Problem4

1.Description

RANSAC is widely used in fitting models from sample points with outliers. Please implement a program to fit a straight 2D line using RANSAC from the following sample points:(-2, 0), (0, 0.9), (2, 2.0), (3, 6.5), (4, 2.9), (5, 8.8), (6, 3.95), (8, 5.03), (10, 5.97),(12, 7.1), (13, 1.2), (14, 8.2), (16, 8.5) (18, 10.1). Please show your result graphically.

2.Solution

RANSAC Is an algorithm that finds the optimal model in a dataset containing outliers by random sampling. The core idea of RANSAC is to fit the model based on the random selection of data points, and set a threshold to judge which points fit the model, and finally obtain the optimal model through the inner points.

The main process of RANSAC is as follows:

• **Input data points**

Given a set of data points (two-dimensional coordinate points).

• **Set parameters**

Set the maximum number of iterations and the distance threshold, respectively, used to control the number of iterations of the algorithm and the standard to judge whether a point is an internal point.

• **Random sampling**

Two points are randomly selected from the data set each time, and a model of a straight line is calculated from these two points (slope and intercept).

• **Inner point judgment**

Calculate the vertical distance of the other point to the line. If the distance is less than the set threshold value, the point is considered to be the inner point. The more inner points, the better the line model is.

• **Update the optimal model**

Repeat the above steps, and record the model with the most internal points each time as the current optimal model.

- **Least squares fitting**

After finally obtaining a model with the most inner points, all the inner points are used to accurately fit the straight lines through the least squares method to obtain the final best model parameters (slope and intercept).

- **Output**

Returns the parameters (slope and intercept), the inner point and the outer point.

3.Result

Setting different iterations will theoretically get different results, but because of the few data points and the simple characteristics of the data points, the results of three or more iterations tend to be stable.
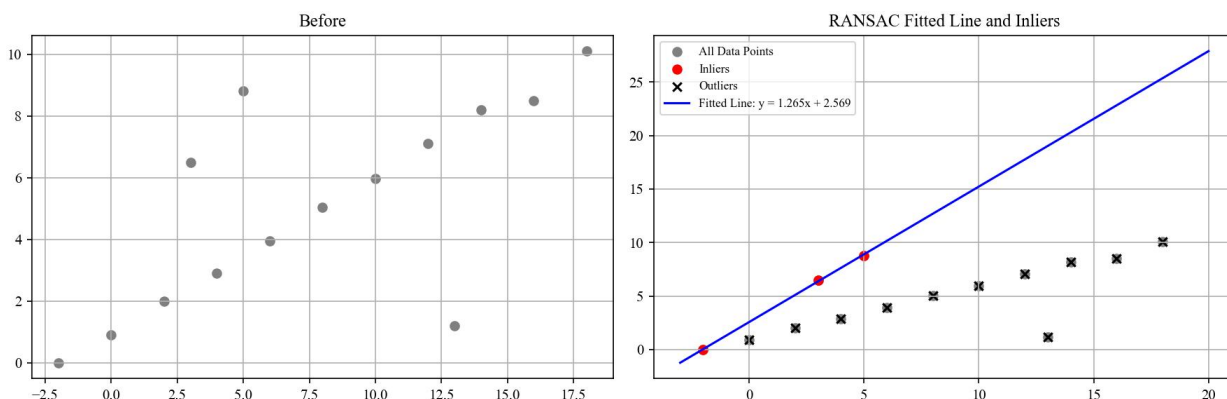
The number of iterations was 1: $\qquad y = -3.600x + 17.300$

The fitting effect was significantly less satisfactory.
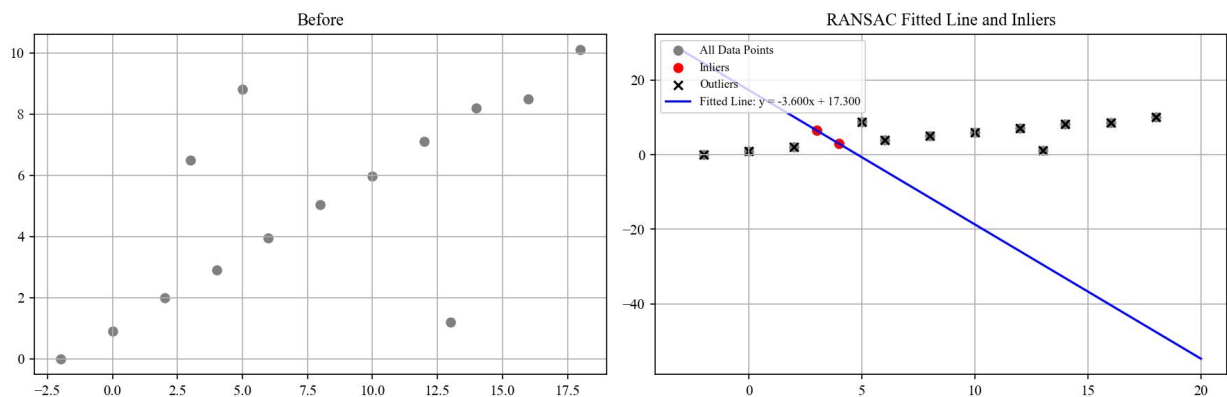
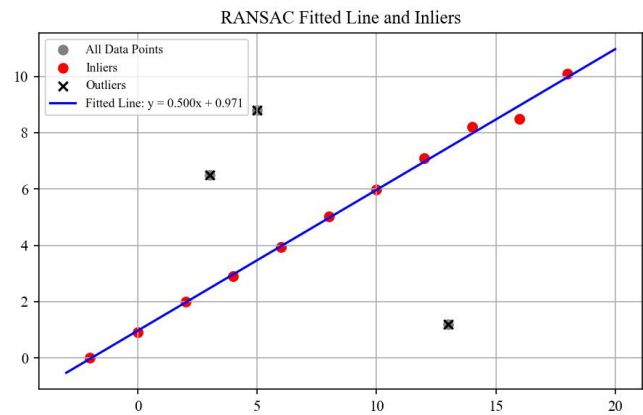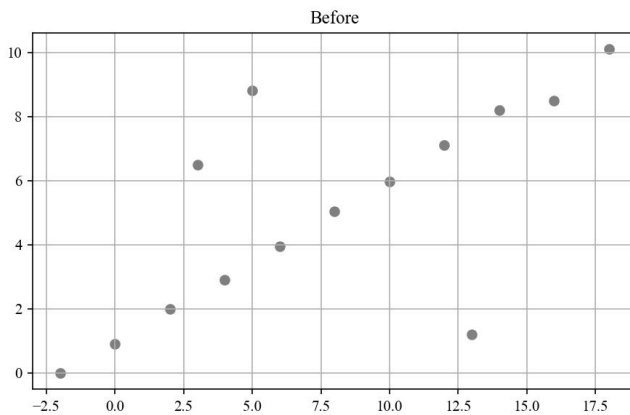The number of iterations was 2: $\qquad y = 1.265x + 2.569$

The effect is getting a little better.

The number of iterations was 3: $y = 0.500x + 0.971$

The fitting effect is already very good.

Finally, the straight line fit from the data point set by the RANSAC algorithm(The number of iterations was 100) is:

$$y = 0.500x + 0.971$$

As shown in the figure below(Image path: Problem4\image\Problem4.png):