12.11 Curve fitting

A quantity y is known to depend on another quantity x. A set of corresponding values has been collected for x and y and is presented in Table 12.4.

1. Fit the 'best' straight line $y = bx + a$ to this set of data points. The objective is to minimise the sum of *absolute deviations* of each observed value of y from the value predicted by the linear relationship.
2. Fit the 'best' straight line where the objective is to minimise the *maximum deviation* of all the observed values of y from the value predicted by the linear relationship.
3. Fit the 'best' quadratic curve $y = cx^2 + bx + a$ to this set of data points using the same objectives as in (1) and (2).

Table 12.4

x	0.0	0.5	1.0	1.5	1.9	2.5	3.0	3.5	4.0	4.5
y	1.0	0.9	0.7	1.5	2.0	2.4	3.2	2.0	2.7	3.5
x	5.0	5.5	6.0	6.6	7.0	7.6	8.5	9.0	10.0	
y	1.0	4.0	3.6	2.7	5.7	4.6	6.0	6.8	7.3	

12.12 Logical design

Logical circuits have a given number of inputs and one output. Impulses may be applied to the inputs of a given logical circuit, and it will respond by giving either an output (signal 1) or no output (signal 0). The input impulses are of the same kind as the outputs, that is, 1 (positive input) or 0 (no input).

In this example, a logical circuit is to be built up of NOR gates. A NOR gate is a device with two inputs and one output. It has the property that there is positive output (signal 1) if and only if *neither* input is positive, that is, both inputs have the value 0. By connecting such gates together with outputs from one gate possibly being inputs into another gate, it is possible to construct a circuit to perform any desired logical function. For example, the circuit illustrated in Figure 12.1 will respond to the inputs A and B in the way indicated by the truth table.

The problem here is to construct a circuit using the *minimum number* of NOR gates that will perform the logical function specified by the truth table in Figure 12.2. This problem, together with further references to it, is discussed in Williams (1974).

'Fan-in' and 'fan-out' are not permitted. That is, more than one output from a NOR gate cannot lead into one input nor can one output lead into more than one input.

It may be assumed throughout that the optimal design is a 'subnet' of the 'maximal' net shown in Figure 12.3.

13.10.3 Objective

The objective is to minimize

$$-\sum_{i,j} B_{ij}\delta_{ij} + \sum_{\substack{i,j,k,l\\ l<k}} C_{ik} D_{jl} \gamma_{ijkl},$$

where B_{ij} is the benefit to be gained from locating department i in city j as given in Part II (for $j = \mathrm{L}$ (London), $B_{ij} = 0$). C_{ik} and D_{ji} are given in the tables in Part II, Section 12.10.

This model has 162 constraints and 69 variables (all 0–1).

Beale and Tomlin formulate their model more compactly. As a consequence, the corresponding linear programming problem is much less constrained than it might be (see Section 10.1). They then expand some of the constraints. Use is then made of the branching strategy to avoid expanding other constraints.

13.11 Curve fitting

This is an application of the goal programming type of formulation discussed in Section 3.3. Each pair of corresponding data values (x_i, y_i) gives rise to a constraint. For cases (1) and (2), these constraints are

$$bx_i + a + u_i - v_i = y_i, \qquad i = 1, 2, \ldots, 19.$$

x_i and y_i are constants (the given values); b, a, u_i and v_i are variables. u_i and v_i give the amounts by which the values of y_i proposed by the linear expression differ from the observed values. It is important to allow a and b to be 'free' variables, that is, they can be allowed to take negative as well as positive values.

In case (1) the objective is to minimize

$$\sum_i u_i + \sum_i v_i.$$

This model has 19 constraints and 40 variables.

In case (2) it is necessary to introduce another variable z together with 38 more constraints:

$$z - u_i \geq 0, \qquad z - v_i \geq 0, \qquad i = 1, 2, \ldots, 19.$$

The objective, in this case, is simply to minimize z. This minimum value of z will clearly be exactly equal to the maximum value of v_i and u_i.

In case (3), it is necessary to introduce a new (free) variable c into the first set of constraints to give

$$cx_i^2 + bx_i + a + u_i - v_i = y, \qquad i = 1, 2, \ldots, 19.$$

The same objective functions as in (1) and (2) will apply.

It is much more usual in statistical problems to minimize the sum of squares of the deviations as the resultant curve often has desirable statistical properties. There are, however, some circumstances in which a sum of absolute deviations is acceptable or even more desirable. Moreover, the possibility of solving this type of problem by linear programming makes it computationally easy to deal with large quantities of data.

Minimizing the maximum deviation has certain attractions from the point of view of presentation. The possibility of a single data point appearing a long way off the fitted curve is minimized.

13.12 Logical design

In order to simplify the formulation, the optimal circuit can be assumed to be a subnet of the maximum shown in Figure 13.1. The following 0–1 integer variables are used:

$$s_i = \begin{cases} 1 & \text{if NOR gate } i \text{ exists, } i = 1, 2, \ldots, 7, \\ 0 & \text{otherwise;} \end{cases}$$

$$t_{i1} = \begin{cases} 1 & \text{if external input } A \text{ is an input to gate } i, \\ 0 & \text{otherwise;} \end{cases}$$

$$t_{i2} = \begin{cases} 1 & \text{if external input } B \text{ is an input to gate } i, \\ 0 & \text{otherwise.} \end{cases}$$

x_{ij} is the output from gate i for the combination of external input signals specified in the jth row of the truth table.

The following constraints are imposed.

A NOR gate can only have an external input if it exists. These conditions are imposed by the constraints

$$s_i - t_{i1} \geq 0, \qquad s_i - t_{i2} \geq 0, \qquad i = 1, 2, \ldots, 7.$$

If a NOR gate has one (or two) external inputs leading into it, only one (or no) NOR gates can feed into it. These conditions are imposed by the constraints

$$s_j + s_k + t_{i1} + t_{i2} \leq 2, \qquad i = 1, 2, 3$$

where j and k are the two NOR gates leading into i in Figure 13.1.

The output signal from NOR gate i must be the correct logical function (NOR) of the input signals into gate i if gate i exists. Let α_{1j} (a constant) be the value of the external input signal A in the jth row of the truth table. Similarly

14.10 Decentralization

The optimal solution is

locate departments A and D in Bristol;
locate departments B, C and E in Brighton.

This results in a yearly benefit of £80 000 but communications costs of £65 100. It is interesting to note that communication costs are also reduced by moving out of London in this problem because they would have been £78 000 if each department had remained in London.

The net yearly benefit (benefits less communication costs) is therefore £14 900.

14.11 Curve fitting

1. The 'best' straight line that minimizes the *sum of absolute deviations* is

$$y = 0.6375x + 0.5812.$$

This is line 1 shown in Figure 14.1. The sum of absolute deviations resulting from this line is 11.46.

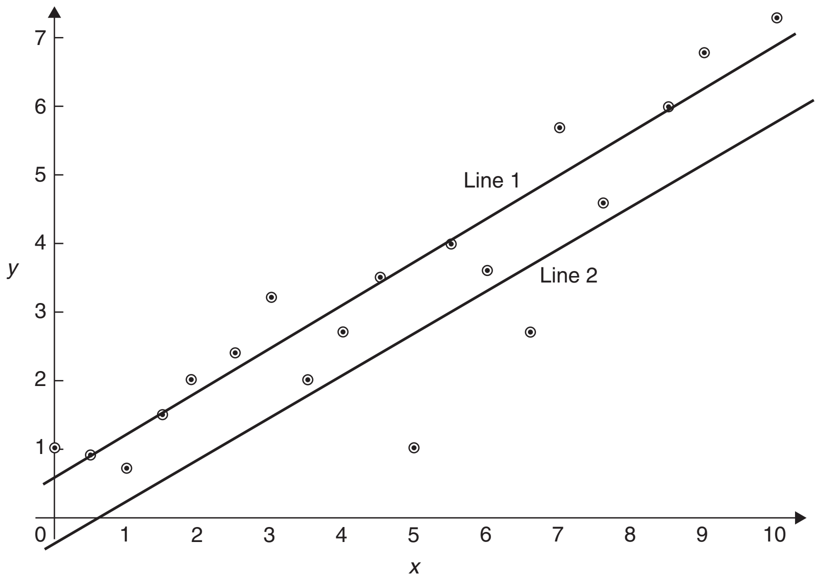

Figure 14.1

2. The ‘best’ straight line that minimizes the *maximum absolute deviation* is

$$y = 0.625x - 0.4.$$

This is line 2 shown in Figure 14.1. The maximum absolute deviation resulting from this line is 1.725. (Points (3.0, 3.2), (5.0, 1.0) and (7.0, 5.7) all have this absolute deviation from the line.) In contrast, line 1 allows point (5.0, 1.0) to have an absolute deviation of 2.77. On the other hand, although line 2 allows no point to have an absolute deviation of more than 1.725, the sum of absolute deviations is 19.95 compared with the 11.47 resulting from line 1.

3. The ‘best’ quadratic curve that minimizes the *sum of absolute deviations* is

$$y = 0.0337x^2 + 0.2945x + 0.9823.$$

This is curve 1 shown in Figure 14.2. The sum of absolute deviations resulting from this curve is 10.45.

4. The ‘best’ quadratic curve that minimizes the *maximum absolute deviation* is

$$y = 0.125x^2 - 0.625x + \ 2.475.$$

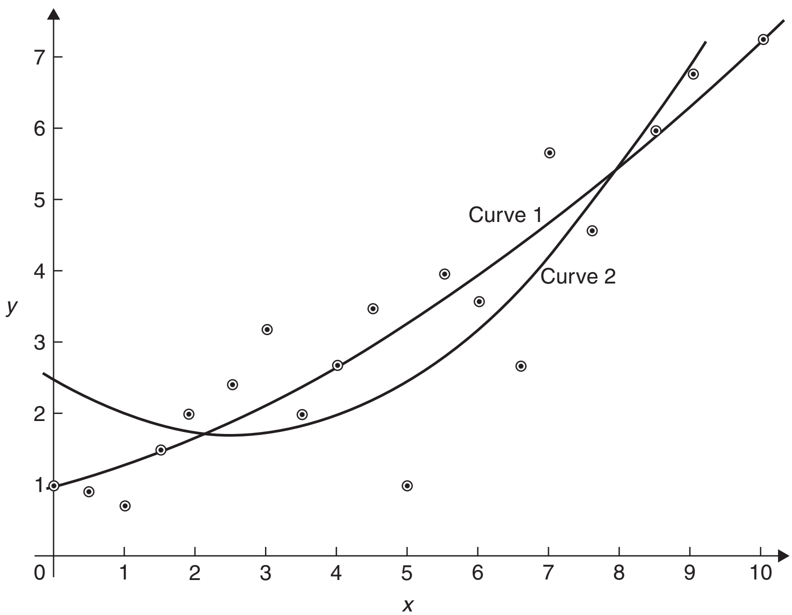

Figure 14.2

This is curve 2 shown in Figure 14.2. The maximum absolute deviation resulting from this curve is 1.475. (Points (0.0, 1.1), (3.0, 3.2), (5.0, 1.0) and (7.0, 5.7) all have this absolute deviation from the curve.)

A way of obtaining analytical solutions to these sorts of problems is described by Williams and Munford (1999).

14.12 Logical design

The optimal solution is shown in Figure 14.3. There are, of course, symmetric alternatives.

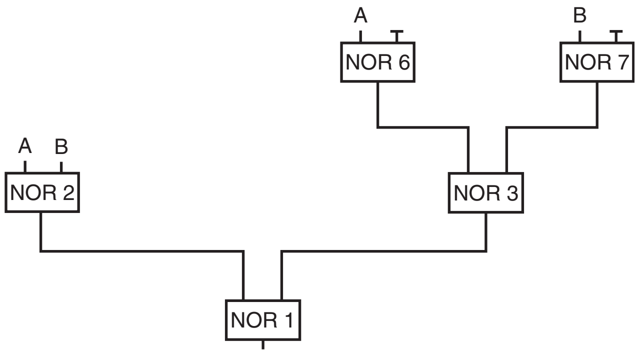

Figure 14.3

14.13 Market sharing

This model proves difficult to solve optimally although a feasible solution is comparatively easy to obtain. Using the original formulation with no extra constraints, a feasible solution was found after 21 nodes by minimizing the sum of percentage deviations. This solution gave a sum of percentage deviations of 8.49 with the maximum such deviation being 3.73% (the Region 3 OIL goal). The best solution obtained in this run was after 360 nodes with a sum of percentage deviations of 4.53, the maximum such deviation being 2.5% (the DELIVERY POINTS goal). After a total of 1995 nodes, the search was completed.

Minimizing the maximum percentage deviation produced a feasible solution after 37 nodes. This solution gave a maximum deviation of 2.5% (the OIL in region 3 goal). The sum of percentage deviations was 8.82. This was proved to be the optimal solution after 3158 nodes.

The second formulation was obtained by adding 228 ‘facets’ for the OIL goals in the three regions. The increased size of this model drastically reduced the number of nodes that could be explored in a given time. After 10 s with the objective of minimizing the sum of percentage deviations, 75 nodes had been