

LEAST SQUARES AND THE CAUCHY SCHWARZ INEQUALITY

March 2008

1. INTRODUCTION

These notes are to fill in some additional details concerning least squares approximations, and to summarize the tools from sections 15.6 and 15.7 of the text which we are using to find local and global maxima and minima. These notes don't have to do with Lagrange multipliers; for this see section 15.8 of the text.

Unlike most of the homework problems, the least squares method is not a toy example. It's the basis for approximating data by linear functions, and this is crucial in physical science and engineering.

Because it is not a toy case, though, it requires more investment to understand. It also requires more investment to explain, which is the reason for these notes.

I have nothing against toy examples, since these are a good way to gain experience with mathematical techniques. So I'd strongly recommend that you try doing all the core homework problems for sections 15.6, 15.7 of the text, and that you discuss these in recitation.

I will focus in lecture on trying to discuss the most important examples, such as the least squares method.

2. THE LEAST SQUARE PROBLEM

Suppose $\{(x_i, y_i)\}_{i=1}^n$ is a set of n points in the $x - y$ plane. The goal of the least squares method is to find the linear function

$$h(x) = mx + b$$

which best approximates these points, in the sense that

$$(2.1) \quad f(m, b) = \sum_{i=1}^n (h(x_i) - y_i)^2 = \sum_{i=1}^n (mx_i + b - y_i)^2$$

is as small as possible.

As discussed in class, the first step is to find the critical points of f . These are where the gradient $\nabla f = (f_m, f_b)$ vanishes. This is because of the following fact.

Critical point rule: If f assumes a local max or min at a point (m, b) and $\nabla f = (f_m, f_b)$ is well defined at (m, b) , then (m, b) is a critical point of f in the sense that $\nabla f = (0, 0)$.

In our case,

$$f_m = \sum_{i=1}^n 2(mx_i + b - y_i)x_i = (2 \sum_{i=1}^n x_i^2)m + (2 \sum_{i=1}^n x_i)b - 2 \sum_{i=1}^n x_i y_i$$

and

$$f_b = \sum_{i=1}^n 2(mx_i + b - y_i) = (2 \sum_{i=1}^n x_i)m + 2nb - 2 \sum_{i=1}^n y_i$$

So $\nabla f = (f_m, f_b) = (0, 0)$ exactly when m and b are solutions of the two linear equations

$$(2.2) \quad Am + Bb = C$$

and

$$(2.3) \quad Em + Fb = G$$

where

$$A = 2 \sum_{i=1}^n x_i^2 \quad ; \quad B = 2 \sum_{i=1}^n x_i \quad ; \quad C = 2 \sum_{i=1}^n x_i y_i$$

$$E = 2 \sum_{i=1}^n x_i \quad ; \quad F = 2n \quad ; \quad G = 2 \sum_{i=1}^n y_i.$$

Example: Suppose $n = 3$ and that $(x_1, y_1) = (1, 0)$, $(x_2, y_2) = (1, 1)$, $(x_3, y_3) = (2, 0)$. Then

$$A = 12 \quad ; \quad B = 8 \quad ; \quad C = 2$$

$$E = 8 \quad ; \quad F = 6 \quad ; \quad G = 2.$$

The system of equations in (2.2) and (2.3) is

$$12m + 8b = 2 \quad ; \quad 8m + 6b = 2$$

This has unique solution $m = -1/2$ and $b = 1$. Try drawing the points $\{(x_i, y_i)\}_{i=1}^3$ and the graph of the best approximation $h(x) = -x/2 + 1$. How does this example compare to the one discussed in class?

3. THREE QUESTIONS ABOUT EQUATIONS (2.2) AND (2.3)

1. When is there a solution (m, b) of both equations, and when is this solution unique?
2. How does one determine whether a solution (m, b) to the two equations leads to a local minimum of $f(m, b)$?
3. How does one find which (m, b) give a global minimum for $f(m, b)$?

I'll put aside question (1) for the moment, and summarize the tools which apply to questions (2) and (3).

In class we discussed the second derivative test when $\nabla f = (0, 0)$. Recall that for this one needs to determine the sign of

$$(3.4) \quad D = f_{mm}f_{bb} - f_{mb}f_{bm}$$

$$(3.5) \quad = AF - BE = 4(n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2) \quad \text{in our case.}$$

Second Derivative Test: If $D < 0$ when the partials are evaluated at (a, b) , then (a, b) is a saddle point for f , and f does not have a local minima or maxima at (a, b) . If $D > 0$ and $f_{mm} > 0$ (or $D > 0$ and $f_{bb} > 0$), then f has a local minimum at (m, b) . If $D > 0$ and $f_{mm} < 0$ (or $D > 0$ and $f_{bb} < 0$), then f has a local maximum at (m, b) .

As far as global minima and maxima are concerned, the main tool is following fact described toward the end of §15.7 of the text:

Extreme Value Theorem: If f is continuous on a closed bounded set S of points (m, b) of \mathbb{R}^2 , then f attains an absolute maximum and an absolute minimum on S .

An example of a closed set is the closed unit square

$$T = \{(m, b) : 0 \leq m \leq 1 \text{ and } 0 \leq b \leq 1\}.$$

If one leaves out the boundary of this set, one gets the interior

$$T_0 = \{(m, b) : 0 < m < 1 \text{ and } 0 < b < 1\}$$

of the square, which is not closed.

The precise definition of a set S being closed is that if $(m_0, b_0) \in \mathbb{R}^2$ is a point which is not in S , then there is a small open disc around (m_0, b_0) which contains no point in S . Try seeing how this definition applies to T and T_0 . For example, T_0 is not closed because the point $(m_0, b_0) = (0, 0)$ is not in T_0 , but every small open disc around $(0, 0)$ does contain points of T_0 .

Warning: A function on a set which is not closed, or on an unbounded set, may have no absolute maximum value or no absolute minimum value, even if it does have local maxima or minima.

To see an example of this, just consider the function

$$F(m, b) = m^3 - m$$

for all $(m, b) \in \mathbb{R}^2$. The local maxima and minima of this function are at those (m, b) where m is a local max or min of $g(m) = m^3 - m$. Using $g'(m) = 3m^2 - 1$ and $g''(m) = 6m$ we see that there is a local minimum at $(1/\sqrt{3}, b)$ for all b and a local maximum at $(-1/\sqrt{3}, b)$ for all b . The second derivative test for functions of two variables is inconclusive at all (m, b) since

$$F_{mm}F_{bb} - F_{mb}F_{bm} = 0$$

for all (m, b) . However, there is no global max or min, since $\lim_{m \rightarrow \pm\infty} g(m) = \pm\infty$.

To deal with this issue, one can use the following fact:

Improved Extreme Value Theorem: Suppose that F is a continuous function on some set S , which may not be closed or bounded. Let S_1 be a closed bounded subset S_1 of S . A point (m, b) is an interior point of S_1 if there is a small open disc containing (m, b) which is contained entirely in S_1 . Suppose there is an interior point $(m_0, b_0) \in S_1$ so that

$$(3.6) \quad F(m_0, b_0) \leq F(m_1, b_1) \quad \text{for all } (m_1, b_1) \in S \text{ which are not interior points of } S_1.$$

Then F attains an absolute minimum on S , and this minimum occurs at at least one interior point of S_1 (which may be different than (m_0, b_0)). If ∇F is well defined at all the interior points of S_1 , then the absolute minimum occurs at a critical point of F in the interior of S_1 . The same statement holds for absolute maxima if \leq in (3.6) is replaced by \geq .

Here is an example; for another one, see §6

Example: Suppose $F(m, b) = m^4 + b^2$ and that $S = \{(m, b) : -1 \leq m \leq 1\}$. Then S is not bounded, since the values of b associated to points in S are not bounded. We try taking S_1 to be the square $\{(m, b) : -1 \leq m \leq 1, -1 \leq b \leq 1\}$. This is closed and bounded, and the point $(m_0, b_0) = (0, 0)$ is an interior point. We have $F(0, 0) = 0 \leq F(m, b)$ for all (m, b) which are not interior points of S_1 , and in fact F attains its absolute minimum at $(m_0, b_0) = (0, 0)$.

Proof of The Improved Extreme Value Theorem: Suppose (3.6) is true. Since S_1 is closed and bounded, F attains an absolute minimum on S_1 at some point $(m_2, b_2) \in S_1$ by the Extreme Value Theorem. Since $(m_0, b_0) \in S_1$, assumption (3.6) implies

$$(3.7) \quad f(m_2, b_2) \leq f(m_0, b_0) \leq f(m_1, b_1) \quad \text{for all } (m_1, b_1) \in S \text{ which are not interior points of } S_1.$$

Every point (m, b) of S is either not in S_1 , in which case $f(m_2, b_2) \leq f(m, b)$ because of (3.7), or is a point of S_1 , in which $f(m_2, b_2) \leq f(m, b)$ because the minimum of f over S_1 occurs at (m_2, b_2) . So $f(m_2, b_2)$ is the absolute minimum of f over all of S . If (m_2, b_2) is not an interior point of S_1 , (3.6) implies $f(m_0, b_0) \leq f(m_2, b_2)$, so (3.7) forces $f(m_0, b_0) = f(m_2, b_2)$. We can thus replace (m_2, b_2) by (m_0, b_0) ; in this way we can assume that the absolute minimum of f occurs at an interior point (m_2, b_2) of S_1 . Then there must be a local minimum at this point, so if ∇f is well defined at (m_2, b_2) it must be the zero vector by the Critical Value Test. The arguments about absolute maxima rather than absolute minima are similar.

4. APPLICATION OF THE SECOND DERIVATIVE TEST AND THE CAUCHY SCHWARZ INEQUALITY

I'm going to put aside question (1) of §2 for the moment and concentrate on question (2) concerning which (m, b) give local minima for the function $f(m, b)$ in equation (2.1).

Because of the Second Derivative Test in §3, we first need to determine the sign of

$$D = f_{mm}f_{bb} - f_{mb}f_{bm} = AF - BE = 4(n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2)$$

if (m, b) is a critical point of f . This is where the Cauchy Schwarz inequality comes in:

Theorem 4.1. (Cauchy-Schwarz) *Suppose $x = (x_1, \dots, x_n)$ and $w = (w_1, \dots, w_n)$ are n -tuples of real numbers. Recall that $x \cdot w = \sum_{i=1}^n x_i w_i$ is the dot product of x and w , and $|x|^2 = x \cdot x$ is the square of the length of x . Then*

$$(4.8) \quad (x \cdot w)^2 \leq (x \cdot x)(w \cdot w).$$

with equality if and only if w is the zero vector or x is a scalar multiple of w . In particular, if neither of x or w is the zero vector, then there is a unique $\theta \in [0, \pi]$ so that

$$(4.9) \quad \cos(\theta) = \frac{x \cdot w}{|x| |w|}$$

Before getting into the proof of this, note that equation (4.9) is familiar from the definition of the angle between two vectors in \mathbb{R}^2 or \mathbb{R}^3 . The point of Cauchy Schwarz is that there is a well defined angle between any two non-zero vectors in \mathbb{R}^n for all n .

Proof. Let's first see why (4.9) follows from (4.8). If (4.8) is true then

$$\left(\frac{x \cdot w}{|x| |w|} \right)^2 = \frac{(x \cdot w)^2}{(x \cdot x)(w \cdot w)} \leq 1$$

So

$$-1 \leq \frac{x \cdot w}{|x| |w|} \leq 1$$

and there will be a unique $\theta \in [0, \pi]$ for which (4.9) is true.

To prove (4.8), notice first that this is certainly true if w is the zero vector $(0, \dots, 0)$, since then both sides of (4.8) are 0. We now suppose that w is not the zero vector, so $w \cdot w > 0$. The idea is to write $x = u + aw$ for some $a \in \mathbb{R}$ and some vector u for which $u \cdot w = 0$. This is just decomposing x into the sum of a multiple of w together with a vector u which is perpendicular to w . To find what a and u should be notice that we want

$$(4.10) \quad 0 = u \cdot w = (x - aw) \cdot w = x \cdot w - a(w \cdot w).$$

Since $w \cdot w \neq 0$, we can let

$$a = \frac{x \cdot w}{w \cdot w} \quad \text{and} \quad u = x - aw$$

Then $x = u + aw$ and (4.10) shows $0 = u \cdot w$.

Now we compute

$$\begin{aligned} x \cdot w &= (u + aw) \cdot w = u \cdot w + a(w \cdot w) = a(w \cdot w) \\ x \cdot x &= (u + aw) \cdot (u + aw) = u \cdot u + 2a(u \cdot w) + a^2(w \cdot w) = u \cdot u + a^2(w \cdot w) \end{aligned}$$

where the last equalities follow from $w \cdot u = u \cdot w = 0$. So

$$(x \cdot w)^2 = a^2(w \cdot w)^2$$

and

$$(x \cdot x)(w \cdot w) = (u \cdot u + a^2(w \cdot w))(w \cdot w) \geq a^2(w \cdot w)^2$$

since $u \cdot u \geq 0$ and $w \cdot w \geq 0$. This proves

$$(x \cdot w)^2 \leq (x \cdot x)(w \cdot w)$$

and that equality holds if and only if $u \cdot u = 0$. But $u \cdot u = 0$ is true if and only if $u = x - aw$ is the zero vector, and this is the case if and only if $x = aw$ is a scalar multiple of w . \square

Corollary 4.2. *Suppose $w = (1, \dots, 1)$ has all its components equal to 1 and that $x = (x_1, \dots, x_n)$. Then*

$$(4.11) \quad (x \cdot w)^2 = \left(\sum_{i=1}^n x_i\right)^2 \leq (x \cdot x)(w \cdot w) = \left(\sum_{i=1}^n x_i^2\right) n.$$

In particular, the constant D in (3.4) always satisfies $D \geq 0$, and $D = 0$ if and only if all the x_i are equal.

Proof. Equation (4.11) is just (4.8) when we plug in $w = (1, \dots, 1)$. From this and (3.4) one gets

$$D/4 = \left(\sum_{i=1}^n x_i^2\right) \cdot n - \left(\sum_{i=1}^n x_i\right)^2 \geq 0.$$

If this is an equality, then Theorem 4.1 shows $x = (x_1, \dots, x_n)$ has to be a scalar multiple of $w = (1, \dots, 1)$. This is the case exactly when x_1, \dots, x_n are equal. \square

We can now say what the Second Derivative Test says about the least squares problem.

Theorem 4.3. *Suppose (m, b) is a solution to the equations (2.2) and (2.3), so that ∇f equals $(0, 0)$ at (m, b) . Then*

- a. *If x_1, \dots, x_n are not all equal, the constant D in (3.4) is positive and $f_{mm} = 2 \sum_{i=1}^n x_i^2 > 0$. So the second derivative test says that f has a local minimum at (m, b) .*
- b. *Suppose there is a constant c so that $x_1 = \dots = x_n = c$. Then D in (3.4) equals 0, so the Second Derivative Test is not conclusive.*

This illustrates how in analyzing where a function takes on minima and maxima, one often has to break the problem into cases.

5. WHEN IS THERE A UNIQUE (m, b) GIVING A LOCAL MINIMUM?

We first address when f has a unique critical point. By the computations in §2, this is the same as asking whether the system of equations (2.2) and (2.3) has a unique solution (m, b) .

The equations (2.2) and (2.3) define lines in the $m - b$ plane which have direction vectors

$$(A, B) = \left(2 \sum_{i=1}^n x_i^2, 2 \sum_{i=1}^n x_i\right)$$

and

$$(E, F) = \left(2 \sum_{i=1}^n x_i, 2n\right)$$

Notice that (E, F) is not the zero vector $(0, 0)$, since its second component is $2n \neq 0$.

The two lines defined by (2.2) and (2.3) will intersect in a unique point unless they are parallel. They will be parallel exactly when (A, B) is a scalar multiple of (E, F) .

So we have to decide whether there is a scalar λ so that

$$(5.12) \quad (A, B) = \lambda(E, F).$$

If such a λ exists, we have to have

$$B = 2 \sum_{i=1}^n x_i = \lambda F = \lambda 2n$$

so $\lambda = \sum_{i=1}^n x_i/n$. We would also have

$$A = 2 \sum_{i=1}^n x_i^2 = \lambda E = \left(\sum_{i=1}^n x_i/n \right) 2 \left(\sum_{i=1}^n x_i \right).$$

Multiplying this equality by n and then subtracting the right side from the left would give

$$D/2 = 2 \left(n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 \right) = 0.$$

Reversing these calculations then completes the proof of the following fact:

Theorem 5.1. *There is a unique solution (m, b) to (2.2) and (2.3) exactly when $D \neq 0$, and in this case (m, b) is the unique critical point of f . By Theorem 4.3, this occurs exactly when x_1, \dots, x_n are not all equal, and then (m, b) is the unique local minimum of f .*

We still have to deal with the case $D = 0$. By Theorem 4.3, if $D = 0$, then all of x_1, \dots, x_n have to be equal to some constant c . The equations (2.2) and (2.3) become

$$(5.13) \quad 2nc^2m + 2ncb = 2c \sum_{i=1}^n y_i$$

and

$$(5.14) \quad 2ncm + 2nb = 2 \sum_{i=1}^n y_i.$$

One sees from this that the first equation is just c times the second equation, so any solution (m, b) of the second equation will also be a solution of the first. Rewriting the second equation, we see that any (m, b) such that

$$h(c) = mc + b = \frac{1}{n} \sum_{i=1}^n y_i$$

will be a solution of both equations. Notice that while there are infinitely many (m, b) for which this is true, the values

$$h(c) = h(x_1) = \dots = h(x_n) = \frac{1}{n} \sum_{i=1}^n y_i$$

will all be equal to the average of the y_i 's, since $x_1 = x_2 = \dots = x_n$. Let's define

$$y_0 = \frac{1}{n} \sum_{i=1}^n y_i.$$

We have

$$f(m, b) = \sum_{i=1}^n (mc + b - y_i)^2 = \sum_{i=1}^n (y_0 - y_i)^2.$$

and notice that this, too, does not depend on the choice of a solution (m, b) to the equations which determine if f has a critical point at (m, b) . I'll now summarize these conclusions:

Theorem 5.2. *Suppose that x_1, \dots, x_n are all equal to a constant c . Then (m, b) is a critical point of f exactly when $mc + b = y_0 = \frac{1}{n} \sum_{i=1}^n y_i$ is the average of the y_i . At such an (m, b) , one has $D = 0$, and*

$$f(m, b) = \sum_{i=1}^n (y_0 - y_i)^2.$$

To decide whether f has a local minimum at the critical points in this case, we are going to have to use some other arguments in the next section. To get a feeling for this case, try setting $n = 3$ and drawing an example to illustrate the various lines associated to $h(x) = mx + b$ which give the least squares approximation. You will see that these lines are exactly the ones which go through the point (c, y_0) .

6. GLOBAL MINIMA FOR f .

We now turn to the issue of finding the global minimum of $f(m, b)$. From the definition

$$f(m, b) = \sum_{i=1}^n (mx_i + b - y_i)^2$$

one sees that f is a polynomial in m and b , so it is continuous for all m and b . However, the domain is all $(m, b) \in \mathbb{R}^2$, and $S = \mathbb{R}^2$ is not bounded. So we have to use the Improved Extreme Value Theorem mentioned at the end of §3 rather than the Extreme Value Theorem.

Suppose first that x_1, \dots, x_n are not all equal. We can reorder the points (x_i, y_i) so $x_1 \neq x_2$. Then

$$(6.15) \quad f(m, b) = \sum_{i=1}^n (mx_i + b - y_i)^2 \geq (mx_1 + b - y_1)^2 + (mx_2 + b - y_2)^2$$

since all the terms in the sum are ≥ 0 . For simplicity, let

$$(6.16) \quad z_1 = mx_1 + b - y_1 \quad ; \quad z_2 = mx_2 + b - y_2$$

so that (6.15) is

$$(6.17) \quad f(m, b) \geq z_1^2 + z_2^2$$

We can solve for m and b in terms of z_1 and z_2 :

$$(6.18) \quad z_1 - z_2 = m(x_1 - x_2) - (y_1 - y_2) \quad \text{so} \quad m = \frac{z_1 - z_2 + y_1 - y_2}{x_1 - x_2}$$

$$(6.19) \quad b = z_1 + y_1 - mx_1 = z_1 + y_1 - x_1 \cdot \frac{z_1 - z_2 + y_1 - y_2}{x_1 - x_2}$$

The numbers x_1, y_1, x_2, y_2 are constants. These formulas express m and b as linear functions of z_1 and z_2 . It follows that there is a constant $L > 0$ which depends on x_1, y_1, x_2, y_2 such that:

$$(6.20) \quad \max(|m|, |b|) < LT + L \quad \text{if} \quad \max(|z_1|, |z_2|) \leq T$$

We now want to use this in the Improved Extreme Value Test in §3. We let $(m_0, b_0) = (0, 0)$ in the test, so that $f(m_0, b_0)$ is some constant c_0 . We want to show that there is a closed bounded set S_1 in $S = \mathbb{R}^2$ so that if (m, b) is not an interior point of S_1 then

$$0 \leq c_0 = f(m_0, b_0) \leq f(m, b)$$

Suppose that (m, b) is a point for which

$$f(m, b) < c_0$$

From (6.17) we conclude that

$$z_1^2 + z_2^2 \leq f(m, b) < c_0$$

so

$$\max(|z_1|, |z_2|) < \sqrt{c_0}.$$

Setting $T = \sqrt{c_0}$ in (6.20), we get from this equation that

$$\max(|m|, |b|) < LT + L = L(\sqrt{c_0} + 1)$$

This means that (m, b) has to be an interior point of the closed bounded square

$$S_1 = \{(m, b) : |m| \leq L(\sqrt{c_0} + 1), |b| \leq L(\sqrt{c_0} + 1)\}$$

We've now shown that the conditions of the Improved Extreme Value Theorem hold, so f has an absolute minimum over all of $S = \mathbb{R}^2$, and this has to occur at a critical point of f inside the square S_1 above. In Theorems 5.1 we found all these critical points when x_1, \dots, x_n are not all equal, so we get the following final result in this case.

Theorem 6.1. *Suppose first that x_1, \dots, x_n are not all equal. Then there is a unique solution (m, b) of the system of equations (2.2) and (2.3). Fixing (m, b) to be this solution, the constant D in (3.4) is positive, and $f_{mm} = 2 \sum_{i=1}^n x_i^2 > 0$. Thus f in (2.1) has a local minimum at (m, b) , and this is in fact a global minimum. Thus in this case there is a unique optimal least squares function $h(x) = mx + b$ for the given data $\{(x_i, y_i)\}_{i=1}^n$.*

We now suppose that there is a constant c such that $x_1 = x_2 = \dots = x_n = c$. Then

$$(6.21) \quad f(m, b) = \sum_{i=1}^n (mc + b - y_i)^2 = \sum_{i=1}^n (h - y_i)^2$$

when $h = h(c)$ and $h(x) = mx + b$. Thus $f(m, b)$ is the value of the function

$$g(h) = \sum_{i=1}^n (h - y_i)^2$$

of a single variable h when we set $h = h(c) = mc + b$.

Treating h as a variable, we have

$$g'(h) = \sum_{i=1}^n 2(h - y_i) = 2nh - 2 \sum_{i=1}^n y_i$$

so $g'(h) = 0$ has the unique solution

$$h = \frac{1}{n} \sum_{i=1}^n y_i = y_0$$

in which y_0 is the average of $\{y_1, \dots, y_n\}$. One has

$$g''(h) = 2n > 0$$

so by the second derivative test for functions of one variable, $g(h)$ has a local minimum at $h = y_0$. Furthermore, $g'(h) < 0$ for $h < y_0$ and $g'(h) > 0$ for $h > y_0$. Thus $g(h)$ is decreasing as h increases when $h < y_0$, and $g(h)$ is increasing as h increases when $h > y_0$. This forces $g(h)$ to have an absolute minimum at $h = y_0$.

Recall now from Theorem 5.2 that when all of x_1, \dots, x_n equal c , the critical points (m, b) of f are those for which $h(c) = mc + b$ equals y_0 . We've just shown these are the (m, b) for which $f(mc + b) = h(mc + b)$ achieves its absolute minimum. Putting all this together gives this result:

Theorem 6.2. *Suppose there is a constant c such that x_1, \dots, x_n are all equal to c . Then equations (2.2) and (2.3) are equivalent to the single equation*

$$mc + b = \frac{1}{n} \sum_{i=1}^n y_i$$

This has infinitely many solutions (m, b) . The function $h(x) = mx + b$ associated to any one of these has $h(x_1) = h(x_2) = \dots = h(x_n)$ equal to the average $y_0 = \frac{1}{n} \sum_{i=1}^n y_i$ of the y_i . The value of $f(m, b)$ at any solution (m, b) as above is equal to $\sum_{i=1}^n (y_0 - y_i)^2$ and this is a global minimum for f .