Notes on the least squares method

Patrick De Leenheer*

January 14, 2009

You have probably heard about a "least squares fit", and when you read papers in mathematical biology this phrase often pops up. These notes provide justification to the existence and uniqueness of the least squares fit to a given data set.

The problem is as follows: Given a data set (x_i, y_i) with $i = 1, 2, \ldots, n$, find the equation of a line $y = ax + b$ that is the best fit (in some sense to be specified later) to these data points. Geometrically, you are trying to draw a straight line in the (x, y)-plane that approximates the n given data points in the best possible way. The best fit is by definition the one that minimizes the sum of squares of the errors between the predicted y-values on the line, and the data y-values. This explains the terminology of "least squares fit". The precise mathematical problem is to

$$\text{Minimize } \sum_{i=1}^{n}(ax_i + b - y_i)^2 \text{ over all } a, b \in \mathbb{R}.$$

Closer inspection of the problem tells us that we are trying to minimize a quadratic function in two unknowns a and b. Let us call this function[1]

$$f(a,b) = \sum_{i=1}^{n}(ax_i+b-y_i)^2 = \left(\sum x_i^2\right)a^2+2ab\left(\sum x_i\right)-2a\left(\sum x_iy_i\right)+nb^2-2b\left(\sum y_i\right)+\left(\sum y_i^2\right)$$

We know that a minimum can only occur at points where:

$$\frac{\partial f}{\partial a} = \frac{\partial f}{\partial b} = 0.$$

Using a more compact notation f_a and f_b for the partial derivatives, let's investigate where these equations take us:

$$\begin{aligned} f_a &= 2\left(\sum x_i^2\right)a + 2\left(\sum x_i\right)b - 2\left(\sum x_iy_i\right) = 0 \\ f_b &= 2\left(\sum x_i\right)a + (2n)b - 2\left(\sum y_i\right) = 0, \end{aligned}$$

or in matrix form:

$$\begin{pmatrix} \sum x_i^2 & \sum x_i \\ \sum x_i & n \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \sum x_iy_i \\ \sum y_i \end{pmatrix}$$

We can solve this by multiplying both sides on the left with the inverse of the matrix, to get

$$\begin{pmatrix} a^* \\ b^* \end{pmatrix} = \frac{1}{n\sum x_i^2 - \left(\sum x_i\right)^2}\begin{pmatrix} n & -\sum x_i \\ -\sum x_i & \sum x_i^2 \end{pmatrix}\begin{pmatrix} \sum x_iy_i \\ \sum y_i \end{pmatrix},$$

provided that

$$n\sum x_i^2 - \left(\sum x_i\right)^2 \neq 0. \tag{1}$$

I claim that (1) is satisfied if and only if the given data points do not lie on a vertical line, i.e. if and only if the n values x_i are not all equal to each other. To prove this we will show that

$$n\sum x_i^2 - \left(\sum x_i\right)^2 = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i - x_j)^2, \tag{2}$$

*Email: deleenhe@math.ufl.edu. Department of Mathematics, University of Florida.

[1] From now on, I'm suppressing the indices in the summation symbols $\sum$.

from which the claim follows immediately. Now,

$$\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i-x_j)^2=\sum_{i=1}^{n}\sum_{j=1}^{n}x_i^2-2x_ix_j+x_j^2=2n\sum x_i^2-2\sum_{i=1}^{n}\sum_{j=1}^{n}x_ix_j=2\left(n\sum x_i^2-\left(\sum x_i\right)^2\right),$$

from which (2) is clear.

Under condition (1) we have thus found a critical point $\left(a^*,b^*\right)$ of the function $f(a,b)$. But that does not necessarily mean that we get a minimum for $f(a,b)$ there! Here's a mouthfull:

A sufficient condition that a minimum occurs at this critical point is that the Hessian matrix of f is positive definite (compare this condition to the perhaps more familiar condition of a positive second derivative at the critical point of a function of a single variable).

But what is a Hessian matrix??

Let us recall the Taylor expansion of f around the critical point:

$$\begin{aligned} f(a,b) &= f(a^*,b^*)+\begin{pmatrix} f_a(a^*,b^*) & f_b(a^*,b^*)\end{pmatrix}\begin{pmatrix} a-a^* \\ b-b^*\end{pmatrix} \\ &\quad +\frac{1}{2!}\begin{pmatrix} a-a^* & b-b^*\end{pmatrix}\begin{pmatrix} f_{aa}(a^*,b^*) & f_{ab}(a^*,b^*) \\ f_{ba}(a^*,b^*) & f_{bb}(a^*,b^*)\end{pmatrix}\begin{pmatrix} a-a^* \\ b-b^*\end{pmatrix} \\ &= f(a^*,b^*)+0+\frac{1}{2!}\begin{pmatrix} a-a^* & b-b^*\end{pmatrix}\begin{pmatrix} f_{aa}(a^*,b^*) & f_{ab}(a^*,b^*) \\ f_{ba}(a^*,b^*) & f_{bb}(a^*,b^*)\end{pmatrix}\begin{pmatrix} a-a^* \\ b-b^*\end{pmatrix}\end{aligned}$$

The linear term vanishes because (a^*,b^*) is a critical point and therefore $f_a=f_b=0$ there. There is no term beyond the quadratic term because $f(a,b)$ is quadratic to begin with, but even if there was a higher order term present, this would not affect the discussion that follows. The last term in the expansion is a quadratic form, given by

$$z^THz,\quad \text{where } z=\begin{pmatrix} a-a^* \\ b-b^*\end{pmatrix} \text{ and } H=\begin{pmatrix} f_{aa}(a^*,b^*) & f_{ab}(a^*,b^*) \\ f_{ba}(a^*,b^*) & f_{bb}(a^*,b^*)\end{pmatrix},$$

The matrix H is called the Hessian matrix, and it is a symmetric matrix (i.e. its $(1,2)$ and $(2,1)$ entries are the same) because the order of the partial derivatives does not matter: $f_{ab}=f_{ba}$.

In general, we say that a quadratic form z^THz (and also the matrix H) is positive definite if $z^THz>0$ for all $z\neq 0$. But can we check if a given symmetric matrix is positive definite? Yes we can! Thanks to the famous Sylvester criterion which provides necessary and sufficient conditions for positive definiteness of a symmetric matrix. These conditions say that the determinants of the 1×1 most upper left submatrix, the 2×2 most upper left submatrix, all the way to the $n\times n$ matrix itself, should be positive. Applied to the matrix H in our case, we find that H is positive definite if and only if

$$\begin{aligned} f_{aa}(a^*,b^*) &= 2\sum x_i^2 \\ f_{aa}(a^*,b^*)f_{bb}(a^*,b^*)-(f_{ab}(a^*,b^*))^2 &= 4\left(n\sum x_i^2-\left(\sum x_i\right)^2\right)\end{aligned}$$

are both positive. Using the above claim once more, we see that this is indeed the case.

In summary, since the Hessian matrix at the critical point $\left(a^*,b^*\right)$ is positive definite, we have just shown that the function $f(a,b)$ does indeed achieve a minimum at this critical point.