的梯度矩阵	265
B.6 行列式的梯度矩阵	266
B.7 黑塞矩阵	268

第 1 章 最小二乘法

最小二乘法 (Least Squares, LS) 自诞生以来, 在诸多领域已经得到了广泛的应用. 本章将首先分别从行空间和列空间的角度给出线性方程组的两个图像, 然后从代数、几何、概率等角度对最小二乘法给出全方位的诠释.

1.1 问题背景

1766 年, 德国有一位名叫约翰·提丢斯 (Johann Daniel Titius, 1729-1796) 的大学教授, 写了下面的数列

$$\frac{3 \times 2^n + 4}{10}, n = -\infty, 0, 1, 2, 3, \dots$$

令人惊奇的是, 他发现这个数列的每一项与当时已知的六大行星 (水星、金星、地球、火星、木星、土星) 到太阳的距离比例 (地球到太阳的距离定义为 1) 有一定的联系. 提丢斯的朋友, 天文学家波得 (Johann Elert Bode, 1747-1826) 深知这一发现的重要意义, 就于 1772 年公布了提丢斯的这一发现. 这串数从此引起了众多科学家的极大重视, 并被称为提丢斯——波得定律 (即太阳系行星与太阳平均距离的经验规则). 当时, 人们还没有发现天王星和海王星, 认为土星就是距太阳最远的行星. 1781 年, 英籍德国人赫歇尔 (William Herschel, 1738-1822) 在接近 19.6 的位置上 (即数列中的第 8 项) 发现了天王星, 从此, 人们就对这一定则深信不疑了. 根据这一定则, 在数列的第 5 项即 2.8 的位置上也应该对应一颗行星, 只是还没有被发现. 于是, 许多天文学家和天文爱好者便以极大的热情, 踏上了寻找这颗新行星的征程.



图 1.1 六大行星到太阳的距离 (近似) 比例, 其中假定地球到太阳的距离为 1

1801 年, 意大利天文学家皮亚齐 (Giuseppe Piazzi, 1746-1826) 终于在相应位置发现了一颗小行星 (即谷神星). 遗憾的是, 经过 40 天的跟踪观测后, 由于谷神星运行至太阳背后, 皮亚齐失去了谷神星的位置. 随后全世界的科学家利用皮亚齐的观测数据开始寻找谷神星, 但是根据大多数人计算的结果来寻找谷神星都没有收获. 时年 24 岁的高斯 (Johann Carl Friedrich Gauss, 1777-1855) 也采用了一种新方法 (即为最小二乘法) 计算了谷神星的轨道. 奥地利天文学家奥尔伯斯 (Heinrich Wilhelm Matthias Olbers, 1758-1840) 根据高斯计算出来的轨道重新发现了谷神星. 高斯使用的最小二乘

法于 1809 年发表在他的著作《天体运动论》中，而法国科学家勒让德（Adrien-Marie Legendre, 1752-1833）于 1806 年独立发现了最小二乘法，但因不为时人所知而默默无闻。两人曾为谁最早创立最小二乘法原理发生争执。最小二乘法自创立以来，在自然科学乃至社会科学的各个领域得到了广泛的应用。

由于最小二乘法大多用于线性方程组的求解，因此，接下来我们首先给出线性方程组的两个图像。

1.2 线性方程组的两个图像

对于下面包含 m 个方程、 n 个未知量的线性方程组

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases}, \quad (1.1)$$

可以将其写成矩阵（向量）形式

$$\mathbf{A}\mathbf{x} = \mathbf{b},$$

其中

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix},$$

为线性方程组的系数矩阵，

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix},$$

分别为待求解的变量组成的向量和方程组右边的常数项组成的向量。

对于线性方程组 (1.1)，一般可以从两个角度来理解：行空间角度和列空间角度。从行空间角度，该方程组可以认为是 n 维空间中 m 个超平面的交集。从列空间角度，该方程组的常向量 $\mathbf{b}$ 可以认为是系数矩阵 $\mathbf{A}$ 的列向量的线性组合，其中 $\mathbf{x}$ 的分量为组合系数。

1.2.1 线性方程组的行空间图像

(1.1) 中的线性方程组由 m 个线性方程构成, 其中每一个方程均代表 n 维空间中的一个超平面. 以其中的第 i 个方程为例, 即

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n = b_i.$$

显然, 该方程是 n 维空间中以系数矩阵 $\mathbf{A}$ 的第 i 个行向量 $[a_{i1} \ a_{i2} \ \cdots \ a_{in}]$ 为法方向的 $(n-1)$ 维超平面. 而线性方程组 (1.1) 的解则可以认为是这 m 个 $(n-1)$ 维超平面的交集, 此即为线性方程组的行空间图像. 其中, 由系数矩阵 $\mathbf{A}$ 的所有行向量张成的线性空间称为 $\mathbf{A}$ 的行空间.

下面以一个简单的线性方程组为例给出线性方程组行空间图像的直观理解, 即

$$\begin{cases} 2x_1 + x_2 = 3 \\ x_1 + 2x_2 = 3 \end{cases} \quad (1.2)$$

显然, 该方程组由二维平面上的两条直线方程构成 (如图 1.2), 即直线方程 $2x_1 + x_2 = 3$ 和直线方程 $x_1 + 2x_2 = 3$. 两条直线的法线分别为系数矩阵的两个行向量. 这两条直线的交点 (1,1) 即为该线性方程组的唯一解.

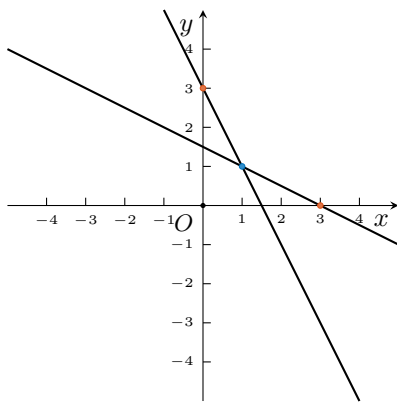


图 1.2 线性方程组的行空间图像

在方程组中的未知数和方程的个数都非常少的时候, 利用方程组的行空间图像来理解线性方程组或许是一种不错的选择 (如图 1.2). 但当方程组的未知数和方程组的个数比较多时, 行空间的方式将很难得到方程组的清晰图像. 比如当 $m = n = 3$ 时, 线性方程组 (1.1) 的解对应于三维空间中三个二维平面的交集. 尽管我们仍然可以通过类似于图 1.2 的方式分别画出三个平面, 然后再探讨它们的交集情况. 但是, 与平面上两条直线的相交问题相比, 三维空间中三个平面的相交情况无论从绘图, 还是从交集结构角度, 都明显要复杂一些. 进一步地, 当方程组的变量的个数和方程的个数更多时, 利用行空间图像的方式来理解线性方程组, 将更为繁琐, 甚至不可接受.

从行空间图像的角度来理解线性方程组,可以得到一些有悖于我们常识的有趣的结论. 比如, 当 $m = n = 4$ 时, (1.1) 变为

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + a_{14}x_4 = b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + a_{24}x_4 = b_2 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + a_{34}x_4 = b_3 \\ a_{41}x_1 + a_{42}x_2 + a_{43}x_3 + a_{44}x_4 = b_4 \end{cases}, \quad (1.3)$$

该方程组包含 4 个变量和 4 个方程. 其中每个方程对应着四维空间的一个三维超平面. 不妨假设该方程组的系数矩阵 $\mathbf{A}$ 非奇异. 从行空间图像角度, (1.3) 的解对应 4 个三维超平面的交集. 如果我们联立前两个方程, 则可以得到四维空间中的一个二维平面 S_1 . 相应地, 后两个方程联立也同样可以得到四维空间中的另一个二维平面 S_2 . 因此, 从行空间图像角度, (1.3) 的解也可以认为对应着这两个二维平面的交集. 由系数矩阵的非奇异性, 我们知道该方程组只有唯一解 $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$. 这意味着, S_1 与 S_2 相交且只交于一点. 因此, 在四维空间中, 两个平面可以只相交于一个点.

从上面两个平面交于一个点的例子可以看出, 在三维空间中不可能出现的事情, 在四维空间中却真实地发生了. 因此, 在大多数情况下, 尽管行空间图像并不是一个处理线性方程组的好的方式, 但它却有助于增进我们对高维世界的理解.

1.2.2 线性方程组的列空间图像

记 $\mathbf{A} = [\mathbf{a}_1 \quad \mathbf{a}_2 \quad \cdots \quad \mathbf{a}_n]$, 则线性方程组 (1.1) 可以表示为

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n = \mathbf{b}, \quad (1.4)$$

即方程组 (1.1) 右边的常数向量 $\mathbf{b}$ 可以表示为系数矩阵 $\mathbf{A}$ 的列向量的线性组合, 且组合系数为待求的未知变量. 此即为线性方程组的列空间图像, 其中由系数矩阵 $\mathbf{A}$ 的所有列向量张成的线性空间称为 $\mathbf{A}$ 的列空间.

从列空间图像的角度, (1.2) 可以重新表示为

$$x_1 \begin{bmatrix} 2 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \end{bmatrix}. \quad (1.5)$$

在 (1.5) 中, 方程组的常数项向量表示为系数矩阵的两个列向量的线性组合, 且组合系数分别为两个待求的变量 x_1, x_2 . 显然 $x_1 = x_2 = 1$ 即为满足 (1.5) 的解. 从图 1.3 可以看出, 我们也可以将 $\mathbf{x} = \begin{bmatrix} x_1 & x_2 \end{bmatrix}^T$ 看作 $\mathbf{b}$ 在 $\eta = \{(2, 1), (1, 2)\}$ 坐标系下的坐标.

接下来, 我们分别从行空间图像角度和列空间图像角度探讨一下线性方程组的解的存在性. 从行空间图像角度, 当线性方程组的各个方程所对应的超平面的交集非空

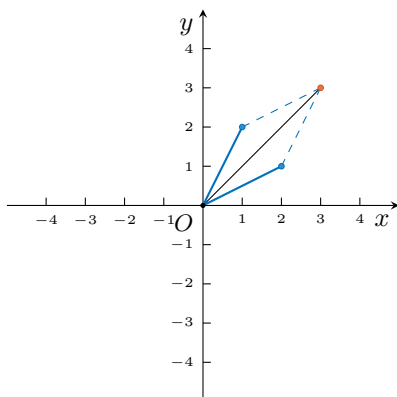


图 1.3 线性方程组的列空间图像

的时候, 线性方程组有解; 反之, 则无解. 从列空间图像角度, 当线性方程组右边的常数项向量位于系数矩阵的列空间时, 方程有解; 否则, 则无解. 从中可以看出, 相对于线性方程组的行空间图像, 线性方程组的列空间图像更为简洁、清晰.

对于一个线性方程组, 当常数项向量位于系数矩阵的列空间时, 可以很容易得到方程组的精确解. 而当常数项向量不能由系数矩阵的列向量线性表出时, 我们可以引入最小二乘法来得到方程组的近似解.

1.3 线性方程组的最小二乘解

由线性方程组的求解理论可知, 在求解线性方程组时, 当方程的个数多于未知数(变量)的个数时, 方程往往无解, 此类方程组称为矛盾方程组或超定方程组. 最小二乘法是求解矛盾方程组的经典方法. 下面分别从行空间图像角度和列空间角度给出最小二乘法的两种解法.

1.3.1 最小二乘法的行空间方法

不妨假设线性方程组 (1.1) 是矛盾方程组. 从行空间图像的角度, 即不存在一组系数 $x_1, x_2, \dots, x_n$, 使得方程组 (1.1) 中的每一个方程都成立. 那么, 退而求其次, 我们希望存在一组系数 $x_1, x_2, \dots, x_n$, 使得方程组的每一个方程都尽量成立, 即对于每一个方程, 我们希望 $a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n = b_i, i = 1, 2, \dots, m$ 尽量成立. 因此, 自然而然, 我们可以用 $\left(\sum_{j=1}^n a_{ij}x_j - b_i\right)^2$ 来衡量第 i 个方程在这组系数下的误差. 此时, 针对线性方程组, 这组系数的求解问题就可以转化为如下优化模型

$$\min_{x_1, x_2, \dots, x_n} f(x_1, x_2, \dots, x_n) = \sum_{i=1}^m \left(\sum_{j=1}^n a_{ij} x_j - b_i \right)^2. \quad (1.6)$$

为了得到上述优化模型的最优解, 我们可以首先求解目标函数的稳定点, 即满足目标函数相对于每个自变量的偏导数都为 0 的点. 由于

$$\frac{\partial f(x_1, x_2, \dots, x_n)}{\partial x_k} = 2 \sum_{i=1}^m a_{ik} \left(\sum_{j=1}^n a_{ij} x_j - b_i \right) \stackrel{\triangle}{=} 0, k = 1, 2, \dots, n,$$

因此, 目标函数稳定点的求解可以转化为如下线性方程组的求解

$$\begin{cases} \sum_{i=1}^m \sum_{j=1}^n a_{i1} a_{ij} x_j = \sum_{i=1}^m a_{i1} b_i \\ \sum_{i=1}^m \sum_{j=1}^n a_{i2} a_{ij} x_j = \sum_{i=1}^m a_{i2} b_i \\ \vdots \\ \sum_{i=1}^m \sum_{j=1}^n a_{in} a_{ij} x_j = \sum_{i=1}^m a_{in} b_i \end{cases}. \quad (1.7)$$

显然 (1.7) 是一个有 n 个未知量 n 个方程的线性方程组, 令

$$\tilde{\mathbf{A}} = \begin{bmatrix} \sum_{i=1}^m a_{i1} a_{i1} & \sum_{i=1}^m a_{i1} a_{i2} & \cdots & \sum_{i=1}^m a_{i1} a_{in} \\ \sum_{i=1}^m a_{i2} a_{i1} & \sum_{i=1}^m a_{i2} a_{i2} & \cdots & \sum_{i=1}^m a_{i2} a_{in} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^m a_{in} a_{i1} & \sum_{i=1}^m a_{in} a_{i2} & \cdots & \sum_{i=1}^m a_{in} a_{in} \end{bmatrix}, \quad \tilde{\mathbf{b}} = \begin{bmatrix} \sum_{i=1}^m a_{i1} b_i \\ \sum_{i=1}^m a_{i2} b_i \\ \vdots \\ \sum_{i=1}^m a_{in} b_i \end{bmatrix},$$

则 (1.7) 的解为

$$\mathbf{x} = \tilde{\mathbf{A}}^{-1} \tilde{\mathbf{b}}. \quad (1.8)$$

尽管 (1.8) 的表达式比较简洁, 但由于 (1.7) 的系数矩阵 $\tilde{\mathbf{A}}$ 过于繁杂, 因此实际应用中我们大多从线性方程组的列空间图像角度给出方程组的最小二乘解.

1.3.2 最小二乘法的列空间方法

当线性方程组 (1.1) 为矛盾方程组时, 从列空间图像的角度, 这意味着常数项向量 $\mathbf{b}$ 不在方程组 (1.1) 的系数矩阵 $\mathbf{A}$ 的列空间中, 或者说 $\mathbf{b}$ 不能由 $\mathbf{A}$ 的列向量线性表出. 同样地, 退而求其次, 我们可以寻求找到一组系数 $x_1, x_2, \dots, x_n$, 使得 $x_1 \mathbf{a}_1 + x_2 \mathbf{a}_2 + \cdots + x_n \mathbf{a}_n = \mathbf{A} \mathbf{x}$ 与向量 $\mathbf{b}$ 尽量接近. 不妨用这两项之差的 2-范数的平

方来衡量它们之间的误差, 这样, 这组系数的求解问题可以转化为如下优化模型

$$\min_{\mathbf{x}} f(\mathbf{x}) = \|\mathbf{Ax} - \mathbf{b}\|^2. \quad (1.9)$$

为了得到 (1.9) 的最优解, 我们同样需要先找到目标函数的稳定点. 首先, 将目标函数展开可以得到

$$f(\mathbf{x}) = \|\mathbf{Ax} - \mathbf{b}\|^2 = (\mathbf{Ax} - \mathbf{b})^T (\mathbf{Ax} - \mathbf{b}) = \mathbf{x}^T \mathbf{A}^T \mathbf{Ax} - 2\mathbf{x}^T \mathbf{A}^T \mathbf{b} + \mathbf{b}^T \mathbf{b}.$$

利用矩阵微积分公式 (相关知识见附录 B) 可以得到

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = 2\mathbf{A}^T \mathbf{Ax} - 2\mathbf{A}^T \mathbf{b}.$$

令 $\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = \mathbf{0}$, 有

$$\mathbf{x} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b} = \mathbf{A}^\dagger \mathbf{b}, \quad (1.10)$$

此即为线性方程组的最小二乘解, 其中 $\mathbf{A}^\dagger = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T$ 称为矩阵 $\mathbf{A}$ 的广义逆.

事实上, 模型 (1.9) 与 (1.6) 是等价的, 所以 (1.9) 可以认为是 (1.6) 的矩阵 (向量) 表达. 注意到

$$\tilde{\mathbf{A}} = \mathbf{A}^T \mathbf{A}, \quad \tilde{\mathbf{b}} = \mathbf{A}^T \mathbf{b}.$$

因此, 两种方式得到的线性方程组的最小二乘解 (1.10) 与 (1.8) 也是等价的. 但毫无疑问, 与基于行空间图像的方法相比, 基于列空间图像的最小二乘解更为简洁、直观、优美.

1.3.3 直线拟合

最小二乘法在诸多领域和方向得到了广泛的应用, 而本节着重介绍最小二乘法在直线拟合中的应用. 已知一组观测点 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ 分布在直角坐标系中 (如图 1.4), 如何用一条直线拟合这些散点呢?

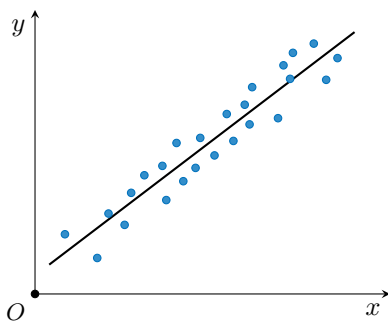


图 1.4 直线拟合示意图

不妨假设待求的直线方程为

$$y = ax + b, \quad (1.11)$$

其中 a 代表直线的斜率, b 为直线的截距, 它们均为待求的未知变量. 显然, 我们希望已知的 n 个观测都满足该直线方程, 即

$$\begin{cases} ax_1 + b = y_1 \\ ax_2 + b = y_2 \\ \vdots \\ ax_n + b = y_n \end{cases}. \quad (1.12)$$

实际应用中, (1.12) 往往为矛盾方程, 即不存在 a, b 使得上述方程组的各个方程都成立. 因而, 退而求其次, 我们接下来转而寻求 a, b , 使得 (1.12) 尽量成立. 借鉴 (1.6), 我们可以建立直线拟合的优化模型为

$$\min_{a,b} f(a,b) = \sum_{i=1}^n (ax_i + b - y_i)^2. \quad (1.13)$$

方便起见, 我们可以类比 (1.9) 将 (1.13) 中的目标函数重新表示为自变量为向量的情形, 这样, (1.13) 可以转化为如下形式

$$\min_{\mathbf{c}} f(\mathbf{c}) = \|\mathbf{X}\mathbf{c} - \mathbf{y}\|^2, \quad (1.14)$$

其中,

$$\mathbf{X} = \begin{bmatrix} \mathbf{x} & \mathbf{1} \end{bmatrix} = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \mathbf{c} = \begin{bmatrix} a \\ b \end{bmatrix}.$$

根据 (1.10), 可得 (1.14) 的最小二乘解为

$$\mathbf{c} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}. \quad (1.15)$$

下面用一个非常简单的例子来直观展示最小二乘法的求解过程.

例 1.1 用直线 $y = ax + b$ 来拟合平面上的三个点: $(1, 1)$, $(2, 1)$ 和 $(3, 3)$.

解 根据给定的三个点的坐标, 显然有 $x_1 = 1, x_2 = 2, x_3 = 3$; $y_1 = 1, y_2 = 1, y_3 = 3$. 记

$$\mathbf{X} = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix},$$

根据公式 (1.15) 可得

$$\begin{aligned} \mathbf{c} &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \\ &= \left(\begin{bmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix} \\ &= \begin{bmatrix} 1 \\ -1/3 \end{bmatrix}, \end{aligned}$$

即, 待求直线方程为 $y = x - \frac{1}{3}$.

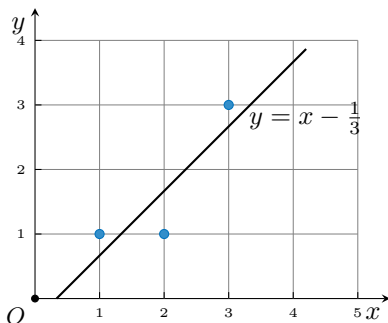


图 1.5 三个点及其对应的拟合直线

1.4 最小二乘法的几何解释

对于方程组 (1.1), (1.4) 给出了其列空间表达, 即 $x_1 \mathbf{a}_1 + x_2 \mathbf{a}_2 + \cdots + x_n \mathbf{a}_n = \mathbf{b}$. 当此方程组为矛盾方程组时, 即当 $\mathbf{b}$ 不能由系数矩阵 $\mathbf{A}$ 的列向量线性表出时, (1.10) 给出的方程组的最小二乘解 $\mathbf{x} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b}$ 即为我们想要的一组系数. 利用这组系数对 $\mathbf{A}$ 的各个列向量进行线性组合, 将得到 $\mathbf{A}$ 的列空间的一个新元素 $\mathbf{Ax}$. 显然, 这个新元素 $\mathbf{Ax}$ 在 $\mathbf{A}$ 的列空间, 而向量 $\mathbf{b}$ 不在 $\mathbf{A}$ 的列空间, 那么它们之间存在什么关系呢?

事实上, 由于 $\mathbf{Ax} = \mathbf{A} (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b}$, 记

$$\mathbf{P}_\mathbf{A} = \mathbf{A} (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T, \quad (1.16)$$

显然 $\mathbf{Ax}$ 与 $\mathbf{b}$ 之间的关系为

$$\mathbf{Ax} = \mathbf{P}_\mathbf{A} \mathbf{b}. \quad (1.17)$$

即, 将 $\mathbf{P}_\mathbf{A}$ 作用于 $\mathbf{b}$, 就能得到 $\mathbf{Ax}$, 那么矩阵 $\mathbf{P}_\mathbf{A}$ 又是一个什么矩阵呢? 它具有什么性质呢? 事实上, $\mathbf{P}_\mathbf{A}$ 就是由矩阵 $\mathbf{A}$ 构建的投影矩阵, 将其作用于任何一个向

量, 都会将该向量投影到矩阵 $\mathbf{A}$ 的列空间. 因此, 当 $\mathbf{x} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b}$ 时, $\mathbf{Ax}$ 就是向量 $\mathbf{b}$ 在 $\mathbf{A}$ 的列空间的投影, 或者说, $\mathbf{Ax}$ 是 $\mathbf{A}$ 的列空间中距离向量 $\mathbf{b}$ 最近的向量. 也即, 当 (1.1) 为矛盾方程组时, 虽然不存在一组系数 $x_1, x_2, \dots, x_n$, 使得 $x_1 \mathbf{a}_1 + x_2 \mathbf{a}_2 + \dots + x_n \mathbf{a}_n = \mathbf{b}$ 成立, 但是方程组的最小二乘解 $\mathbf{x} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b}$, 会使得 $x_1 \mathbf{a}_1 + x_2 \mathbf{a}_2 + \dots + x_n \mathbf{a}_n$ (即 $\mathbf{Ax}$) 最大程度靠近 $\mathbf{b}$.

总结而言, 从几何角度, 最小二乘法相当于将常数项向量 $\mathbf{b}$ 往系数矩阵 $\mathbf{A}$ 的列空间上进行正交投影.

接下来, 我们给出投影矩阵的严格定义.

定义 1.1 (投影矩阵) 具有对称性的幂等矩阵, 即为投影矩阵.

从定义可以看出, 当一个矩阵 $\mathbf{P}$ 为投影矩阵时, 根据对称性, 必然有 $\mathbf{P}^T = \mathbf{P}$. 根据幂等性, 必然有 $\mathbf{P}^2 = \mathbf{P}$. 投影矩阵为幂等矩阵, 意味着, 在一个空间上投影一次和投影两次甚至多次, 结果是一样的.

在实际应用中, 我们还经常用到投影矩阵 $\mathbf{P}$ 的正交补投影矩阵 $\mathbf{P}^\perp$, 其中

$$\mathbf{P}^\perp = \mathbf{I} - \mathbf{P}, \quad (1.18)$$

即正交补投影矩阵等于单位阵与投影矩阵之差. 对于 (1.16) 中的矩阵 $\mathbf{P}_\mathbf{A}$, 容易验证其满足投影矩阵的定义, 即 $\mathbf{P}_\mathbf{A}^T = \mathbf{P}_\mathbf{A}$, $\mathbf{P}_\mathbf{A}^2 = \mathbf{P}_\mathbf{A}$. 相应地 $\mathbf{P}_\mathbf{A}$ 的正交补投影矩阵为 $\mathbf{P}_\mathbf{A}^\perp = \mathbf{I} - \mathbf{P}_\mathbf{A}$, 将 $\mathbf{P}_\mathbf{A}^\perp$ 作用于任意列向量, 则将该向量投影到矩阵 $\mathbf{A}$ 的列空间的正交补空间. 值得注意的是, 上述投影矩阵 $\mathbf{P}_\mathbf{A}$ 的定义一般要求矩阵 $\mathbf{A}$ 为列满秩矩阵. 当 $\mathbf{A}$ 为行满秩矩阵的时候, 可先将其转置为列满秩矩阵 $\mathbf{A}^T$, 相应的投影矩阵为 $\mathbf{P}_{\mathbf{A}^T} = \mathbf{A}^T (\mathbf{A} \mathbf{A}^T)^{-1} \mathbf{A}$, 将 $\mathbf{P}_{\mathbf{A}^T}$ 作用于任意向量, 则可以将该向量投影到矩阵 $\mathbf{A}$ 的行空间.

仍以上述的直线拟合为例. 显然, 这 n 个观测对 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ 可以认为是二维空间的 n 个散点. 同时, 也可以将这些观测看作 n 维样本空间的两个向量 $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_n]^T$, $\mathbf{y} = [y_1 \ y_2 \ \dots \ y_n]^T$. 利用 $y = ax + b$ 对这组散点进行直线拟合, 相当于寻找参数 a, b 使得向量方程 $\mathbf{y} = a\mathbf{x} + b\mathbf{1}$ 成立或者尽量成立. 从样本空间来看, 其实就是找到向量 $\mathbf{x}$ 与 $\mathbf{1}$ 的最佳线性组合, 使得该组合与向量 $\mathbf{y}$ 尽量接近. 这相当于将向量 $\mathbf{y}$ 投影到向量 $\mathbf{x}$ 与 $\mathbf{1}$ 所张成的平面上, 而投影点 $\mathbf{y}'$ 即为该平面上距离点 $\mathbf{y}$ 最近的点 (图 1.6).

1.5 最小二乘法的概率解释

本节将从概率角度, 对最小二乘法给出进一步解释. 仍以平面上一组观测点 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ 的直线拟合为例. 在模型 (1.14) 中, 目标函数 $f(a, b) = \sum_{i=1}^n (ax_i + b - y_i)^2$ 代表因变量 y 的总体观测误差. 注意到, 这个总体观测误差由 n

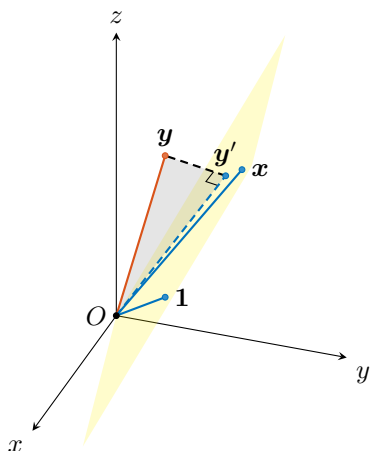


图 1.6 最小二乘法的几何解释, 其中 $\mathbf{x} = [1 \ 2 \ 3]^T$, $\mathbf{y} = [1 \ 1 \ 3]^T$, $\mathbf{1} = [1 \ 1 \ 1]^T$, 最小二乘法对应着三个向量之间的投影关系. 当以 $y = ax + b$ 来拟合散点时, 对应到样本空间, 相当于寻求向量 $\mathbf{y}$ 到向量 $\mathbf{x}$ 和 $\mathbf{1}$ 所张成平面 ($\text{span}(\mathbf{x}, \mathbf{1})$) 上的投影 $\mathbf{y}'$

个部分构成. 其中的每一项 $(ax_i + b - y_i)^2$ 均代表相应观测点的因变量的观测误差. 即对于每个观测点的因变量的观测误差, 都用模型解 $(ax_i + b)$ 与观测值 y_i 的差的平方来衡量. 这正是相应的方法命名为最小二乘法而不是最小一乘或者最小三乘的原因所在. 接下来, 从概率的角度对这一问题进行进一步阐述.

在直线拟合问题中, 由于 (1.12) 一般为矛盾方程, 因此我们可以把 (1.12) 写为如下的形式

$$\begin{cases} y_1 = ax_1 + b + \varepsilon_1 \\ y_2 = ax_2 + b + \varepsilon_2 \\ \vdots \\ y_n = ax_n + b + \varepsilon_n \end{cases}, \quad (1.19)$$

其中, ε_i 为第 i 个因变量 y_i 的观测误差, 它对应着不能被线性模型刻画的因素. 由于误差的不确定性, 其中的每一个观测误差 ε_i 都可以看作一个随机变量. 假设 ε_i 服从均值为 0, 标准差为 σ 的正态分布, 即 $\varepsilon_i \sim N(0, \sigma^2)$, 则其概率密度函数为¹

$$f(\varepsilon_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\varepsilon_i^2}{2\sigma^2}\right). \quad (1.20)$$

¹在本书, 模型误差函数和概率密度函数都用 $f(\cdot)$ 来表示, 请读者注意区分.

这意味着, 在给定 x_i 和参数 a, b 的情况下, 因变量 y_i 也服从同样的正态分布, 即

$$f(y_i | x_i; a, b) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(ax_i + b - y_i)^2}{2\sigma^2}\right). \quad (1.21)$$

我们可以定义所有观测数据关于参数 a, b 的似然 (Likelihood) 函数如下

$$L(a, b) = f(y_1, y_2, \dots, y_n | x_1, x_2, \dots, x_n; a, b). \quad (1.22)$$

显然, 似然函数 $L(a, b)$ 为给定 $x_1, x_2, \dots, x_n$ 和参数 a, b 情况下, $y_1, y_2, \dots, y_n$ 的联合概率密度函数. 不妨假设所有的误差项 ε_i 独立同分布, 那么联合概率密度函数等于所有因变量的概率密度函数的乘积, 即

$$f(y_1, y_2, \dots, y_n | x_1, x_2, \dots, x_n; a, b) = \prod_{i=1}^n f(y_i | x_i; a, b), \quad (1.23)$$

因此,

$$L(a, b) = \prod_{i=1}^n f(y_i | x_i; a, b) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(ax_i + b - y_i)^2}{2\sigma^2}\right). \quad (1.24)$$

选择合适的参数 a, b , 使得观测数据出现的可能性最大 (即 $L(a, b)$ 最大), 即为关于参数 a, b 的最大似然问题. 为了便于求解, 定义 $l(a, b) = \ln(L(a, b))$, 显然最大化 $L(a, b)$ 与最大化 $l(a, b)$ 是等价的. 由于

$$\begin{aligned} l(a, b) &= \ln\left(\prod_{i=1}^n f(y_i | x_i; a, b)\right) = \ln\left(\prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(ax_i + b - y_i)^2}{2\sigma^2}\right)\right) \\ &= \sum_{i=1}^n \ln\left(\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(ax_i + b - y_i)^2}{2\sigma^2}\right)\right) \\ &= n \ln \frac{1}{\sqrt{2\pi}\sigma} - \sum_{i=1}^n \frac{(ax_i + b - y_i)^2}{2\sigma^2}, \end{aligned} \quad (1.25)$$

并且 $\frac{1}{2\sigma^2}$ 为常数. 因此, 最大化 $l(a, b)$, 就相当于最小化 $\sum_{i=1}^n (ax_i + b - y_i)^2$. 而 $\sum_{i=1}^n (ax_i + b - y_i)^2$ 正好为 (1.13) 的目标函数, 即 $f(a, b) = \sum_{i=1}^n (ax_i + b - y_i)^2$. 因此, 在观测误差满足独立同高斯分布的前提下, 直线拟合问题的目标函数中的平方项是一个必然的结果.

从上面的推导可以看出, 各个观测点的模型误差满足独立同分布的高斯分布是最小二乘法能够行之有效的前提. 如果此条件不能得到满足, 用最小二乘法拟合数据很多情况下将不能得到合理的结果.

由于各个观测互不影响, 因此不同观测点的观测误差之间的独立性一般都是成立的, 那么接下来我们重点关注每一个观测点的模型误差是否都满足高斯分布. 事实上, 概率论中的中心极限定理为这个前提条件的成立提供了一定的理论保证.

定理 1.1 (中心极限定理) 假设随机变量 $\delta_1, \delta_2, \dots, \delta_n$ 相互独立, 在一定条件下², 它们的平均 $\frac{1}{n} \sum_{i=1}^n \delta_i$ 随着 n 的增大趋于高斯分布.

对于上述直线拟合问题中的观测数据的每一个误差项 ε_i , 都可以认为它由多种不相干的因素构成, 而每一个因素也可以认为是一个随机变量, 不妨将其记为 $\varepsilon_{ij}, j = 1, 2, \dots, k$, 其中 k 为随机因素的个数. 因此, ε_i 可以表示为多个随机变量的平均, 即 $\varepsilon_i = \frac{1}{k} \sum_{j=1}^k \varepsilon_{ij}$. 而当这些随机因素的个数比较大时, 根据中心极限定理, 他们的平均 ε_i 必然趋于高斯分布. 因此, 一般情况下, 我们假设直线拟合中不能被模型刻画的观测误差满足高斯分布是合理的.

1.6 最小二乘法在应用中的问题

前面的几节给出最小二乘法的相关理论结果, 但在实际应用中, 针对不同的应用场景, 最小二乘法还存在着各种问题. 接下来, 我们将对其中的变量问题、约束问题、病态问题、异常问题、目标函数问题等几个常见的问题展开讨论.

1.6.1 变量问题

在本小节, 我们仍以直线拟合为例, 探讨最小二乘法中的变量问题. 在 1.3.3 节中, 对于给定的一组散点 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, 通过最小二乘法, 我们给出了拟合这组散点的直线方程 $y = ax + b$. 这样做, 其实隐含了一个前提条件: x 为自变量, y 为因变量. 那么我们不免要问, 对于同样的这组散点, 如果把 y 当做自变量, x 当做因变量, 即用 $x = a'y + b'$ 来拟合这组散点是否可以得到同样的结果呢? 或者说, 用 $x = a'y + b'$ 来拟合这组散点, 和用 $y = ax + b$ 来拟合这组散点是否会得到同一条直线呢? 容易验证, 如果它们为同一条直线, 必有下列等式成立

$$\begin{cases} a' = \frac{1}{a} \\ b' = -\frac{b}{a} \end{cases}. \quad (1.26)$$

接下来我们把 y 当做自变量, x 当做因变量, 用 $x = a'y + b'$ 来拟合这组散点. 同样记

$$\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^T, \quad \mathbf{y} = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix}^T, \quad \mathbf{1} = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}^T,$$

这组散点的直线拟合方程 $x = a'y + b'$ 的确定, 相当于寻求系数 a', b' , 使得下面的方

² 在 Lindeberg-Levy 定理中这个条件为随机变量具有相同的分布, 在 Lindeberg-Feller 定理中这个条件为 Lindeberg 条件, 该条件不要求随机变量具有相同的分布.

程组尽量成立

$$\begin{cases} a'y_1 + b' = x_1 \\ a'y_2 + b' = x_2 \\ \vdots \\ a'y_n + b' = x_n \end{cases}. \quad (1.27)$$

其对应的向量形式为

$$a'\mathbf{y} + b'\mathbf{1} = \mathbf{x},$$

令

$$\mathbf{Y} = [\mathbf{y} \quad \mathbf{1}], \quad \mathbf{c}' = [a' \quad b']^T,$$

则最佳系数 a', b' 的确定问题转化为如下优化模型

$$\min_{\mathbf{c}'} g(\mathbf{c}') = \|\mathbf{Y}\mathbf{c}' - \mathbf{x}\|^2. \quad (1.28)$$

借鉴 (1.14) 可以得到 (1.28) 的最小二乘解为

$$\mathbf{c}' = (\mathbf{Y}^T \mathbf{Y})^{-1} \mathbf{Y}^T \mathbf{x}. \quad (1.29)$$

至此, 我们通过自变量和因变量位置的互换给出了两种直线拟合的方式. 那么, 这两种直线拟合的方式是否等价呢? 下面我们仍然用 1.3.3 节的例子 1.1 中的三个散点 $(1, 1), (2, 1), (3, 3)$ 来对此问题进行验证. 记

$$\mathbf{Y} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 3 & 1 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix},$$

根据公式 (1.29), 可以得到

$$\mathbf{c}' = (\mathbf{Y}^T \mathbf{Y})^{-1} \mathbf{Y}^T \mathbf{x} = \begin{bmatrix} 0.75 \\ 0.75 \end{bmatrix},$$

对应的直线方程为 (如图 1.7)

$$x = 0.75y + 0.75,$$

即 $a' = b' = 0.75$, 显然 $a' \neq \frac{1}{a}, b' \neq -\frac{b}{a}$, 即等式 (1.26) 不成立. 这说明, 这两种直线拟合得到的结果是不等价的. 从图 1.7 也可以看出, 利用 $x = a'y + b'$ 得到的拟合直线与利用 $y = ax + b$ 拟合的直线并不是同一条直线.

接下来, 我们从模型的目标函数本身来对上述现象进行进一步解释. 当用 $y = ax + b$ 来拟合散点时, 模型的目标函数 (见 (1.13) 或者 (1.14)) 反映的是各个散点在 Y 轴方向到拟合直线的距离的平方和 (图 1.8a). 而当用 $x = a'y + b'$ 来拟合散点时,

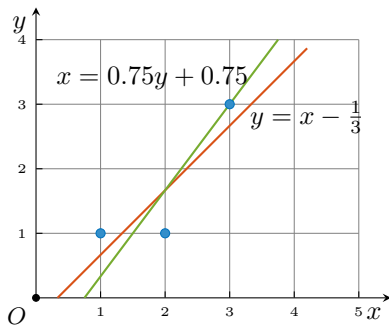


图 1.7 分别用 $y = ax + b$ 和 $x = a'y + b'$ 拟合散点，得到不同的直线方程

模型的目标函数（见 (1.28)）则反映的是各个散点在 X 轴方向到拟合直线的距离的平方和（图 1.8b）。即在 $y = ax + b$ 中， x 为自变量， y 为因变量，此时的最小二乘解把因变量 y 的观测误差降到最小。而在 $x = a'y + b'$ ， y 为自变量， x 为因变量，此时的最小二乘解则使得因变量 x 的观测误差降至最小。总的来说，最小二乘法总是使得因变量的观测误差达到最小。

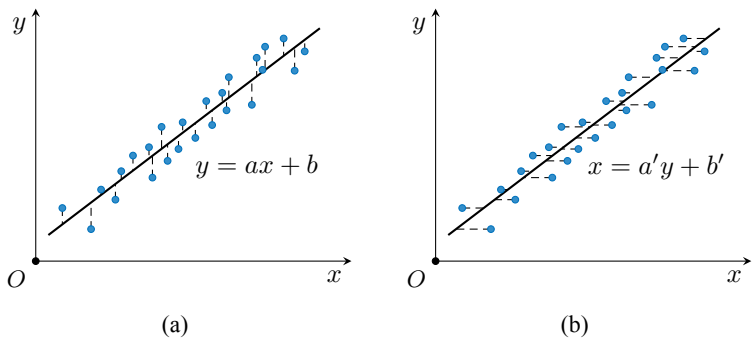


图 1.8 当 x 和 y 互为自变量和因变量时直线拟合的直观解释 (a) x 为自变量 (b) y 为自变量

从上面的讨论可以看出，对于给定的观测数据，选哪个变量作为自变量，哪个变量作为因变量将会对最终的最小二乘结果产生较大的影响。那么，在实际的应用中，该如何界定自变量与因变量呢？一个基本的原则就是，应当选择观测误差小甚至没有观测误差的变量作为自变量；相应地，观测误差大的变量应当被选择为因变量。

真实的数据可能存在所有的观测变量误差都比较大的情况，此时用前面任何一种最小二乘法都可能导致较大的误差。这种情况下，各个变量地位相当，没有明显的因果关系，解决此类问题一般需要引入总体最小二乘法。图 1.9 给出两个变量时总体最小二乘法的直观解释。它最终得到的直线将会使各个散点到该直线距离（即垂直距

离或垂线距离) 的平方和最小. 由于总体最小二乘可以归结为主成分分析, 我们将在第2章对其展开进一步讨论, 这里就不再赘述.

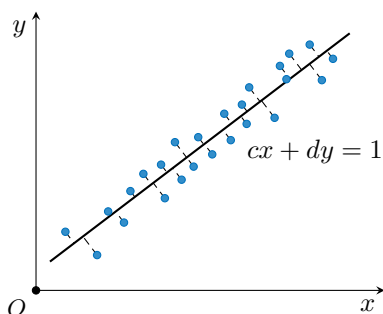


图 1.9 总体最小二乘法的直观解释, 此时变量 x 和 y 都有较大的误差, 且没有明显的因果关系

此外, 从几何角度, 当以 $x = a'y + b'$ 来拟合这组散点时, 在样本空间中相当于寻找向量 $\mathbf{x} = [x_1 \ x_2 \ \cdots \ x_n]^T$ 到向量 $\mathbf{y} = [y_1 \ y_2 \ \cdots \ y_n]^T$ 与 $\mathbf{1} = [1 \ 1 \ \cdots \ 1]^T$ 所张成的平面的投影点 $\mathbf{x}'$ (图 1.10). 显然, 这与图 1.6 所表示的是完全不同的几何关系. 这再次说明了用 $y = ax + b$ 进行直线拟合与用 $x = a'y + b'$ 进行直线拟合一般情况下将得到不同的结果.

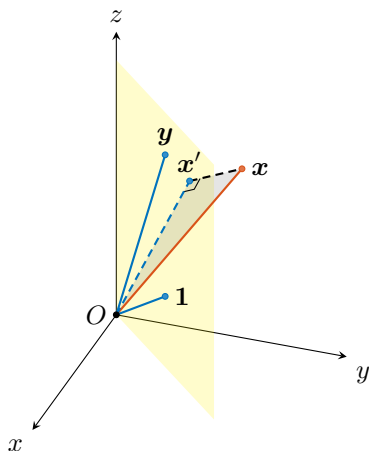


图 1.10 最小二乘法的几何解释: 其中 $\mathbf{x} = [1 \ 2 \ 3]^T$, $\mathbf{y} = [1 \ 1 \ 3]^T$, $\mathbf{1} = [1 \ 1 \ 1]^T$, 最小二乘法对应着三个向量之间的投影关系. 当以 $x = a'y + b'$ 来拟合散点时, 对应到样本空间, 相当于寻求向量 $\mathbf{x}$ 到向量 $\mathbf{y}$ 和 $\mathbf{1}$ 所张成平面 ($\text{span}(\mathbf{y}, \mathbf{1})$) 上的投影 $\mathbf{x}'$

1.6.2 约束问题

上述所有最小二乘法的理论和实例, 均对未知量没有任何约束. 实际应用中, 这些变量可能会有实际的物理意义, 因此需要对它们添加相应的约束以满足相应的物理性质.

仍然假设 (1.1) 为矛盾方程组, 同时假设未知变量 $\mathbf{x}$ 需要满足 m 个等式约束 $h_i(\mathbf{x}) = 0, i = 1, 2, \dots, m$ 和 p 个不等式约束 $g_j(\mathbf{x}) \leq 0, j = 1, 2, \dots, p$, 则相应的最优化模型为

$$\begin{cases} \min_{\mathbf{x}} & f(\mathbf{x}) = \|\mathbf{Ax} - \mathbf{b}\|^2 \\ \text{s.t.} & h_i(\mathbf{x}) = 0, \quad i = 1, 2, \dots, m \\ & g_j(\mathbf{x}) \leq 0, \quad j = 1, 2, \dots, p \end{cases} \quad (1.30)$$

一般情况下, 此模型不存在类似于 (1.10) 的解析解. 在经典的最优化理论里面提供了解决这类问题的诸多方法, 并且很多方法都被收录在诸多应用软件中, 因此这里就不再对此问题展开讨论. 下面用一个简单的例子, 从几何上给约束最小二乘一个直观的解读.

假设系数矩阵 $\mathbf{A}$ 由三维空间中的两个向量组成, 即 $\mathbf{A} = [\mathbf{a}_1 \quad \mathbf{a}_2]$. 为了直观起见, 不妨假设 $\mathbf{a}_1, \mathbf{a}_2$ 都在平面 Oxy 内 (如图 1.11 所示), 并且分别用点 A_1, A_2 表示. 同时假设常数项向量 $\mathbf{b}$ 在 Oyz 平面, 并用点 B 表示.

当没有任何约束项时, 线性方程组的求解问题等价于求解下面的最优化模型,

$$\min_{\mathbf{x}} f(\mathbf{x}) = \|\mathbf{Ax} - \mathbf{b}\|^2. \quad (1.31)$$

正如 (1.10) 所给出的, 此模型存在解析解, 并且此模型的解等价于点 B 往 OA_1, OA_2 所张成的平面的投影, 对应投影点 B_1 .

如果只考虑等式约束, 并假设只有一个等式约束 $\mathbf{1}^T \mathbf{x} = 1$, 其中 $\mathbf{1} = [1 \quad 1]^T$, 即 $\mathbf{x}$ 的两个分量之和为 1. 此时, 线性方程组的求解问题可以转化为如下等式约束的最优化问题

$$\begin{cases} \min_{\mathbf{x}} & f(\mathbf{x}) = \|\mathbf{Ax} - \mathbf{b}\|^2 \\ \text{s.t.} & \mathbf{1}^T \mathbf{x} = 1 \end{cases}. \quad (1.32)$$

此模型也存在解析解 (具体请参考第 6.3 节), 并且此模型的解等价于点 B 往 A_1, A_2 两个点所在直线的投影, 对应投影点 B_2 .

如果只考虑不等式约束, 并假设只有一个不等式约束 $\mathbf{x} \geq \mathbf{0}$, 即 $\mathbf{x}$ 的所有分量都是非负的. 此时, 线性方程组的求解问题可以转化为如下非负约束的最优化问题

$$\begin{cases} \min_{\mathbf{x}} & f(\mathbf{x}) = \|\mathbf{Ax} - \mathbf{b}\|^2 \\ \text{s.t.} & \mathbf{x} \geq \mathbf{0} \end{cases}. \quad (1.33)$$

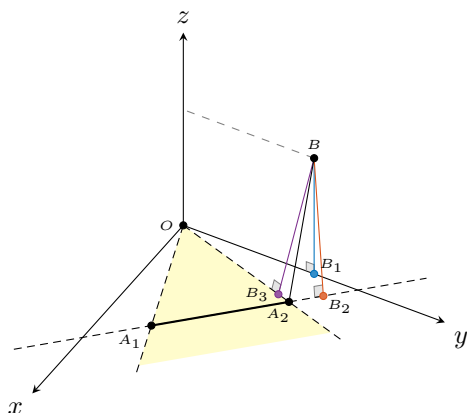


图 1.11 约束最小二乘法: 点 B 的 4 种不同投影方式对应着 4 种不同约束的最小二乘法, 其中 B_1 对应着模型的无约束最小二乘解, B_2 对应着模型的等式约束最小二乘解, B_3 对应着模型的非负约束最小二乘解, $B_4(A_2)$ 对应着模型非负、等式约束最小二乘解

此模型不存在解析解, 但可以通过最优化理论中的罚函数、有效集等方法得到该模型的最优解. 此模型的解等价于点 B 往两个点 A_1, A_2 所对应的向量所夹的黄色区域的投影, 对应投影点 B_3 (此时的投影点可以理解点 B 到黄色区域的最短距离点).

如果同时考虑上述的等式约束和非负约束, 此时, 线性方程组的求解问题可以转化为如下两个约束的最优化问题

$$\begin{cases} \min_{\mathbf{x}} & f(\mathbf{x}) = \|\mathbf{Ax} - \mathbf{b}\|^2 \\ \text{s.t.} & \mathbf{1}^T \mathbf{x} = 1, \mathbf{x} \geq \mathbf{0} \end{cases} \quad (1.34)$$

此模型也不存在解析解, 但是也可以用罚函数、有效集等方法得到该模型的最优解. 此模型的解等价于点 B 往 A_1, A_2 两个点所连接的线段的投影, 对应投影点 B_4 (此时的投影点可以理解点 B 到该线段的最短距离点). 值得注意的是, 此时的投影点 B_4 与点 A_2 是重合的, 因此在图 1.11 中我们并没有标出 B_4 .

1.6.3 病态问题

对于一个线性方程组 $\mathbf{Ax} = \mathbf{b}$, 当其系数矩阵的条件数很大时, 此方程组为病态方程组. 此时, $\mathbf{A}$ 或者 $\mathbf{b}$ 的微小扰动都可能导致方程组的解 $\mathbf{x}$ 发生较大的变动. 比如线性方程组

$$\begin{cases} 5x + 7y = 0.5 \\ 7x + 10y = 0.7 \end{cases}, \quad (1.35)$$

的解为 $x = 0.1, y = 0$. 如果我们对 (1.35) 中右边的常数项 $\mathbf{b}$ 进行微调, 使得线性方程组变为

$$\begin{cases} 5x + 7y = 0.51 \\ 7x + 10y = 0.69 \end{cases},$$

则方程的解变为 $x = 0.27, y = -0.12$. 可见 $\mathbf{b}$ 的微小变动引起了方程的解的较大变动. 之所以会产生这个现象, 正是因为该方程组的系数矩阵具有较大的条件数. 接下来我们首先给出矩阵的条件数的定义.

定义 1.2 (矩阵的条件数) 矩阵 $\mathbf{A}$ 的条件数等于 $\mathbf{A}$ 的范数与 $\mathbf{A}^{-1}$ 的范数的乘积 (矩阵范数相关定义见附录 A), 即 $\text{cond}(\mathbf{A}) = \|\mathbf{A}\| \|\mathbf{A}^{-1}\|$.

可以验证, 当取矩阵的 2-范数时, 矩阵 $\mathbf{A}$ 的条件数等于该矩阵的最大奇异值与最小奇异值之比, 即

$$\text{cond}(\mathbf{A}) = \frac{\sigma_{\max}}{\sigma_{\min}},$$

其中 $\sigma_{\max}$ 为矩阵 $\mathbf{A}$ 的最大奇异值, 而 $\sigma_{\min}$ 为 $\mathbf{A}$ 的最小奇异值.

在 (1.35) 中, 计算可得, 系数矩阵 $\mathbf{A}$ 有两个奇异值, 分别为 $\sigma_1 = 14.933, \sigma_2 = 0.067$. 显然, 矩阵 $\mathbf{A}$ 的两个奇异值相差悬殊, 这必然导致 $\mathbf{A}$ 具有较大的条件数

$$\text{cond}(\mathbf{A}) = \frac{\sigma_1}{\sigma_2} = 222.9955.$$

相应地, 线性方程组 (1.35) 为病态方程组.

为了缓解线性方程组中系数矩阵的病态问题, 常用的方法是在优化模型 (1.9) 中加入一个正则项 (这也是岭回归[1]的思想), 得到如下最优化模型

$$\min_{\mathbf{x}} f(\mathbf{x}) = \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 + \lambda \|\mathbf{x}\|^2. \quad (1.36)$$

这里 λ 为一个需要人为设定的正数. 为了得到 (1.36) 的解, 我们可以先计算目标函数对自变量的导数, 并令其等于零向量

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = 2\mathbf{A}^T \mathbf{A} \mathbf{x} - 2\mathbf{A}^T \mathbf{b} + 2\lambda \mathbf{x} \stackrel{\text{令}}{=} \mathbf{0},$$

可得

$$\mathbf{x} = (\mathbf{A}^T \mathbf{A} + \lambda \mathbf{I})^{-1} \mathbf{A}^T \mathbf{b}. \quad (1.37)$$

其中 $\mathbf{I}$ 为相应阶数的单位矩阵.

当矩阵 $\mathbf{A}$ 的条件数很大时, $\mathbf{A}^T \mathbf{A}$ 必然包含一个相对特别小的特征值, 而对应的逆矩阵 $(\mathbf{A}^T \mathbf{A})^{-1}$ 必然对某些方向 (比如属于该特征值的特征向量方向) 的向量具有极大的放大作用, 这正是相应的线性方程组为病态方程组的原因所在. 而 (1.37) 中扰动项 $\lambda \mathbf{I}$ 的加入, 使得 $\mathbf{A}^T \mathbf{A} + \lambda \mathbf{I}$ 的条件数一般情况下都要远小于 $\mathbf{A}^T \mathbf{A}$ 的条件数, 从而可以一定程度克服线性方程组的病态问题.

1.6.4 异常问题

在直线拟合问题中, 当因变量和自变量的观测点的分布呈现一定的线性关系的时候, 可以用最小二乘法得到很好的拟合结果. 但在实际应用中, 当观测点受到噪声污染的时候 (如图 1.12), 直接利用最小二乘法对包括噪声点在内的所有观测点进行直线拟合在很多情况下难以得到理想的结果.

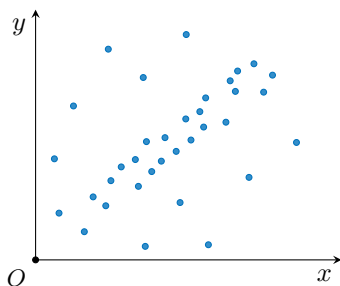


图 1.12 受噪声污染的观测数据

为了解决噪声污染下的直线拟合问题, 研究人员发展了各种策略, 其中常用的包括加权最小二乘法、Huber 回归、随机采样一致性 (Random Sample Consensus, RANSAC) [2] 等方法. 接下来, 简要介绍一下 RANSAC 的大体思路.

利用 RANSAC 进行直线拟合可以分为初筛和拟合两个阶段. 在初筛阶段, 随机给出包含 s 个散点的原始观测点的一个子集, 然后对这 s 个点利用最小二乘法进行直线拟合. 设定一个阈值 d , 记录下所有的观测点中与该直线的距离小于 d 的点 (称为内点), 并将所有的内点存储在一个集合 S_i 中. 重复以上步骤, 当出现内点个数增加的情形, 则更新 S_i . 对最终的内点集合 S_i 进行直线拟合即为我们想要的最终结果. 图 1.13 ~ 图 1.15 给出了初筛过程的简单示意图, 从中可以看出, 图 1.15 包含了更多的内点, 利用这些内点进行直线拟合显然会得到我们想要的结果.

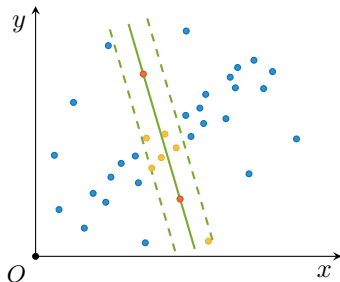


图 1.13 随机采样一致性. 随机子集的散点标为红色, 距离随机子集的拟合直线的距离小于给定阈值的点标为黄色 (本图中有 6 个点)

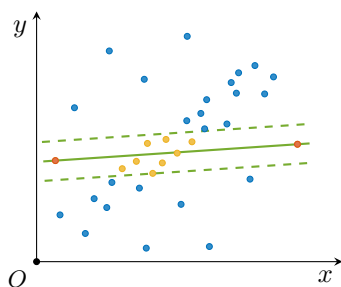


图 1.14 随机采样一致性. 随机子集的散点标为红色, 距离随机子集的拟合直线的距离小于给定阈值的点标为黄色 (本图中有 8 个点)

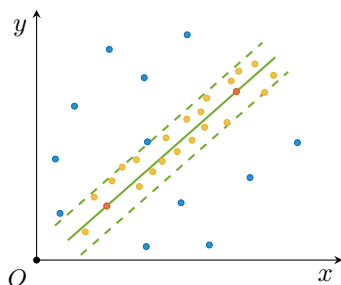


图 1.15 随机采样一致性. 随机子集的散点标为红色, 距离随机子集的拟合直线的距离小于给定阈值的点标为黄色 (本图中有 21 个点)

1.6.5 目标函数问题

最小二乘法不但可以用于直线拟合, 而且可以用于曲线拟合; 不仅可以处理单变量的观测数据, 同时也可以处理多变量的观测数据 (对应曲面拟合).

给定一组观测数据 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, 下面给出用 m 次多项式来拟合这些散点的过程

$$y = f(x) = a_0 x^m + a_1 x^{m-1} + \dots + a_{m-1} x + a_m.$$

注意到, 参数 $\mathbf{a} = \begin{bmatrix} a_0 & a_1 & \dots & a_m \end{bmatrix}^T$ 的确定等价于下面的最优化问题

$$\min_{\mathbf{a}} g(\mathbf{a}) = \|\mathbf{X}\mathbf{a} - \mathbf{y}\|^2, \quad (1.38)$$

其中

$$\begin{aligned} \mathbf{x}_i &= \begin{bmatrix} x_1^i & x_2^i & \cdots & x_n^i \end{bmatrix}^T, i = 1, 2, \cdots, m, \\ \mathbf{y} &= \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix}^T, \\ \mathbf{1} &= \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}^T, \\ \mathbf{X} &= \begin{bmatrix} \mathbf{x}_m & \cdots & \mathbf{x}_1 & \mathbf{1} \end{bmatrix}, \end{aligned}$$

则很容易得到 (1.38) 的最小二乘解为

$$\mathbf{a} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}. \quad (1.39)$$

比较 (1.15) 可以看出, 直线拟合和曲线拟合并没有本质的区别.

在多数情况下, 对于一组给定的观测点, 我们首先需要选取适当的目标函数, 然后利用最小二乘法对其进行直线或者曲线拟合. 但有些情况下, 我们很难用一个简单的函数来描述给定的散点分布, 如图 1.16. 对于这种情况的曲线拟合, 该如何进行处理呢?

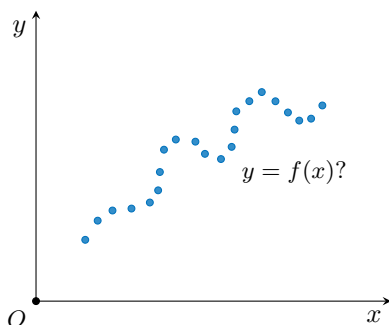


图 1.16 未知目标函数表达式的曲线拟合问题

第一种常用的方法是采用多项式对上述散点进行拟合. 如果多项式的次数设定合理, 在多数情况下, 多项式拟合都会得到不错的结果. 但在多项式拟合中, 多项式的次数是一个需要人为设定的量, 不合理的次数会导致过拟合或者欠拟合现象.

第二种常用的方法则是利用局部最小二乘法对任意形状的散点分布进行曲线拟合. 假设最终的拟合函数为 $y = f(x)$, 自变量 x 的每一个位置处, 相应的 y 值只需要利用该位置附近的若干观测点进行直线拟合确定. 该方法隐含了一个前提, 即数据的分布在局部是线性或者接近线性的. 在观测点足够多, 分布足够密的时候, 利用局部最小二乘法一般都会得到不错的结果.

1.7 小 结

至此，本章的内容总结为以下 6 条：

1. 从行空间角度，线性方程组的解为各个方程所对应的超平面的交集.
2. 从列空间角度，线性方程组的常数项向量为系数矩阵的各个列向量的线性表出，且表出系数即为待求的解.
3. 最小二乘法是求解矛盾方程组的常用方法.
4. 代数上，线性方程组的最小二乘解对应着矩阵的广义逆操作.
5. 几何上，线性方程组的最小二乘解等价于线性方程组的常数项向量在系数矩阵的列空间的正交投影.
6. 概率上，最小二乘法基于线性方程组模型误差服从高斯分布.

参考文献

- [1] Shalabh. Theory of Ridge Regression Estimation with Applications[J]. Journal of the Royal Statistical Society Series A: Statistics in Society, 2022, 185(2): 742-743.
- [2] FISCHLER Martin-A., BOLLES Robert-C. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography[G] //FISCHLER Martin-A., FIRSCHEIN Oscar. Readings in Computer Vision. San Francisco (CA): Morgan Kaufmann, 1987: 726-740.