Part 2

1

But then we will go on to see how to do this

By linear descent.

Then we'll find a better algorithm, and

then we will explore how to do this sort

of problem, where it is not so easy

to do explicitly.

So if we differentiate the 1st row w.r.t $m$

then the first thing to worry about is all the

sums over the data items $i$

$$\nabla \chi^2 = \begin{bmatrix} \frac{\partial \chi^2}{\partial m} \\ \frac{\partial \chi^2}{\partial c} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -2\sum_i x_i(y_i - mx_i - c) \\ -2\sum_i (y_i - mx_i - c \end{bmatrix}$$

But actually it turns out that we don't

need to worry about these sums, because

we're not differentiating the $x_i$ and $y_i$'s

themselves

$mx_i + mx_2$ ....

is the same as if we differentiate w.r.t. m we'll just get ~~$x$~~ $x_1 + x_2 + x_3$ ....

So we don't have to worry about those sums

Then its easy right, differentiate a square, that drops its power by 1, and we multiply by 2

and then we take the differential of the inside bracket, w.r.t. m, is $-x_i$

we can take the that $-2$ outside the sum altogether in fact.

For the second row, it's easier, cause the differential w.r.t. C is just $-1$.

So we just get the 2 down from power, ... and minus sign. and all looks quite easy.

Keeping on looking at second row, ~~and~~ then sum of C times number data items, we can take out of the sum altogether, and then just got the sum of $y_i$, and sum of m times $x_i$ and if we divide that by data items we get result that C ....

$$\therefore\ y = y(x_i; a_i) = mx_i + C$$

$$\chi^2 = \sum_i r_i^2 = \sum_i (y_i - mx_i - c)^2$$

$$\nabla \chi^2 = \begin{bmatrix} \frac{\partial \chi^2}{\partial m} \\ \frac{\partial \chi^2}{\partial c} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -2\sum_i x_i (y_i - mx_i - c) \\ -2\sum_i (y_i - mx_i - c) \end{bmatrix}$$

$$C = \bar{y} - m\bar{x}$$

with $\bar{y}$ and $\bar{x}$ been the average

we can carry on in that way, and generate an answer from, and show maths here don't need to show it blow by blow

$$(c)\quad \sigma_c \simeq \sigma_m \sqrt{\bar{x}^2 + \frac{1}{n} \sum_i (x-\bar{x})^2}$$

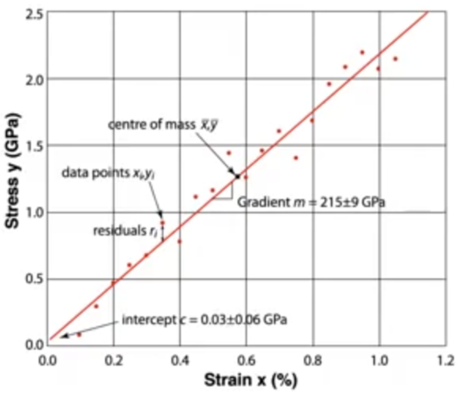
$$m = \frac{\sum(x-\bar{x})y}{\sum(x-\bar{x}y)^2} \qquad \sigma_m^2 \simeq \frac{\chi^2}{\sum(x-\bar{x})^2(n-2)}$$

It a bit tricky to see

We can also find the estimate for the uncertainties, called $\sigma_c$ and $\sigma_m$

Its very important when you do a fit to get an idea of the uncertainties in these fitting parameters and quote those in your fit (:

Coming back to our fitted data, we'll pull it out again see pics




$$c = \bar{y} - m\bar{x} \qquad \sigma_c \simeq \sigma_m \sqrt{\bar{x}^2 + \frac{1}{n}\sum_i (x-\bar{x})^2}$$

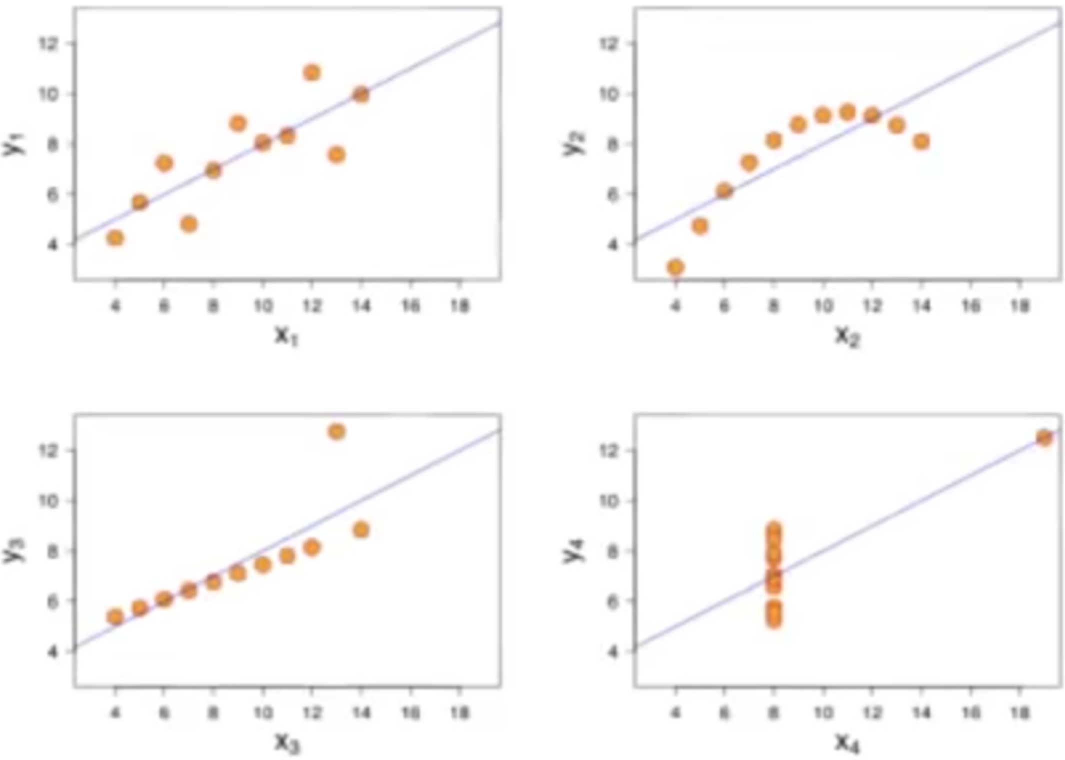
$$m = \frac{\sum(x-\bar{x})y}{\sum(x-\bar{x})^2} \qquad \sigma_m^2 \simeq \frac{\chi^2}{\sum(x-\bar{x})^2(n-2)}$$

The amazing thing is how accurate this sort of thing is, it's really cool.

We have quite noisy data, but a gradient of 215, the uncertainty is only 9, which is 5%, really amazing.

You should always plot your fit and visually compare it, as a sanity check, we can see why this is a good idea here.

This is Anscombe's quartet, see (pic) wikipedia




**Anscombe's quartet**

all these 4 data sets, have the same $\chi^2$,
means, best fit lines and uncertainties in fitting parameters
But obviously have different data
In right hand two cases, probably fitting a
straight line is wrong thing to do
Bottom left, if you remove flying data point, the
gradient different, and intercept, its only
the top left that fits doing the right thing
altogether.

And there's another subtlety if we go back
and look at c intercept, we can see that the
intercept, depends on gradient m
So what we said earlier, when we looked
at plot of $\chi^2$

So there's a way to re-cast the problem, which
is to look at the deviations ~~of~~ from centre of
mass of data $\bar{x}$ instead.

and then the intercept $b$, rather than $c$, is the location of the centre of mass in $y$, $\bar{y}$

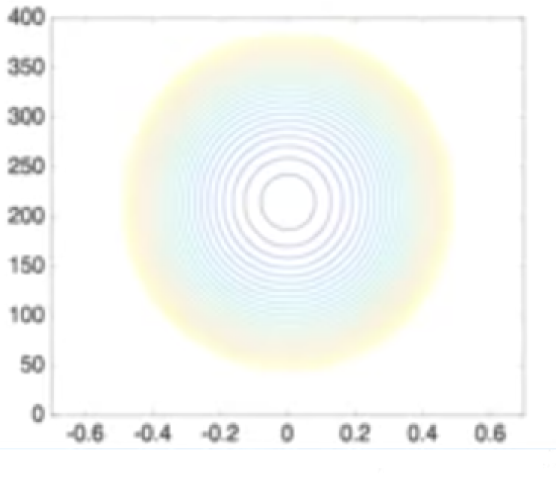
and then the constant term in fit $b$, that constant $b$, doesn't depend on the gradient any more and neither does the uncertainty include the term from the uncertainty $m$

In fact if I plot out the contour plot $\chi^2$ and when I do that I find, that it's got stanked, it a nice circular looking thing (pic)

$$y = (m \pm \sigma_m)(x - \bar{x}) + (b \pm \sigma_b)$$

$$m = \frac{\sum (x-\bar{x}) y}{\sum (x-\bar{x})^2} \qquad \sigma_m^2 \simeq \frac{\chi^2}{\sum (x-\bar{x})^2 (n-2)}$$

$$b = \bar{y} \qquad \sigma_b^2 \simeq \frac{\chi^2}{n(n-2)}$$



7

So, I have removed the interaction between m and Constant term

So, its mathematically more reasonable, well posed/stated problem

So, that's essence of regression, of how to fit a line to some data

and this is a totally really useful life skill whatever your professional job

What we will do in next few sessions, is look at how to do this in more complicated cases, with more complicated functions

and how to extend the idea of regression to those cases

The main thing, really, that we have defined here that is important to remember, is that goodness of fit of the estimated $\chi^2$ the sum of squares, the deviations of fit from data.

and X2 is going to be really usefull to us going forward.

8