Module 6

General non linear least squares

In this session we going to look at how to finally learn how to fit an distribution of heights data

Then in following exercise I am going to do it myself.

Then in next session, we'll wrap up and see how to do it practically in matlab and python

So we looking at how to fit a function that is arbitrarily complicated, compared to the simplest case of LA. $y = mx + c$ (that we looked at last time)

Of course, there is intermediate possibilities between very complicated and simple as possible

But this gives you the general case.

Before how we move on to look at a computer using tools instead of writing our own.

So lets say we have a function y, of some variable x, and that function has parameters a_k, where k goes from $1 \ldots\ldots m$

eg $y(x; a_k) = (x-a_1)^2 + a_2$

This function is not linear in a_1; in double a_1 I do not double the function

So its a non linearly least squares we going to do.

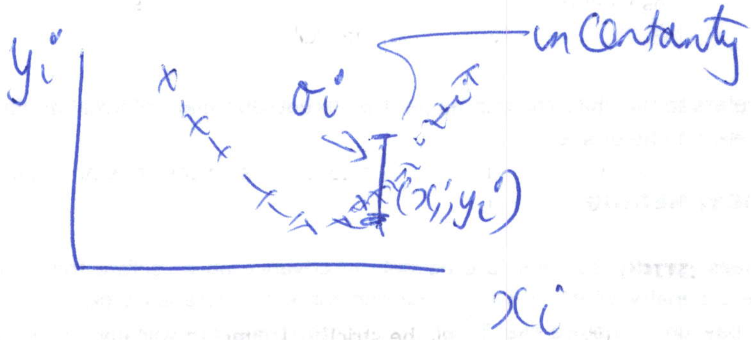
Say I want to fit the parameters a_k to some data

$i = 1 \ldots\ldots n$

(y_i, x_i, σ_i)

and I have pairs of data y_i, and x_i, so for every x, I have y.

and I have an associated uncertainty, σ_i
that is the more uncertain I am about data point y_i, the bigger the uncertainty σ_i will be
I can sketch it out, something like:



Then I am going to define a goodness of fit parameter χ^2 as being the sum of all data points i, of difference y_i and model of x_i with its parameter a_k

$$\chi^2 = \sum_{i=1}^{n} \frac{\left[y_i - y(x_i; a_k)\right]^2}{\sigma^2}$$

and divide all that by σ^2, and take squares of difference

So what I am doing here, is penalizing each difference by the uncertainty σ^2

So, the uncertainty data point have low weight in my sum of χ^2, so they dont affect the fit too much.

If we do not know what the σ are we can assign them all to be 1, and ~~just~~ they just drop out.

But if we have an idea about the uncertainty, then this gives us a way to include it

And my minima for χ^2 is when grad for χ^2 is equal to 0

In general case I may be able to write down an expression for grad like, but I may not be able to solve it algebraically

$$\nabla \chi^2 = 0$$

So instead I am going to look to solve:

$$\boxed{\nabla \chi^2 = 0}$$

By steepest descent, going down the contours simply by updating the vector of fitting ~~parameters~~ a parameters a

So I have my vector a; I am going to say that my next iteration is going to be my current iteration, minus some constant times the grad of χ^2

Ⓐ $$a_{next} = a_{cur} - cst \, \nabla \chi^2$$

Ⓐ

So I am going down the gradient Ⓐ here by an amount (cst) given by the constant (going down steepest descent), then I going to make my next guess what fitting parameters should be and write it down as vector.

And I'll keep doing it until I reach this criteria: $\nabla \chi^2 = 0$, which means I found minima, or flattening that.

until χ^2 stops changing, which should be the same thing or I just give up, so many iterations, so get bored, something has gone wrong

So to do the grad, I need to do:

$$\frac{d\chi^2}{da_k} = \sum_{i=1}^{n} -2 \frac{y_i - \cancel{y_i} \, y[x_i; a_k]}{\sigma^2} \frac{dy}{da_k}$$

(for each k's)

this will be my differential, -2 we can just take out.

So now we need to update (page 5) ①

$$\therefore a_{next} = a_{cur} - cst \, \nabla \chi^2$$

$$= a_{cur} + \sum_{i=1}^{n} \frac{[y_i - y[x_i; a_k]]}{\sigma^2} \frac{dy}{da_k}$$

evaluated at a_{curr}

(as I do not know a_{next} yet)

Lets use the formula for an example

lets differentiate wrt a_1

$$y(x; a_k) = (x - a_1)^2 + a_2$$

$$\therefore \frac{dy}{da_1} = -2(x - a_1) \qquad \frac{dy}{da_2} = 1$$

So thats the steepest descent formula for fitting a nonlinear function

When we try to minimise the sum of squares of residuals

This is therefor called nonlinear least squares fitting

There are lots of more complicated methods than steepest descent for solving these sorts of problems, which will look at later,

But first give this a go, and code it up for sandpit problem

8

So that's the simplest ~~form~~ version ~~for doing a general~~ ~~general~~ of how to do a general fitting finding a minimum ~~of~~ or least value of squares of residuals for a model that's not linear in both function and fitting parameters, that's called GENERALISED NON LINEARISED LEAST ~~SQUARES~~ SQUARES FITTING.