

百分比距离与其他距离最小二乘法拟合研究

1 百分比距离原理

以一元线性回归为例, 设有一组 n 个试验数据 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, 欲寻求线性函数(理论分析证明二者存在线性相关性) : $y = ax + b$. 式中 a, b 均为待常数. 残差平方和定义为

$$D = \sum_{i=1}^n \frac{y_i}{y_i} \frac{ax_i - b}{y_i}^2$$

对上式求偏导, 并另偏导为0, 则

$$\begin{aligned} \frac{\partial Q}{\partial a} &= -2 \sum (y_i - ax_i - b)x_i/y_i^2 = 0 \\ \frac{\partial Q}{\partial b} &= -2 \sum (y_i - ax_i - b)/y_i^2 = 0 \end{aligned}$$

即

$$\begin{aligned} a \sum x_i^2/y_i^2 + b \sum x_i/y_i^2 &= \sum x_i/y_i \\ a \sum x_i/y_i^2 + b \sum 1/y_i^2 &= \sum 1/y_i \end{aligned}$$

可以求得该一元二次方程的解 $\begin{cases} a_3 = \frac{\sum x_i/y_i \sum 1/y_i^2 - \sum 1/y_i \sum x_i/y_i^2}{\sum x_i^2/y_i^2 \sum 1/y_i^2 - \sum 1/y_i \sum x_i/y_i^2} \\ b_3 = \frac{\sum x_i/y_i - a_3 \sum x_i^2/y_i^2}{\sum x_i/y_i^2} \end{cases}$

2. 实例对比

使用程序软件: R

数据集S1: 100个数据点, 符合曲线 $y = 2x + 1$ 变化趋势, 考虑数据的正常误差是影响因素(包括操作人员、机械等)的微小变化造成的, 这类误差即随机误差, 往往呈现正态分布, 因此加一个正态随机偏差, 符合 $N(0, 5)$ 分布。

数据集S2: 196个数据点, 符合曲线 $y = 2x + 1$ 变化趋势, 偏差符合 $N(0, 10)$ 分布, 越靠近原点越集中。

数据集S3: 278个数据点, 符合曲线 $y = 2x + 1$ 变化趋势, 偏差符合 $N(0, 10)$ 分布, 越靠近原点越集中。

用三种方法对以上数据集进行拟合, 拟合结果如下表1所示, 其中radio指y最大绝对值最小值比:

表1 拟合结果参数表

数据集	相关系数	a_1	a_2	a_3	b_1	b_2	b_3	radio
S1	0.9844	2.0659	2.1175	1.9987	-1.6543	-4.2661	1.7386	75.32
S2	0.9828	2.0451	2.0869	1.8197	-1.0395	-2.4085	6.3645	227.3

S3	0.9830	1.9892	2.0048	0.4717	0.6151	-0.7382	37.5939	1011.2
----	--------	--------	--------	--------	--------	---------	---------	--------

拟合图像如下图所示。

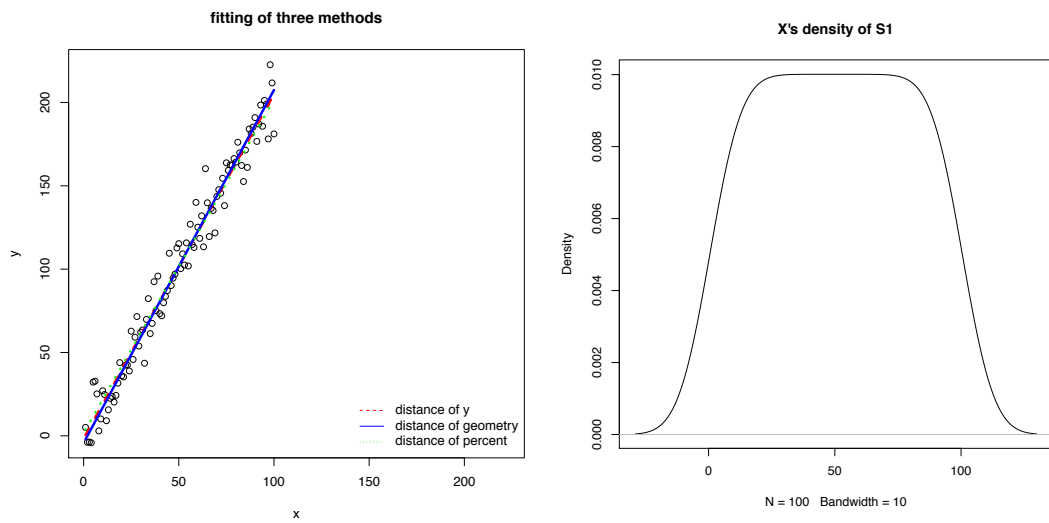


图1. S1图像与数据点的概率分布

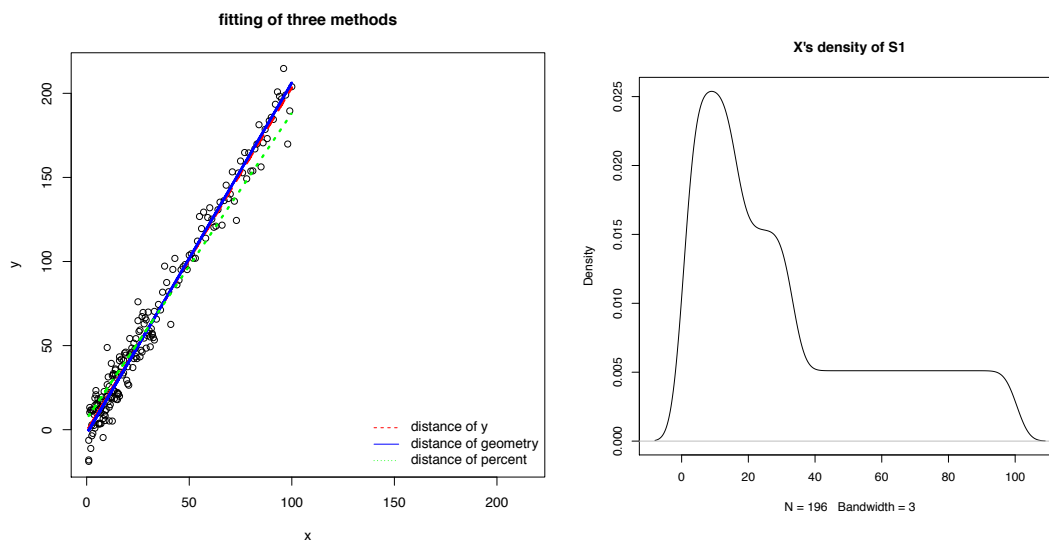


图2. S2图像与数据点的概率分布

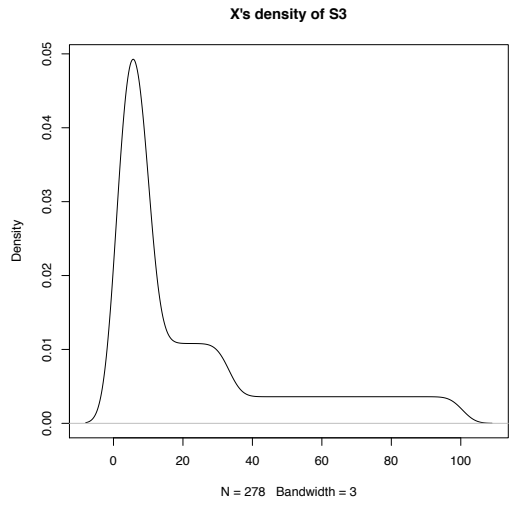
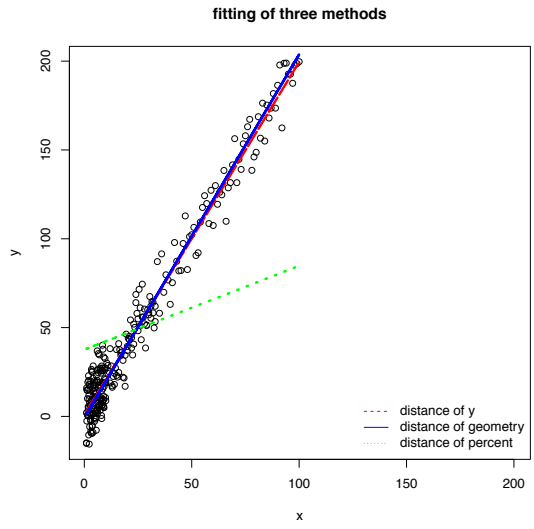


图3 S1图像与数据点的概率分布