

# Least squares

Sample Subtitle

Juan V. Vía

August 15, 2021

# Example

## Showing why least squares

We have a variable  $y$ . We know that it's dependent of another variable  $x$  in some way. But we don't know how, exactly. So we go to the field and measure certain points. Those that we can reach. Six of them.

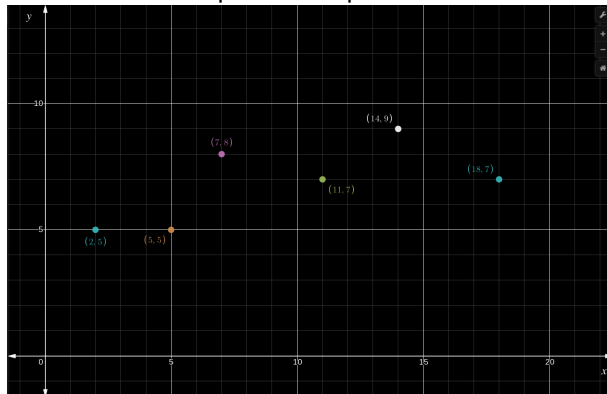
$$(2, 5), (5, 5), (7, 8), (11, 7), (14, 9), (18, 7)$$

That is: at  $x = 2$  we measure  $y = 5$ , at  $x = 5$  we measure  $y = 5$  again, but at  $x = 7$  we got  $y = 8$ , and so on.

# Example

Showing why least squares

Back to desk we plot these points.

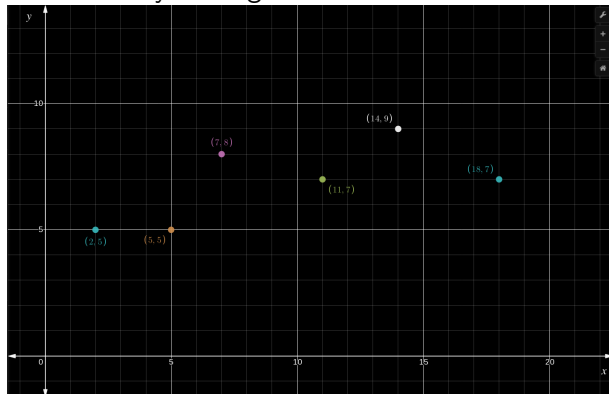


There is a model lurking in these points? A line perhaps?

# Example

## Showing why least squares

Let's start by tracing a line with "best fit" that data

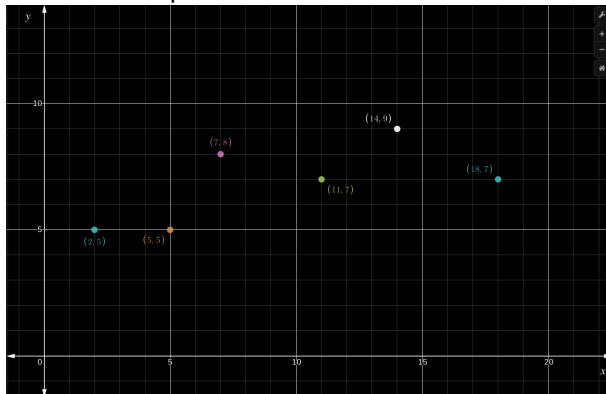


Why a line? To warm the modeling machine in our minds we are considering a line, a one degree polynomial.

# Example

Showing why least squares

What line to plot? How to choose that line?



Enter the math...

# Example

## Showing why least squares

A very common way to express a line in the plane is isolating  $y$  as a polynomial function of  $x$ :

$$y = mx + b$$

From this equation we can see that  $m$  is the *slope* of the line and  $b$  is the *y-intercept* (the value of  $y$  when  $x = 0$ ).

Will be handy for us to write this equation using vector notation:

$$y = \begin{bmatrix} x & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix}$$

And switching terms:

$$\begin{bmatrix} x & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = y$$

# Example

## Showing why least squares

Ok. Fine. This form

$$\begin{bmatrix} x & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = y$$

will be a useful one because we know  $x$  and  $y$  in six points. For example take the first point  $(2, 5)$

$$\begin{bmatrix} 2 & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = 5$$

# Example

## Showing why least squares

And yes, your guess is true. We can incorporate the second point (5, 5) and get

$$\begin{bmatrix} 2 & 1 \\ 5 & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} 5 \\ 5 \end{bmatrix}$$

Doing the product you can workout  $m$  and  $b$  from this equation.  $m = 0$  and  $b = 5$ . Because the matrix is square. And invertible by the way.



# Example

## Showing why least squares

Going further, incorporate the third point.

$$\begin{bmatrix} 2 & 1 \\ 5 & 1 \\ 7 & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} 5 \\ 5 \\ 8 \end{bmatrix}$$

This invalidate the previous result.  $m$  cannot be 0 more. Neither  $b$  with 5.

Actually, henceforth the equation is over-determined and have no solution.

# Example

Showing why least squares

Fourth point.

$$\begin{bmatrix} 2 & 1 \\ 5 & 1 \\ 7 & 1 \\ 11 & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} 5 \\ 5 \\ 8 \\ 7 \end{bmatrix}$$

# Example

Showing why least squares

Fifth point.

$$\begin{bmatrix} 2 & 1 \\ 5 & 1 \\ 7 & 1 \\ 11 & 1 \\ 14 & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} 5 \\ 5 \\ 8 \\ 7 \\ 9 \end{bmatrix}$$

# Example

Showing why least squares

Sixth and last point.

$$\begin{bmatrix} 2 & 1 \\ 5 & 1 \\ 7 & 1 \\ 11 & 1 \\ 14 & 1 \\ 18 & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} 5 \\ 5 \\ 8 \\ 7 \\ 9 \\ 7 \end{bmatrix}$$

This is the equation derived from our measurement.

# Example

Showing why least squares

Time to name things.

# Example

Showing why least squares

Call  $A$  to the matrix

$$A = \begin{bmatrix} 2 & 1 \\ 5 & 1 \\ 7 & 1 \\ 11 & 1 \\ 14 & 1 \\ 18 & 1 \end{bmatrix}$$

# Example

Showing why least squares

Call  $x$  to the column vector

$$x = \begin{bmatrix} m \\ b \end{bmatrix}$$

# Example

Showing why least squares

Call  $b$  to the column vector

$$b = \begin{bmatrix} 5 \\ 5 \\ 8 \\ 7 \\ 9 \\ 7 \end{bmatrix}$$



# Example

Showing why least squares

Thus, the equation

$$\begin{bmatrix} 2 & 1 \\ 5 & 1 \\ 7 & 1 \\ 11 & 1 \\ 14 & 1 \\ 18 & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} 5 \\ 5 \\ 8 \\ 7 \\ 9 \\ 7 \end{bmatrix}$$

becomes

$$Ax = b$$

# Example

## Showing why least squares

In other words: if you want to know the values  $m$  and  $b$  in the equation  $y = mx + b$  of the line we are searching find the  $x$  such  $Ax = b$ .

# Example

## Showing why least squares

In other words: if you want to know the values  $m$  and  $b$  in the equation  $y = mx + b$  of the line we are searching find the  $x$  such  $Ax = b$ , but **that equation have no solution!**

# Example

## Showing why least squares

In other words: if you want to know the values  $m$  and  $b$  in the equation  $y = mx + b$  of the line we are searching find the  $x$  such  $Ax = b$ , but **that equation have no solution!** because the matrix  $A$  is tall, so  $Ax = b$  is over-determined, there are more equations than variables to choose. There is no  $x$  such the norm of the difference between  $Ax$  and  $b$  equals zero

$$\|Ax - b\| \neq 0$$

# Example

## Showing why least squares

In other words: if you want to know the values  $m$  and  $b$  in the equation  $y = mx + b$  of the line we are searching find the  $x$  such  $Ax = b$ , but **that equation have no solution!** because the matrix  $A$  is tall, so  $Ax = b$  is over-determined, there are more equations than variables to choose. There is no  $x$  such the norm of the difference between  $Ax$  and  $b$  equals zero

$$\|Ax - b\| \neq 0$$

**We need the least squares approximate solution**

# Example

## Showing why least squares

In the least squares approximate solution of  $Ax = b$  we, provided that  $\|Ax - b\| \neq 0$ , select  $x$  that the *norm* of the *residual*  $r = Ax - b$  is minimal.

# Example

## Showing why least squares

In the least squares approximate solution of  $Ax = b$  we, provided that  $\|Ax - b\| \neq 0$ , select  $x$  that the *norm* of the *residual*  $r = Ax - b$  is minimal.

It's all there: <http://vmls-book.stanford.edu/> and in many other sources. The least squares approximate solution is very known.

# Example

## Showing why least squares

In the least squares approximate solution of  $Ax = b$  we, provided that  $\|Ax - b\| \neq 0$ , select  $x$  that the *norm* of the *residual*  $r = Ax - b$  is minimal.

It's all there: <http://vmls-book.stanford.edu/> and in many other sources. The least squares approximate solution is very known.

Let's see it.



# Example

Showing why least squares

All start with our equation  $Ax = b$ .

# Example

Showing why least squares

Take the original matrix  $A$ :

$$A = \begin{bmatrix} 2 & 1 \\ 5 & 1 \\ 7 & 1 \\ 11 & 1 \\ 14 & 1 \\ 18 & 1 \end{bmatrix}$$

# Example

Showing why least squares

Take the original matrix  $A$ :

$$A = \begin{bmatrix} 2 & 1 \\ 5 & 1 \\ 7 & 1 \\ 11 & 1 \\ 14 & 1 \\ 18 & 1 \end{bmatrix}$$

Transpose it, and get  $A^T$ :

$$A^T = \begin{bmatrix} 2 & 5 & 7 & 11 & 14 & 18 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

# Example

Showing why least squares

Multiply  $A^T$  by  $A$

$$A^T A = \begin{bmatrix} 2 & 5 & 7 & 11 & 14 & 18 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 5 & 1 \\ 7 & 1 \\ 11 & 1 \\ 14 & 1 \\ 18 & 1 \end{bmatrix} = \begin{bmatrix} 719 & 57 \\ 57 & 6 \end{bmatrix}$$

# Example

Showing why least squares

Invert the product

$$(A^T A)^{-1} = \begin{bmatrix} 719 & 57 \\ 57 & 6 \end{bmatrix}^{-1} = \begin{bmatrix} \frac{2}{355} & -\frac{19}{355} \\ -\frac{19}{355} & \frac{719}{1065} \end{bmatrix}$$

# Example

Showing why least squares

Multiply inverse and transpose

$$\begin{aligned}(A^T A)^{-1} A^T &= \begin{bmatrix} \frac{2}{355} & -\frac{19}{355} \\ -\frac{19}{355} & \frac{719}{1065} \end{bmatrix} \begin{bmatrix} 2 & 5 & 7 & 11 & 14 & 18 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} -\frac{3}{71} & -\frac{9}{355} & -\frac{1}{71} & \frac{3}{355} & \frac{9}{355} & \frac{17}{355} \\ \frac{121}{213} & \frac{434}{1065} & \frac{64}{213} & \frac{92}{1065} & -\frac{79}{1065} & -\frac{307}{1065} \end{bmatrix}\end{aligned}$$

Let's call this product  $(A^T A)^{-1} A^T$  the *pseudo-inverse*

# Example

Showing why least squares

Finally multiply the pseudo-inverse by  $b$

$$x = (A^T A)^{-1} A^T b \quad (1)$$

$$= \begin{bmatrix} -\frac{3}{71} & -\frac{9}{355} & -\frac{1}{71} & \frac{3}{355} & \frac{9}{355} & -\frac{17}{355} \\ \frac{121}{213} & \frac{434}{1065} & \frac{64}{213} & \frac{92}{1065} & -\frac{79}{1065} & -\frac{307}{1065} \end{bmatrix} \begin{bmatrix} 5 \\ 5 \\ 8 \\ 7 \\ 9 \\ 7 \end{bmatrix} \quad (2)$$

$$= \begin{bmatrix} \frac{61}{355} \\ \frac{5539}{1065} \end{bmatrix} \quad (3)$$

$$= \begin{bmatrix} 0.171831 \\ 5.20094 \end{bmatrix} \quad (4)$$

# Example

## Showing why least squares

That's it

$$x = \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} \frac{61}{355} \\ \frac{5539}{1065} \end{bmatrix} \text{ or } \begin{bmatrix} 0.171831 \\ 5.20094 \end{bmatrix}$$

and  $m = 0.171831$

and  $b = 5.20094$

and the line is

$$y = 0.171831x + 5.20094$$

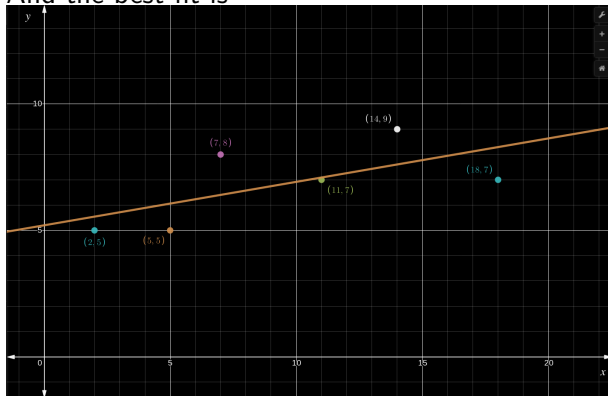
and **that** is the line which **best fit** our data points.



# Example

Showing why least squares

And the best fit is



# Least squares

## Recapitulation

You have a bunch of data points and suspect that a polynomial function  $y = Ax + B$  from this points is a good model.

You build a matrix  $A$  and a vector  $b$  from the polynomial and the points.

The polynomial coefficients are a vector  $x$  so  $Ax = b$

Then  $x = (A^T A)^{-1} A^T b$

# Least squares

Don't worry, its XXI century

So the key is calculate  $x = (A^T A)^{-1} A^T b$ .

But you don't need to do this by hand like Gauss did.

See the code, in this case Javascript code, using the mathjs package.

---

```
const { multiply, transpose, inv, matrix, format } =  
  require( "mathjs")
```

```
const solve = (A, b) => {  
  const transposed = transpose(A)  
  const product    = multiply(transposed, A)  
  const inverse     = inv(product)  
  const pseudoInverse = multiply(inverse,transposed)  
  const x           = multiply(pseudoInverse, b)  
  return x  
}
```

```
const A = matrix([[2,1],[5,1],[7,1],[11,1],[14,1],[18,1]])  
const b = matrix([5,5,8,7,9,7])
```

```
console.log(format(solve(A, b),5))
```

---

# Least squares

## The polynomial function

Get a terminal and do the magic

# Least squares

## The polynomial function

Get a terminal and do the magic

---

```
$ node ls.js  
[0.17183, 5.2009]
```

---

# Least squares

## Going further

Mind blowing is coming

To avert catastrophic damage in our brains it's time to rename things

In the example we found a line  $y = mx + b$ , now:

$m$  becomes  $a_1$

$b$  becomes  $a_0$

$x$  becomes  $x_1$

$y$  remains the same

The found line, then, becomes

$$y = a_1x_1 + a_0 = 0.17183x_1 + 5.2009$$

**Nothing has changed**, only the names.

# Least squares

## The polynomial function

In the example we picked a line for regression.

A line in a plane can be expressed as a polynomial function  $P(x_1)$  where  $y$  (the dependent variable, a real number) is a function of  $x_1$  (the independent variable, a real number)

$$y: \mathbb{R} \rightarrow \mathbb{R} = P(x_1) = a_1 x_1 + a_0$$

But there are others  $P(x_1)$ s, for example:

$$y: \mathbb{R} \rightarrow \mathbb{R} = P(x_1) = a_2 x_1^2 + a_1 x_1 + a_0$$

What about using that quadratic function as the model for the measured points?

Let's see...



# Least squares

## A curve as a model

That is the curve:  $y = a_2x_1^2 + a_1x_1 + a_0$

And those are the points:  $(2, 5), (5, 5), (7, 8), (11, 7), (14, 9), (18, 7)$

Replacing  $x_1$  and  $y$  in each point, and rearranging:

$$4a_2 + 2a_1 + a_0 = 5 \quad (5)$$

$$25a_2 + 5a_1 + a_0 = 5 \quad (6)$$

$$49a_2 + 7a_1 + a_0 = 8 \quad (7)$$

$$121a_2 + 11a_1 + a_0 = 7 \quad (8)$$

$$196a_2 + 14a_1 + a_0 = 9 \quad (9)$$

$$324a_2 + 18a_1 + a_0 = 7 \quad (10)$$

# Least squares

## A curve as a model

The curve:  $y = a_2x_1^2 + a_1x_1 + a_0$

In matrix form:

$$\begin{bmatrix} 4 & 2 & 1 \\ 25 & 5 & 1 \\ 49 & 7 & 1 \\ 121 & 11 & 1 \\ 196 & 14 & 1 \\ 324 & 18 & 1 \end{bmatrix} \begin{bmatrix} a_2 \\ a_1 \\ a_0 \end{bmatrix} = \begin{bmatrix} 5 \\ 5 \\ 8 \\ 7 \\ 9 \\ 7 \end{bmatrix}$$

In this case

$$x = \begin{bmatrix} -0.028205 \\ 0.73633 \\ 3.2181 \end{bmatrix}$$

and the curve is  $y = -0.028205x_1^2 + 0.73633x_1 + 3.2181$

# Least squares

## Adding one dimension

Until now we were working with polynomials with one independent variable ( $x_1$ ) and one dependent variable ( $y$ )

$$y: \mathbb{R} \rightarrow \mathbb{R} = P(x_1)$$

What about a polynomial with two independent variables?

Nothing, the least squares method stays the same. Note that is all about finding coefficients. So for example take

$$y: \mathbb{R}^2 \rightarrow \mathbb{R} = P(x_1, x_2)$$

where

$$y = a_5 x_1^2 + a_4 x_2^2 + a_3 x_1 x_2 + a_2 x_1 + a_1 x_2 + a_0$$

# Least squares

## Adding one dimension

That function ( $y = a_5x_1^2 + a_4x_2^2 + a_3x_1x_2 + a_2x_1 + a_1x_2 + a_0$ ) is useful when you need a model to represent height at a certain point given with two coordinates ( $x_1, x_2$ ).

In this case we are looking for six coefficients

$$x = \begin{bmatrix} a_5 \\ a_4 \\ a_3 \\ a_2 \\ a_1 \\ a_0 \end{bmatrix}$$

so we need at least six points ( $x_1, x_2, y$ ) to uniquely determinate those coefficients.

# Least squares

## Adding one dimension

If you have more points than six you form the matrix  $A$ , form the vector  $b$  as usual, and resolve  $x$  using the least squares approximate solution:

$$x = (A^T A)^{-1} A^T b$$

# Least squares

## Multidimensional input

Of course you can extend the application of least squares to problems where the input have more than two dimension

$$y: \mathbb{R}^n \rightarrow \mathbb{R} = P(x_1, x_2, \dots, x_n)$$

Take for example a set of measured temperatures in certain points (ten or more)  $(x_1, x_2, x_3, t)$  in the interior of a body. If we want a quadratic model then the polynomial will be:

$$y = f(x_1, x_2, x_3) : \mathbb{R}^3 \rightarrow \mathbb{R} \quad (11)$$

$$y = a_9 x_1^2 + a_8 x_2^2 + a_7 x_3^2 + a_6 x_1 x_2 + a_5 x_1 x_3 + \quad (12)$$

$$a_4 x_2 x_3 + a_3 x_1 + a_2 x_2 + a_1 x_3 + a_0 \quad (13)$$

# Least squares

## Multidimensional input

Take the measurements, build the matrix  $A$  evaluating  $x_1, x_2, x_3$ , their products and their squares, build the column vector  $b$  evaluating  $y$ . Be aware that  $x$  (in  $Ax = b$ ) will be

$$x = \begin{bmatrix} a_9 \\ a_8 \\ a_7 \\ a_6 \\ a_5 \\ a_4 \\ a_3 \\ a_2 \\ a_1 \\ a_0 \end{bmatrix}$$

evaluate  $x = (A^T A)^{-1} A^T b$  and you are done.

# Least squares

## Multidimensional output

What about  $\mathbb{R}^n \rightarrow \mathbb{R}^m$ ?

Almost the same, with certain precautions.

Take the case of  $y : \mathbb{R}^3 \rightarrow \mathbb{R}^3$

where

$$y_1 : \mathbb{R}^3 \rightarrow \mathbb{R} = P_1(x_1, x_2, x_3) \quad (14)$$

$$y_2 : \mathbb{R}^3 \rightarrow \mathbb{R} = P_2(x_1, x_2, x_3) \quad (15)$$

$$y_3 : \mathbb{R}^3 \rightarrow \mathbb{R} = P_3(x_1, x_2, x_3) \quad (16)$$



# Least squares

## Multidimensional output

We can isolate and resolve each function e.g., taking the first coordinate of  $y(y_1)$  :

from the measurements and the selected polynomial model form the matrix  $A_1$  so:

$$A_1 x = b_1$$

$$x = (A_1^T A_1)^{-1} A_1^T b_1$$

$x$  here denotes  $[a_{11} \ a_{21} \ \dots \ a_{n1}]^T$  (the second subscript indicates that those are the coefficients of  $P_1$ , the polynomial of  $y_1$ )

Doing the same for  $y_2$  and  $y_3$  the model for the transformation  $y : \mathbb{R}^3 \rightarrow \mathbb{R}^3$  is complete.

But these results came from isolated functions. What if I want a least squares solution for the transformation as a whole?

Let's see it.

# Least squares

## Multidimensional output

What if I want a least squares solution for the transformation as a whole?

Instead of calculate

$$(A_1^T A_1)^{-1} A_1^T b_1 \text{ for } P_1$$

$$(A_2^T A_2)^{-1} A_2^T b_2 \text{ for } P_2$$

$$(A_3^T A_3)^{-1} A_3^T b_3 \text{ for } P_3$$

We can integrate the evaluation of the column vector  $x$  (the coefficients) using the following calculation:

$$x = (A_1^T A_1 + A_2^T A_2 + A_3^T A_3)^{-1} (A_1^T b_1 + A_2^T b_2 + A_3^T b_3)$$

**And that is the way to map, for example, from one color set to another given a set of example points.**