

Find a least squares formula to find the best fit for a set of data points,  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ , to one of these functions:

1.  $y = ax^2 + bx + c$
2.  $y = a \ln x + b$
3.  $y = ae^x + b$

Then, create a small (5-10) set of interesting data points to exercise your formula. Plot your points and your function on a graph. Show your work!

Here is a solution for  $y = ax^3 + bx^2 + cx + d$

We want to minimize the sum of the squares of the errors  $e_i$ :

(Here I use  $x$  and  $y$  instead of  $x_i$  and  $y_i$  just to keep the clutter down.)

$$\begin{aligned} f(a, b, c, d) &= \sum e_i^2 = \sum (ax^3 + bx^2 + cx + d - y)^2 \\ &= \sum (a^2x^6 + 2abx^5 + 2acx^4 + b^2x^4 + 2adx^3 - 2ax^3y + 2bcx^3 + 2bdx^2 - 2bx^2y + c^2x^2 \\ &\quad + 2cdx - 2cxy + d^2 - 2dy + y^2) \end{aligned}$$

$$f_a = \sum (2ax^6 + 2bx^5 + 2cx^4 + 2dx^3 - 2x^3y)$$

$$f_b = \sum (2ax^5 + 2bx^4 + 2cx^3 + 2dx^2 - 2x^2y)$$

$$f_c = \sum (2ax^4 + 2bx^3 + 2cx^2 + 2dx - 2xy)$$

$$f_d = \sum (2ax^3 + 2bx^2 + 2cx + 2d - 2y)$$

Set these derivatives equal to zero, divide each by two, move the constant (without a,b,c,d) to the right, write the summation and sums and differences of summations, and we obtain this system of equations:

$$\begin{bmatrix} a \sum x^6 + b \sum x^5 + c \sum x^4 + d \sum x^3 = \sum x^3y \\ a \sum x^5 + b \sum x^4 + c \sum x^3 + d \sum x^2 = \sum x^2y \\ a \sum x^4 + b \sum x^3 + c \sum x^2 + d \sum x = \sum xy \\ a \sum x^3 + b \sum x^2 + c \sum x + d \sum 1 = \sum y \end{bmatrix}$$

One can find the solution to each variable a,b,c,d:

$$a = \frac{\begin{vmatrix} \sum x^3y & \sum x^5 & \sum x^4 & \sum x^3 \\ \sum x^2y & \sum x^4 & \sum x^3 & \sum x^2 \\ \sum xy & \sum x^3 & \sum x^2 & \sum x \\ \sum y & \sum x^2 & \sum x & \sum 1 \end{vmatrix}}{\begin{vmatrix} \sum x^6 & \sum x^5 & \sum x^4 & \sum x^3 \\ \sum x^5 & \sum x^4 & \sum x^3 & \sum x^2 \\ \sum x^4 & \sum x^3 & \sum x^2 & \sum x \\ \sum x^3 & \sum x^2 & \sum x & \sum 1 \end{vmatrix}}$$

Note that the top matrix is the bottom coefficient matrix with the a-column replaced with the constant column. Do the same to find formulas for b,c, and d.