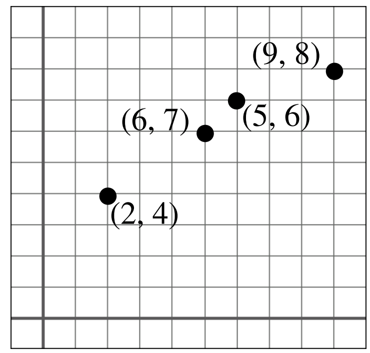
Calculus 120, section 7.5 Least Squares

notes prepared by Tim Pilachowski

So far, we have dealt with equations, and found both derivatives and integrals of those functions. In the real world, the equations didn't fall out of the sky but were developed from data—observations made about phenomenon. The curve is found that fits the data.

Hopefully you remember how to find the equation of a line given two points: calculate slope (m), substitute to find the y-coordinate of the y-intercept (b), and write the equation ($y = mx + b$). When there are more than two points, and they don't conveniently line up for us, we use partial derivatives to minimize the difference (error) between the observed data and the *line of best fit*. The process is called *regression analysis* and the method is called *least squares*. We'll adopt the statistics convention and use the formula $y = Ax + B$ for the line of regression.



Example A: A manufacturer has collected preliminary data relating number of units produced (x, measured in hundreds) and cost (y, measured in \$1000's). Find the equation that best represents cost as a function of number of units produced. *Answer*: $\hat{y} = 0.58x + 3.06$

The data are pictured in the scatterplot to the right. The line of regression will go "through the middle", but the question becomes, "Of all the lines that we might draw that seem to fit the data, which is the one that has the least error between the observed (actual) y-value and the regression's (predicted) $\hat{y}$-value = $Ax + B$?"

x_i	observed y_i	regression $\hat{y}$	error E_i	$(E_i)^2$
2	4			
5	6			
6	7			
9	8			

Since some of the points will lie above the regression line, and some will lie below it, some of the errors will be positive and some will be negative. Finding the sum of the errors would result in a "canceling" effect. Instead, to keep the value of each error and retain its effect we'll find the sum of the squared errors.

This is the function for which we want a minimum. We'll use the techniques of the previous sections: partial derivatives. We'll need the chain rule.

The process above involved only four data points. If we had 400, or 4000, the same process would become very unwieldy very quickly. We can develop a general formula which can be applied to N = any number of points.

x_i	observed y_i	regression $\hat{y}$	error E_i	$(E_i)^2$
x_1	y_1	$Ax_1 + B$	$y_1 - Ax_1 - B$	$(y_1 - Ax_1 - B)^2$
x_2	y_2	$Ax_2 + B$	$y_2 - Ax_2 - B$	$(y_2 - Ax_2 - B)^2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_N	y_N	$Ax_N + B$	$y_N - Ax_N - B$	$(y_N - Ax_N - B)^2$

Using the symbol $\sum$ to mean "the sum of", the least-squares error function we want to minimize is

$$\text{sum of squared errors} = f(A, B) = \sum (y - Ax - B)^2$$

Note: I really *should* write $f(A, B) = \sum_{i=1}^{N} (y_i - Ax_i - B)^2$, but chose the version above for simplicity's sake.

Using the sum rule and chain rule, we get

$$\frac{\partial f}{\partial A} = \sum 2(y - Ax - B)(-x) = \sum 2(-xy + Ax^2 + Bx) = 0$$

$$\frac{\partial f}{\partial B} = \sum 2(y - Ax - B)(-1) = \sum 2(-y + Ax + B) = 0$$

Note that since we have N points and thus have N terms in our sum, we can replace $\sum B$ with NB.

Dividing by two to get easier numbers and rearranging gives us

$$\begin{cases} -\sum xy + A\sum x^2 + B\sum x = 0 \\ -\sum y + A\sum x + NB = 0 \end{cases} \rightarrow \begin{cases} A\sum x^2 + B\sum x = \sum xy \\ A\sum x + NB = \sum y \end{cases}$$

Solving the second equations for B we get $B = \dfrac{\sum y - A * \sum x}{N}$.

Substituting into the first equation and solving for A we get $A = \dfrac{N * \sum xy - \sum x * \sum y}{N * \sum x^2 - \left(\sum x\right)^2}$

Example A revisited: Use the formulas above to find the least-squares regression equation.

	x	y	xy	x^2
	2	4		
	5	6		
	6	7		
	9	8		
$\Sigma =$				