11 Diagnosis of Outliers

Overview

Gauss - Jacobi combinatorial algorithm can be employed to detect outliers among data samples. Computing the positional norm of the solution of the different subsets, the common station(s) of the subsets having mostly deviating norm from the median, can be suspected as outlier(s). This non-statistical outlier detection method is demonstrated in case of a planar ranging problem.

11-1 Problem Definition

Outliers are those observations that are inconsistent with the rest of the observation samples. They often degrade the quality of the estimated parameters and render them unreliable for any meaningful inferences. Here a non-statistical algebraic approach to outlier diagnosis is presented, that uses Gauss-Jacobi combinatorial algorithm.

11-2 Illustrative Example

Let us recall the planar ranging problem considered in Section 1-2-2. The prototype of the equations is,

\[
\text{In}[90]:= \text{Clear}[^\text{Global`*}^] \\
\text{In}[91]:= e = (x_i - x_0)^2 + (y_i - y_0)^2 - s_i^2 \quad / \quad \text{Expand} \\
\text{Out}[91]= x_0^2 + y_0^2 - s_i^2 - 2 x_0 x_i + x_i^2 - 2 y_0 y_i + y_i^2
\]

where \( i = 1 \ldots m \). Let us consider four known stations, \( m = 4 \). The numerical values are,

\[
\text{In}[92]:= \text{data2DN} = \{x_1 \rightarrow 48177.62, y_1 \rightarrow 6531.28, s_1 \rightarrow 611.023, \\
x_2 \rightarrow 49600.15, y_2 \rightarrow 7185.19, s_2 \rightarrow 1530.432, x_3 \rightarrow 49830.93, y_3 \rightarrow 5670.69, \\
s_3 \rightarrow 1323.884, x_4 \rightarrow 47863.91, y_4 \rightarrow 5077.24, s_4 \rightarrow 1206.524\};
\]

For this case, the distance \( s_2 \) is falsified such that the observed value is recorded as 1530.432 m instead of the correct value 1529.482 m.

The number of the combinations are,

\[
\text{In}[93]:= n = 2; m = 4; \\
\text{In}[94]:= \text{mn = Binomial[m, n]} \\
\text{Out}[94]= 6
\]

The pairs of the combinations are,

\[
\text{In}[95]:= \text{qs = Partition[Map[# & , Flatten[Subsets[Range[m], (n)]]], n]} \\
\text{Out}[95]= \{\{1, 2\}, \{1, 3\}, \{1, 4\}, \{2, 3\}, \{2, 4\}, \{3, 4\}\}
\]

The corresponding data list,
Let us combine the two lists, and let us compute the value of the least square objective for all pairs. The objective function is,

\[
\text{obj} = \sum_{i=1}^{m} \left[ (x_0^2 + y_0^2 - s_i^2 - 2x_0 x_i + x_i^2 - 2y_0 y_i + y_i^2)^2 + (x_0^2 + y_0^2 - s_i^2 - 2x_0 x_i + x_i^2 - 2y_0 y_i + y_i^2)^2 + (x_0^2 + y_0^2 - s_i^2 - 2x_0 x_i + x_i^2 - 2y_0 y_i + y_i^2)^2 \right]
\]

The values of the objectives are,

\[
\text{obj} = \{6.78456 \times 10^6, 8.11877 \times 10^6, 8.26366 \times 10^6, 6.15198 \times 10^6, 2.10599 \times 10^{11}, 3.47129 \times 10^{13}\}
\]

and

\[
\text{obj} = \{2.00489 \times 10^{13}, 3.3126 \times 10^{12}, 1.75753 \times 10^{13}, 1.45782 \times 10^{14}, 4.16504 \times 10^9, 8.25303 \times 10^6\}
\]

Let us combine the two lists,

\[
\text{combi} = \{6.78456 \times 10^6, (48.564.5, 6058.37), \{8.11877 \times 10^6, (48.565.3, 6058.96)\}, \{8.26366 \times 10^6, (48.565.3, 6058.97)\}, \{6.15198 \times 10^6, (48.565., 6057.98)\}, \{2.10599 \times 10^{11}, (48.698.4, 5948.64)\}, \{3.47129 \times 10^{13}, (48.991.2, 4647.21)\}\}
\]
We shall select the solution from the two possible ones of each combinations, which gives the smaller objective value,

\[
\text{solxynm} = \text{Table}[\text{If}[\text{combi1}[[i, 1]] > \text{combi2}[[i, 1]], \text{combi2}[[i, 2]], \text{combi1}[[i, 2]]], \{i, 1, \text{mn}\}]
\]

Now we can compute the positional norms,

\[
\text{d} = \text{Map}[(2 \times \text{Norm}[\#] \&), \text{solxynm}]
\]

The median,

\[
\text{M} = \text{Median}[\text{d}]
\]

Deviations from the median,

\[
\text{d} - \text{M} \rightarrow \text{ListPlot}[(\text{d}, \text{Joined} \to \text{True}, \text{Frame} \to \text{True}, \text{Axes} \to \text{False}, \text{PlotRange} \to \text{All})]
\]

From the combinations,

\[
\text{qs} = \text{Table}[[\text{objV2}[[1]], \text{solxynm}[[1]], \{1, 1, \text{mn}\}]
\]

it can be clearly realized that in those cases, when the station 2 is included in the combinations, namely 1: (1, 2), 4: (2, 3) and 5: (2, 4) the deviations are larger than those of the combinations without station 2.

**Remark**: In case of GPS Positioning in Section 2- 3- 2 the result of the Gauss-Jacobi combinatorial algorithm indicated the poor geometry of the 10th combination, see Figs. 2.1 and 2.2.

**Conclusions**

This method can be useful, but probably only for relatively modest number of stations, since with increasing number of stations, the number of combinations is running away (combinatoric explosion).
the number of stations, the number of combinations is running away (combinatorial explosion).