
Subject: Re: Newbie's question

Posted by [Paolo Grigis](#) on Thu, 20 Oct 2005 07:33:33 GMT

[View Forum Message](#) <> [Reply to Message](#)

ChiChiRuiz@gmail.com wrote:

> Hi there,

>

> I have a scatter plot which has the shape of a parabola, like $y=x^2$.

> I want to find the best curve fit to the scatter plot, so I used the

> function "curvefit" with no weights and with initial guesses (1.0, 2.0)

> i.e. $y = 1.*x^{(2.)}$. So, here's the problem...when I use only the right

> half of the data points (i.e. x and y values are positive), I get the

> curvefit returns parameter (0.5, 1.5), which means, the best fit curve

> is $y=.5*x^{(1.75)}$. I know the fit should be symmetric, so the same curve

> SHOULD fit the other half. Now onto the left half side of the data

> set, curvefit does not work anymore, and here's why, $x^{(1.5)}=x^{(3/2)}$

> and when x is a negative number, IDL returns "NaN" because it can't

> take the square root of a negative number, hence the entire procedure

> will not work. I ended up having to throw away half of my data points,

> and I'm not very comfortable with that. Any idea how to go around it

> or suggest another function to do the same thing?

What about fitting the data to the function $a*abs(x)^b$?

Paolo

>

> Besides, I've thought about using "polyfit", but if I remember

> correctly, polyfit only takes in one x value vs. one y value. Scatter

> plot has one x value vs. several y values. I don't think it'll

> work in my case, but I may be wrong...

>

> TIA (thanks in advance)

>

> Angie

>

Subject: Re: Newbie's question

Posted by [peter.albert@gmx.de](#) on Thu, 20 Oct 2005 11:02:10 GMT

[View Forum Message](#) <> [Reply to Message](#)

>

> Besides, I've thought about using "polyfit", but if I remember

> correctly, polyfit only takes in one x value vs. one y value. Scatter

> plot has one x value vs. several y values. I don't think it'll

> work in my case, but I may be wrong...

Hi Angie,

are you sure you do have more y than x values in your data arrays, or do they just appear like that in the scatter plot, because you have many identical x values? Besides, if you have more y than x values, I wonder how you actually do the scatter plot. And, well, you used CURVEFIT, so I guess you actually do have all the appropriate data points. In this case, you should just give POLY_FIT a try. Don't bother about y values scattering for one and the same x value. That's just what curve fitting is about, isn't it?

Cheers,

Peter

>
> TIA (thanks in advance)
>
> Angie

Subject: Re: Newbie's question

Posted by [Paul Van Delst\[1\]](#) on Thu, 20 Oct 2005 13:44:44 GMT

[View Forum Message](#) <> [Reply to Message](#)

ChiChiRuiz@gmail.com wrote:

> Hi there,
>
> I have a scatter plot which has the shape of a parabola, like $y=x^2$.
> I want to find the best curve fit to the scatter plot, so I used the
> function "curvefit" with no weights and with initial guesses (1.0, 2.0)
> i.e. $y = 1.*x^{(2.)}$. So, here's the problem...when I use only the right
> half of the data points (i.e. x and y values are positive), I get the
> curvefit returns parameter (0.5, 1.5), which means, the best fit curve
> is $y=.5*x^{(1.75)}$. I know the fit should be symmetric, so the same curve
> SHOULD fit the other half. Now onto the left half side of the data
> set, curvefit does not work anymore, and here's why, $x^{(1.5)}=x^{(3/2)}$
> and when x is a negative number, IDL returns "NaN" because it can't
> take the square root of a negative number, hence the entire procedure
> will not work. I ended up having to throw away half of my data points,
> and I'm not very comfortable with that. Any idea how to go around it
> or suggest another function to do the same thing?

Try Craig Markwardt's MPFIT suite (google will find it). It is a much more robust curve fitter than IDL's CURVEFIT.

cheers,

paulv

--

Paul van Delst
CIMSS @ NOAA/NCEP/EMC

Subject: Re: Newbie's question

Posted by ChiChiRuiz@gmail.com on Thu, 20 Oct 2005 13:53:52 GMT

[View Forum Message](#) <> [Reply to Message](#)

Thank you for your inputs. I'll try poly_fit and find MPFIT!

Subject: Re: Newbie's question

Posted by [Paul Van Delst\[1\]](#) on Thu, 20 Oct 2005 14:10:27 GMT

[View Forum Message](#) <> [Reply to Message](#)

ChiChiRuiz@gmail.com wrote:

> Thank you for your inputs. I'll try poly_fit and find MPFIT!

BTW, the MPFIT code has a drop in replacement for CURVEFIT -- I think it's called MPCURVEFIT. Anyway, another plus is that it's much much much faster than IDL's CURVEFIT -- at least it was for the fits I was performing:

$z(x,y) = c0(y) + c1(y).x^{c2(y)} + c3(y).x^{c4(y)}$ (where $c4 \sim 2*c2$)

I fit 1000's of channels of data in a flash with Craig's MP Code. CURVEFIT didn't converge most of the time, and when it did it took forever (i.e. go and get a coffee, come back, and it would still be working on a single channel).

cheers,

paulv

--

Paul van Delst
CIMSS @ NOAA/NCEP/EMC

Subject: Re: Newbie's question

Posted by [Paolo Grigis](#) on Thu, 20 Oct 2005 15:10:41 GMT

[View Forum Message](#) <> [Reply to Message](#)

While it is surely a very good thing to use MPFIT instead of CURVEFIT, this won't help solving your NaN problem, since you were applying a "bad" function on your negative data.

Change your model function to something which gives real results also for negative inputs, and then you should

be able to successfully fit your data (no matter what fitting routine you are using).

Ciao,
Paolo

ChiChiRuiz@gmail.com wrote:

> Thank you for your inputs. I'll try poly_fit and find MPFIT!
>

Subject: Re: Newbie's question

Posted by [James Kuyper](#) on Thu, 20 Oct 2005 18:41:08 GMT

[View Forum Message](#) <> [Reply to Message](#)

ChiChiRuiz@gmail.com wrote:

> Hi there,
>
> I have a scatter plot which has the shape of a parabola, like $y=x^2$.
> I want to find the best curve fit to the scatter plot, so I used the
> function "curvefit" with no weights and with initial guesses (1.0, 2.0)
> i.e. $y = 1.*x^{(2.)}$. So, here's the problem...when I use only the right
> half of the data points (i.e. x and y values are positive), I get the
> curvefit returns parameter (0.5, 1.5), which means, the best fit curve
> is $y=.5*x^{(1.75)}$. I know the fit should be symmetric, so the same curve
> SHOULD fit the other half. Now onto the left half side of the data
> set, curvefit does not work anymore, and here's why, $x^{(1.5)}=x^{(3/2)}$
> and when x is a negative number, IDL returns "NaN" because it can't
> take the square root of a negative number, hence the entire procedure
> will not work. I ended up having to throw away half of my data points,
> and I'm not very comfortable with that. Any idea how to go around it
> or suggest another function to do the same thing?

The fundamental problem is that curve fitting routines generally require that the dependent variable is a well-defined and continuous function of the curve's parameters. x^a is well-defined for negative numbers, only if it is treated as a complex-valued expression. It's continuous only if you use an unconventional branch cut, one that doesn't run along the negative real axis. If you have no idea what a branch cut is, you shouldn't even be attempting to do a fit of this type.

That's just a symptom of a deeper and simpler problem: you should try to fit data to a function, unless you have an understanding of the data that suggests that a function of that type is to be expected.

Of course, sometimes you have to fit the data without having any

theoretical basis for the fit. As long as you have reason to believe that the dependent variable is a sufficiently continuous function of the independent variables, you can usually fit it to a polynomial series ("sufficiently" and "usually" are weasel words to cover many different complicated issues that would require a small book to explain them properly).

There's many different polynomial series you can fit to - the general rule is that if you use a sufficiently large number of terms to fit your data, the remaining error in the fit will be dominated by a term proportional to the first term in the series that you didn't use. For instance, in a simple power series, if you fit $y = a + b*x + c*x^2$, then the first term you left out is x^3 , so you should expect the errors to be roughly proportional to x^3 ; they'll be smallest near $x == 0$. Similarly, if you fit to shifted power series like $y = a + b*(x-x_0) + c*(x-x_0)^2$, where x_0 is fixed, then the first term you left out was $(x-x_0)^3$. Therefore, your errors will tend to be smallest near $x == x_0$.

- > Besides, I've thought about using "polyfit", but if I remember
- > correctly, polyfit only takes in one x value vs. one y value. Scatter
- > plot has one x value vs. several y values. I don't think it'll
- > work in my case, but I may be wrong...

POLY_FIT is a suitable routine for performing such a fit. I don't understand what you're saying about why you don't think you can use it, but your reason sounds incorrect. You normally send polyfit a complete set of x values, and a complete set of corresponding y values.

```
x = (INDGEN(32)-16)/16.0
y = (x-2.0)*x*(x+2.0); Cubic function
```

```
fit = POLY_FIT(x,y,2,YFIT=yfit); Quadratic fit
plot,x,y,psym=2
oplot,x,yfit; Fairly good fit of quadratic curve to cubic data.
plot,x,yfit-y;
```

Subject: Re: Newbie's question
Posted by ChiChiRuiz@gmail.com on Thu, 20 Oct 2005 20:18:52 GMT
[View Forum Message](#) <> [Reply to Message](#)

Poly_fit doesn't really give me what I need. I don't need the coefficients of a quadratic equation, I want to know the best fit of the scatter plot to some power of x. I know it's not exactly power square, but it should be in that neighborhood. Even if I shift all data to the positive axis, i.e. $y = a*(x-x_0)^b$, any x values less than x_0 is still considered "negative". I don't know what else...maybe I'll try change of variable or something... thank you for your help.

Subject: Re: Newbie's question
Posted by [James Kuyper](#) on Thu, 20 Oct 2005 22:08:33 GMT
[View Forum Message](#) <> [Reply to Message](#)

ChiChiRuiz@gmail.com wrote:

> Poly_fit doesn't really give me what I need. I don't need the
> coefficients of a quadratic equation, I want to know the best fit of
> the scatter plot to some power of x. I know it's not exactly power
> square, but it should be in that neighborhood. Even if I shift all
> data to the positive axis, i.e. $y = a * (x - x_0)^b$, any x values less than
> x_0 is still considered "negative". I don't know what else...maybe I'll
> try change of variable or something... thank you for your help.

What leads you to believe that y is some power of x? Is it simply a guess based upon the shape of the curve, or do you have some theoretical reason for expecting a power relationship?

Theories that lead to a power-law relationship without fixing the power to be a specific rational number generally apply only to data where the dependent variable is guaranteed to be positive. The numerical problems you have trying to fit such a relationship to data where x is sometimes negative are directly related to the reasons why theories tend not to imply the existence of such relationships.

For that same reason, if you're merely guessing at what the shape of the curve is, rather than getting it from a theory, I suspect that your guess is a bad one.

One possibility: the relationship isn't $y = a * x^b$; it's actually $y = a * |x|^b$. I've seen situations where that is a reasonable model. That will avoid the problems you've been having.

This is really a scientific problem, not a numerical one; figure out the right model for your data and the curve-fitting routines shouldn't have any problem fitting it to your data.

Subject: Re: Newbie's question
Posted by [JD Smith](#) on Fri, 21 Oct 2005 00:07:18 GMT
[View Forum Message](#) <> [Reply to Message](#)

On Thu, 20 Oct 2005 13:18:52 -0700, ChiChiRuiz@gmail.com wrote:

> Poly_fit doesn't really give me what I need. I don't need the
> coefficients of a quadratic equation, I want to know the best fit of the
> scatter plot to some power of x. I know it's not exactly power square,
> but it should be in that neighborhood. Even if I shift all data to the
> positive axis, i.e. $y = a * (x - x_0)^b$, any x values less than x_0 is still

> considered "negative". I don't know what else...maybe I'll try change of
> variable or something... thank you for your help.

Fitting to a single power law is a time honored tradition in many of the precision-limited fields of physics (e.g. astronomy). The typical approach is to fit a straight line to the log/log representation of the data. The slope of the line is the exponent b. If your data have negative values by artificial choice (e.g. time offset, etc.) simply shift that choice to make them positive.

JD

Subject: Re: Newbie's question
Posted by [James Kuyper](#) on Fri, 21 Oct 2005 12:55:53 GMT
[View Forum Message](#) <> [Reply to Message](#)

JD Smith wrote:

> On Thu, 20 Oct 2005 13:18:52 -0700, ChiChiRuiz@gmail.com wrote:
>
>> Poly_fit doesn't really give me what I need. I don't need the
>> coefficients of a quadratic equation, I want to know the best fit of the
>> scatter plot to some power of x. I know it's not exactly power square,
>> but it should be in that neighborhood. Even if I shift all data to the
>> positive axis, i.e. $y = a * (x - x_0)^b$, any x values less than x_0 is still
>> considered "negative". I don't know what else...maybe I'll try change of
>> variable or something... thank you for your help.
>
> Fitting to a single power law is a time honored tradition in many of
> the precision-limited fields of physics (e.g. astronomy).

True, but following that tradition is only appropriate when there's a specific reason to expect a power law of some kind.

> ... The typical
> approach is to fit a straight line to the log/log representation of
> the data. The slope of the line is the exponent b. If your data have
> negative values by artificial choice (e.g. time offset, etc.) simply
> shift that choice to make them positive.

The key point is that you need to know the appropriate amount to shift them. If the fact that you have negative numbers is "artificial", that implies that you may know the amount that needs to be added. Otherwise, adding an arbitrary amount could produce meaningless results. However, making a fit to the form $y = a * (x - x_0)^b$, with x_0 constrained to be less than the minimum value of x, could be a suitable approach.

Subject: Re: Newbie's question

Posted by ChiChiRuiz@gmail.com on Fri, 21 Oct 2005 16:58:21 GMT

[View Forum Message](#) <> [Reply to Message](#)

I agreed that it's more a scientific problem, rather than a numerical one. It'd just never crossed my mind that it would be this complicated. The x, y arrays are values from different images over the same pixel location, because of the stats analysis to produce these values, they "SHOULD" have a $y=x^2$ relationship, but due to large analytical errors, I know it's not exactly $y=x^2$. I just want to get a general idea for the scatter plot.
