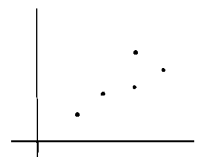
LS. Least Squares Interpolation

1. The least-squares line.

Suppose you have a large number n of experimentally determined points, through which you want to pass a curve. There is a formula (the Lagrange interpolation formula) producing a polynomial curve of degree $n-1$ which goes through the points exactly. But normally one wants to find a simple curve, like a line, parabola, or exponential, which goes approximately through the points, rather than a high-degree polynomial which goes exactly through them. The reason is that the location of the points is to some extent determined by experimental error, so one wants a smooth-looking curve which averages out these errors, not a wiggly polynomial which takes them seriously.

In this section, we consider the most common case — finding a line which goes approximately through a set of data points.



Suppose the data points are

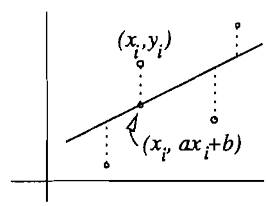
$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$$

and we want to find the line

$$y = ax + b \tag{1}$$

which "best" passes through them. Assuming our errors in measurement are distributed randomly according to the usual bell-shaped curve (the so-called "Gaussian distribution"), it can be shown that the right choice of a and b is the one for which the sum D of the squares of the deviations

$$D = \sum_{i=1}^{n} \big(y_i - (ax_i + b)\big)^2 \tag{2}$$



is a *minimum*. In the formula (2), the quantities in parentheses (shown by dotted lines in the picture) are the **deviations** between the observed values y_i and the ones $ax_i + b$ that would be predicted using the line (1).

The deviations are squared for theoretical reasons connected with the assumed Gaussian error distribution; note however that the effect is to ensure that we sum only positive quantities; this is important, since we do not want deviations of opposite sign to cancel each other out. It also weights more heavily the larger deviations, keeping experimenters honest, since they tend to ignore large deviations ("I had a headache that day").

This prescription for finding the line (1) is called the **method of least squares**, and the resulting line (1) is called the **least-squares** line or the **regression** line.

To calculate the values of a and b which make D a minimum, we see where the two partial derivatives are zero:

$$\begin{aligned} \frac{\partial D}{\partial a} &= \sum_{i=1}^{n} 2(y_i - ax_i - b)(-x_i) \;=\; 0 \\ \frac{\partial D}{\partial b} &= \sum_{i=1}^{n} 2(y_i - ax_i - b)(-1) \;=\; 0\,. \end{aligned} \tag{3}$$

These give us a pair of *linear* equations for determining a and b, as we see by collecting terms and cancelling the 2's:

(4)
$$\begin{aligned} \left(\sum x_i^2\right) a \;+\; \left(\sum x_i\right) b \;&=\; \sum x_i y_i \\ \left(\sum x_i\right) a \;+\; n\,b \;&=\; \sum y_i\;. \end{aligned}$$

(Notice that it saves a lot of work to differentiate (2) using the chain rule, rather than first expanding out the squares.)

The equations (4) are usually divided by n to make them more expressive:

(5)
$$\begin{aligned} \bar{s}\,a \;+\; \bar{x}\,b \;&=\; \frac{1}{n}\sum x_i y_i \\ \bar{x}\,a \;+\; b \;&=\; \bar{y}\;, \end{aligned}$$

where $\bar{x}$ and $\bar{y}$ are the average of the x_i and y_i, and $\bar{s} = \sum x_i^2/n$ is the average of the squares.

From this point on use linear algebra to determine a and b. It is a good exercise to see that the equations are always solvable unless all the x_i are the same (in which case the best line is vertical and can't be written in the form (1)).

In practice, least-squares lines are found by pressing a calculator button, or giving a MatLab command. Examples of calculating a least-squares line are in the exercises in your book and these notes. Do them from scratch, starting from (2), since the purpose here is to get practice with max-min problems in several variables; don't plug into the equations (5). Remember to differentiate (2) using the chain rule; don't expand out the squares, which leads to messy algebra and highly probable error.

2. Fitting curves by least squares.

If the experimental points seem to follow a curve rather than a line, it might make more sense to try to fit a second-degree polynomial

(6)
$$y = a_0 + a_1 x + a_2 x^2$$

to them. If there are only three points, we can do this exactly (by the Lagrange interpolation formula). For more points, however, we once again seek the values of a_0, a_1, a_2 for which the sum of the squares of the deviations

(7)
$$D \;=\; \sum_1^n \left(y_i - (a_0 + a_1 x_i + a_2 x_i^2)\right)^2$$

is a minimum. Now there are three unknowns, a_0, a_1, a_2. Calculating (remember to use the chain rule!) the three partial derivatives $\partial D/\partial a_i$, $i = 0, 1, 2$, and setting them equal to zero leads to a square system of three linear equations; the a_i are the three unknowns, and the coefficients depend on the data points (x_i, y_i). They can be solved by finding the inverse matrix, elimination, or using a calculator or MatLab.

If the points seem to lie more and more along a line as $x \to \infty$, but lie on one side of the line for low values of x, it might be reasonable to try a function which has similar behavior, like

(8)
$$y \;=\; a_0 + a_1 x + a_2\,\frac{1}{x}$$

and again minimize the sum of the squares of the deviations, as in (7). In general, this method of least squares applies to a trial expression of the form

(9) $$y = a_0 f_0(x) + a_1 f_1(x) + \ldots + a_r f_r(x),$$

where the $f_i(x)$ are given functions (usually simple ones like $1, x, x^2, 1/x, e^{kx}$, etc. Such an expression (9) is called a **linear combination** of the functions $f_i(x)$. The method produces a square inhomogeneous system of linear equations in the unknowns $a_0, \ldots, a_r$ which can be solved by finding the inverse matrix to the system, or by elimination.

The method also applies to finding a linear function

(10) $$z = a_1 + a_2 x + a_3 y$$

to fit a set of data points

(11) $$(x_1, y_1, z_1), \ldots, (x_n, y_n, z_n) .$$

where there are two independent variables x and y and a dependent variable z (this is the quantity being experimentally measured, for different values of (x, y)). This time after differentiation we get a 3×3 system of linear equations for determining a_1, a_2, a_3 .

The essential point in all this is that the unknown coefficients a_i should occur *linearly* in the trial function. Try fitting a function like ce^{kx} to data points by using least squares, and you'll see the difficulty right away. (Since this is an important problem — fitting an exponential to data points — one of the Exercises explains how to adapt the method to this type of problem.)

Exercises: Section 2G