

# 试题一：随机抽样一致性算法

在一组含有粗差的数据中，循环进行随机抽样和一致性评估，实现参数的稳健估计。

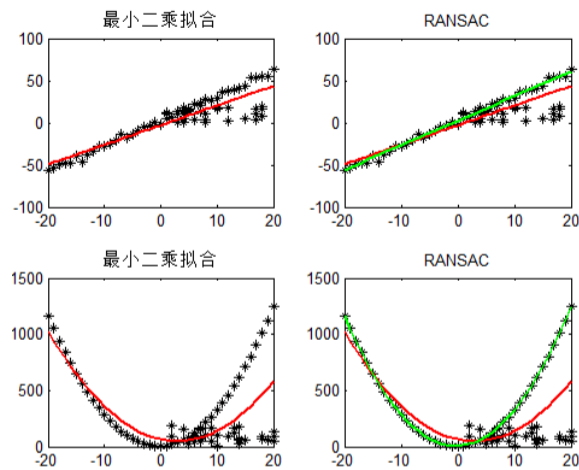


图 1 最小二乘与 RANSAC 算法对比

## 一、数据文件读取

编写程序，读取文件“data.txt”，每行记录包括“ID、X、Y”等内容，基本格式见表 1。

表 1 数据格式说明

数据内容	格式说明
n, k, t, d	模型参数
P1, 1, 2	点名，X 坐标，Y 坐标
P2, 2, 3	n, k, t, d 分别代表适应于模型的最少数 据个数，算法的迭代次数，模型的阈值， 模型合适的最小数据组数
P3, 3, 4	
.....	

【程序正确性】给出点 P<sub>17</sub> 的 X、Y 坐标。

## 二、程序算法

RANSAC(Random Sample Consensus,随机采样一致)算法是从一组含有“外点”(outliers)的数据中正确估计数学模型参数的迭代算法。其假设一组数据是由“外点”和“内点”构成，随机选取一组最小数据样本（n，模型的最少数据个数）来确定一个基础模型（在本文中为线性），通过该模型对全部数据进行验证，将与模型拟合程度符合阈值（t）的数据点标记为内点，重复上述模型多次，比较不同迭代得到的模型选择最佳模型。

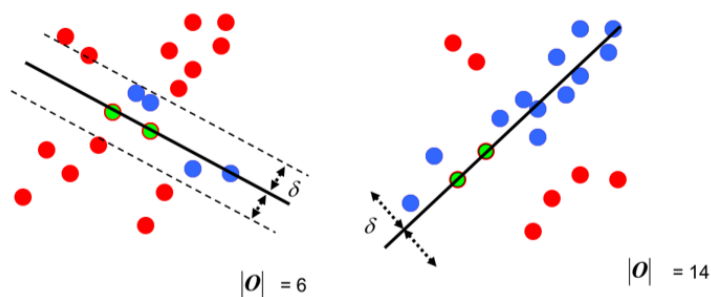


图 2 RANSAC 算法说明

## 1. 直线方程

从一组数据中随机不重复地选择  $n$ （此处为 2）个点，记为  $P_1(x_1, y_1)$  和  $P_2(x_2, y_2)$ ，对于直线拟合问题，直线的一般方程为：

$$Ax + By + C = 0 \quad (1.1)$$

其中  $(A, B, C)$  为模型参数，可由以下公式计算得到：

$$\begin{cases} A_0 = y_2 - y_1 \\ B_0 = x_1 - x_2 \\ C_0 = x_2 y_1 - x_1 y_2 \end{cases} \quad (1.2)$$

同时需要对该直线方程进行归一化处理，即

$$\begin{cases} A = \frac{A_0}{\sqrt{A_0^2 + B_0^2}} \\ B = \frac{B_0}{\sqrt{A_0^2 + B_0^2}} \\ C = \frac{C_0}{\sqrt{A_0^2 + B_0^2}} \end{cases} \quad (1.3)$$

**【程序正确性】** 给出由  $P_1$  和  $P_2$  组成的直线方程参数  $A_0$ 、 $B_0$ 、 $C_0$ 、 $A$ 、 $B$ 、 $C$  以及由  $P_5$  和  $P_{17}$  组成的直线方程参数  $A$ 、 $B$ 、 $C$ 。

## 2. 内点判断

对于数据集  $\text{data}$  中所有的已知点  $P_i(x_i, y_i)$ ，计算该点到直线的距离

$$d_i = |Ax_i + By_i + C| \quad (1.4)$$

若该点距离小于阈值  $(d_i < t)$ ，则标记为内点。

【程序正确性】给出直线  $P_1P_2$  的内点个数以及直线  $P_5P_{17}$  的内点个数。

### 3. 最佳模型

在该数据集中随机选取  $n$ （此处为 2）个不重复的样本点构成直线方程，当该直线对应的内点数量超过预设的最小阈值( $n_{th} > d$ )时，即判定为有效直线。通过迭代（最大迭代次数不超过  $k$  次）从所有有效直线模型中选取内点数量最多的解作为最佳直线。

【程序正确性】给出最佳直线的参数  $A_0$ 、 $B_0$ 、 $C_0$ 、 $A$ 、 $B$ 、 $C$  以及内点数量并判断点  $P_5$  和点  $P_{16}$  是否属于内点（是=1，否=0）。

### 4. 最小二乘拟合

将内点集合记作  $Q$ ，对该点集应用最小二乘法拟合直线，首先计算该组数据的平均值：

$$\begin{cases} \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i^2 \\ \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i^2 \end{cases} \quad (1.5)$$

计算协方差项

$$\begin{cases} S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2 \\ S_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2 \\ S_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \end{cases} \quad (1.6)$$

计算方程参数

$$\begin{cases} \hat{a} = \frac{S_{xy}}{S_{xx}} \\ \hat{b} = \bar{y} - \hat{a}\bar{x} \end{cases} \quad (1.7)$$

对方程系数归一化处理

$$\begin{cases} A = \frac{A_0}{\sqrt{A_0^2 + B_0^2}} \\ B = \frac{B_0}{\sqrt{A_0^2 + B_0^2}} \end{cases} \quad (1.8)$$

得到最终直线方程

$$Ax + By + C = 0$$

(1.9)

【程序正确性】给出最小二乘拟合的协方差  $S_{xx}$ 、 $S_{yy}$ 、 $S_{xy}$ ，直线方程参数  $A$ 、 $B$ 、 $C$ （注意考虑斜距等于零的情况）以及点  $P_{87}$  到该直线的距离。

5. 统计指标

计算内点个数占总点数的比例

$$\eta = \frac{n}{N}$$

(1.10)

计算最小二乘拟合后直线方程的均方根误差

$$RMSE = \frac{\sum_{i=1}^n d_i^2}{n}$$

(1.11)

【程序正确性】给出内点个数占比（小数形式如 0.1）和均方根误差。

三、计算报告的显示与保存

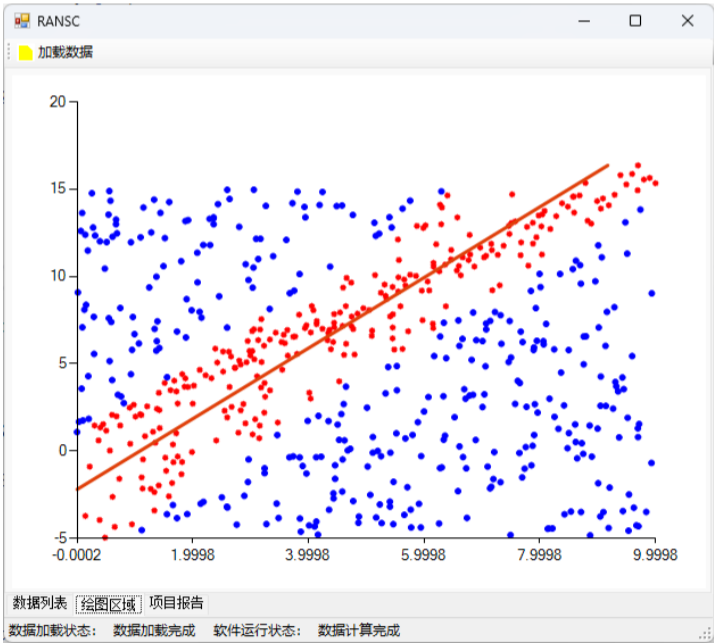
1. 程序正确性

根据读取的数据文件，编程完成相关算法，按照格式要求输出结果，如下表所示。

序号	说明	输出格式要求
1	点 $P_{17}$ 的 $x$ 坐标	*.*****（保留 5 位小数）
2	点 $P_{17}$ 的 $y$ 坐标	*.*****（保留 5 位小数）
3	直线 $P_1P_2$ 的参数 $A_0$	*.*****（保留 5 位小数）
4	直线 $P_1P_2$ 的参数 $B_0$	*.*****（保留 5 位小数）
5	直线 $P_1P_2$ 的参数 $C_0$	*.*****（保留 5 位小数）
6	直线 $P_1P_2$ 的参数 $A$	*.*****（保留 5 位小数）
7	直线 $P_1P_2$ 的参数 $B$	*.*****（保留 5 位小数）
8	直线 $P_1P_2$ 的参数 $C$	*.*****（保留 5 位小数）
9	直线 $P_5P_{17}$ 的参数 $A$	*.*****（保留 5 位小数）
10	直线 $P_5P_{17}$ 的参数 $B$	*.*****（保留 5 位小数）
11	直线 $P_5P_{17}$ 的参数 $C$	*.*****（保留 5 位小数）
12	直线 $P_1P_2$ 的内点个数	*（整数）
13	直线 $P_5P_{17}$ 的内点个数	*（整数）
14	最佳直线的参数 $A_0$	*.*****（保留 5 位小数）

15	最佳直线的参数 $B_0$	*.***** (保留 5 位小数)
16	最佳直线的参数 $C_0$	*.***** (保留 5 位小数)
17	最佳直线的参数 $A$	*.***** (保留 5 位小数)
18	最佳直线的参数 $B$	*.***** (保留 5 位小数)
19	最佳直线的参数 $C$	*.***** (保留 5 位小数)
20	点 $P_5$ 是否为最佳直线的内点	0/1
21	点 $P_{16}$ 是否为最佳直线的内点	0/1
22	最小二乘拟合直线的协方差 $S_{xx}$	*.***** (保留 5 位小数)
23	最小二乘拟合直线的协方差 $S_{yy}$	*.***** (保留 5 位小数)
24	最小二乘拟合直线的协方差 $S_{xy}$	*.***** (保留 5 位小数)
25	最小二乘拟合直线的参数 $A$	*.***** (保留 5 位小数)
26	最小二乘拟合直线的参数 $B$	*.***** (保留 5 位小数)
27	最小二乘拟合直线的参数 $C$	*.***** (保留 5 位小数)
28	点 $P_{87}$ 到拟合直线的距离	*.***** (保留 5 位小数)
29	内点个数占比 (小数)	*.***** (保留 5 位小数)
30	均方根误差 RMSE	*.***** (保留 5 位小数)

2. 参考界面



张祖全  
二〇二五年七月