Subject: Re: fitting mixed gaussians
Posted by David Fanning on Tue, 17 May 2005 18:32:59 GMT
View Forum Message <> Reply to Message

Ben Tupper writes:

> I would like to employ a "mixed gaussian" fitting routine that fits not
> just one peak in a distribution but rather fits any number of peaks.
> According to my old Gonzales and Woods Image Processing book this
> technique is used in Bayesian classifier - in the two class problem the
> saddle of the two fitted curves is used as the point around which the
> classes are separated. I just to segment simple grayscale images with it
> usiing the histogram as the data to model.
>
> Using GAUSSFIT or MPFIT it's very easy to fit one (the tallest) peak
> shown in the example (see code example below.) But how the heck to fit
> the second peak? I have scoured the internet and find much written
> about it - but it is all in Greek. Really. Yikes.
>
> Is it possible to make the curve fitting routines I have on hand to fit
> two gaussian models?

Craig is going to fill you in on all the details of this,
using MPFIT, but I just wanted to mention that Rob Dimeo's
PAN program is super for allowing you to fit as many curves
as you like onto a set of data interactively.

http://www.ncnr.nist.gov/staff/dimeo/panweb/pan.html

Just draw and adjust the initial curve parameters on
the data itself, then go fit it with a click of the button.
Of course, it is simply a wrapper for MPFIT. The tutorial
Rob has provided is helpful for getting started.

Cheers,

David

--
David Fanning, Ph.D.
Fanning Software Consulting, Inc.
Coyote's Guide to IDL Programming: http://www.dfanning.com/

Subject: Re: fitting mixed gaussians
Posted by Rob Dimeo on Tue, 17 May 2005 19:55:39 GMT
View Forum Message <> Reply to Message

Thanks for the plug, David. Yep, PAN is just a big GUI wrapper for MPCURVEFIT. In fact the kind of fitting described in the original post is done routinely at our neutron scattering facility where the whole model function can be composed of many Gaussians, Lorentzians, Lognormals, etc.

I should also mention that PAN has been updated to include an alternative error estimation based on the *bootstrap Monte-Carlo* method which can be useful for fitting functions that are highly non-linear in the fit parameters. There is a discussion of this in Numerical Recipes for those intereseted.

Rob

David Fanning wrote:
> Ben Tupper writes:
>
>> I would like to employ a "mixed gaussian" fitting routine that fits not
>> just one peak in a distribution but rather fits any number of peaks.
>> According to my old Gonzales and Woods Image Processing book this
>> technique is used in Bayesian classifier - in the two class problem the
>> saddle of the two fitted curves is used as the point around which the
>> classes are separated. I just to segment simple grayscale images with it
>> usiing the histogram as the data to model.
>>
>> Using GAUSSFIT or MPFIT it's very easy to fit one (the tallest) peak
>> shown in the example (see code example below.) But how the heck to fit
>> the second peak? I have scoured the internet and find much written
>> about it - but it is all in Greek. Really. Yikes.
>>
>> Is it possible to make the curve fitting routines I have on hand to fit
>> two gaussian models?
>
> Craig is going to fill you in on all the details of this,
> using MPFIT, but I just wanted to mention that Rob Dimeo's
> PAN program is super for allowing you to fit as many curves
> as you like onto a set of data interactively.
>
> http://www.ncnr.nist.gov/staff/dimeo/panweb/pan.html

>
> Just draw and adjust the initial curve parameters on
> the data itself, then go fit it with a click of the button.
> Of course, it is simply a wrapper for MPFIT. The tutorial
> Rob has provided is helpful for getting started.
>
> Cheers,
>
> David
>
> --
> David Fanning, Ph.D.
> Fanning Software Consulting, Inc.
> Coyote's Guide to IDL Programming: http://www.dfanning.com/

Subject: Re: fitting mixed gaussians
Posted by btt on Wed, 18 May 2005 14:54:36 GMT
View Forum Message <> Reply to Message

Rob wrote:
> Thanks for the plug, David. Yep, PAN is just a big GUI wrapper for
> MPCURVEFIT. In fact the kind of fitting described in the original post
> is done routinely at our neutron scattering facility where the whole
> model function can be composed of many Gaussians, Lorentzians,
> Lognormals, etc.
>

>>
>> Just draw and adjust the initial curve parameters on
>> the data itself, then go fit it with a click of the button.
>> Of course, it is simply a wrapper for MPFIT. The tutorial
>> Rob has provided is helpful for getting started.
>>
Thanks David and Rob,

Yes I looked at these - the trouble is that I want to perform this on literally thousands of images (each image has one object in it). Manual seeding is not practical. I think what I am after is described here...

http://tinyurl.com/9horr

I have started to translate the MatLab code to IDL - but it is clearly over my head.

Thanks again,
Ben

Subject: Re: fitting mixed gaussians
Posted by jcami on Wed, 18 May 2005 20:35:53 GMT
View Forum Message <> Reply to Message

> Yes I looked at these - the trouble is that I want to perform this on

> literally thousands of images (each image has one object in it).
Manual
> seeding is not practical. I think what I am after is described
here...
>
> http://tinyurl.com/9horr
>
> I have started to translate the MatLab code to IDL - but it is
clearly
> over my head.

I would start by writing a function that creates multiple Gaussians.
Something like

```
FUNCTION multi_gauss, X, P

; P contains a set of 3 parameters per Gaussian, plus an offset and a
slope
; for the linear part of the model.

n_gauss = (n_elements(P)-2)/3
profile = dblarr(n_elements(X))
for loop=0,n_gauss-1 do begin
 this_P = P[loop*3:loop*3+2]
 this_u = (X-this_P[1])/this_P[2]
 profile = profile + this_P[0] * exp(-5d-1 * this_u^2)
endfor
profile = profile + P[n_g * 3]
profile = profile + P[n_g * 3 + 1] * X

return, profile

END
```

You can feed this sort of function to mpfitfun; in your peak fitting
routine
you can then start with a single Gaussian, subtract the fit from the
data and look for a second Gaussian (again using something like
mpfitpeak); if you find a second Gaussian, you use the parameters for
both Gaussians as initial estimates to the multi_gauss routine and
mpfitfun should do the rest.

Subject: Re: fitting mixed gaussians
Posted by Craig Markwardt on Thu, 19 May 2005 00:37:28 GMT
View Forum Message <> Reply to Message

Ben Tupper <btupper@bigelow.org> writes:

> Rob wrote:
>> Thanks for the plug, David. Yep, PAN is just a big GUI wrapper for
>> MPCURVEFIT. In fact the kind of fitting described in the original post
>> is done routinely at our neutron scattering facility where the whole
>> model function can be composed of many Gaussians, Lorentzians,
>> Lognormals, etc.
>>
>
>>>
>>> Just draw and adjust the initial curve parameters on
>>> the data itself, then go fit it with a click of the button.
>>> Of course, it is simply a wrapper for MPFIT. The tutorial
>>> Rob has provided is helpful for getting started.
>>>
> Thanks David and Rob,
>
> Yes I looked at these - the trouble is that I want to perform this on
> literally thousands of images (each image has one object in it).
> Manual seeding is not practical. I think what I am after is described
> here...
>
> http://tinyurl.com/9horr
>
> I have started to translate the MatLab code to IDL - but it is clearly
> over my head.

Long ago I had somebody write to me, asking about how to do this kind of thing. I don't think he ever succeeded.

You are going to have problems like:

1. uniqueness - there are essentially an infinite number of ways to add gaussians to reproduce the data; for example, why not have one gaussian per sample?

2. robustness - the problem is so unconstrained that there is significant potential for screwed up fits.

I would recommend constraining the analysis as much as possible based on your problem domain, for example if you know that peaks must be positive, or the natural widths of the peaks, etc.

One technique would be to find the tallest peak, or perhaps the N

tallest peaks in the data, and fit gaussians to those. Then subtract that fit and see what the next highest peaks are, and fit another set of gaussians. You keep doing that until you reach some noise threshold (i.e. the errors on the amplitudes are comparable to the amplitudes). Sometimes this is known as "CLEAN" in image space.

But be prepared for the fitting to fail or give whacked results at least 30% of the time.

Good luck,
Craig

--
---------- ----------
Craig B. Markwardt, Ph.D. EMAIL: craigmnet@REMOVEcow.physics.wisc.edu
Astrophysics, IDL, Finance, Derivatives | Remove "net" for better response
---------- ----------

Subject: Re: fitting mixed gaussians
Posted by btt on Fri, 20 May 2005 14:10:06 GMT
View Forum Message <> Reply to Message

Craig Markwardt wrote:
> Ben Tupper <btupper@bigelow.org> writes:
>
>
>> Rob wrote:
>>
>>> Thanks for the plug, David. Yep, PAN is just a big GUI wrapper for
>>> MPCURVEFIT. In fact the kind of fitting described in the original post
>>> is done routinely at our neutron scattering facility where the whole
>>> model function can be composed of many Gaussians, Lorentzians,
>>> Lognormals, etc.
>>>
>>
>>>> Just draw and adjust the initial curve parameters on
>>>> the data itself, then go fit it with a click of the button.
>>>> Of course, it is simply a wrapper for MPFIT. The tutorial
>>>> Rob has provided is helpful for getting started.
>>>>
>>
>> Thanks David and Rob,
>>
>> Yes I looked at these - the trouble is that I want to perform this on
>> literally thousands of images (each image has one object in it).
>> Manual seeding is not practical. I think what I am after is described
>> here...

>>
>> http://tinyurl.com/9horr
>>
>> I have started to translate the MatLab code to IDL - but it is clearly
>> over my head.
>
>
> Long ago I had somebody write to me, asking about how to do this kind
> of thing. I don't think he ever succeeded.
>
> You are going to have problems like:
>
> 1. uniqueness - there are essentially an infinite number of ways to
> add gaussians to reproduce the data; for example, why not have
> one gaussian per sample?
>
> 2. robustness - the problem is so unconstrained that there is
> significant potential for screwed up fits.
>
> I would recommend constraining the analysis as much as possible based
> on your problem domain, for example if you know that peaks must be
> positive, or the natural widths of the peaks, etc.
>
> One technique would be to find the tallest peak, or perhaps the N
> tallest peaks in the data, and fit gaussians to those. Then subtract
> that fit and see what the next highest peaks are, and fit another set
> of gaussians. You keep doing that until you reach some noise
> threshold (i.e. the errors on the amplitudes are comparable to the
> amplitudes). Sometimes this is known as "CLEAN" in image space.
>
> But be prepared for the fitting to fail or give whacked results at
> least 30% of the time.
>

Thank you. Yes there do seem to be a lot of pitfalls. As long as the peaks are distinct, it seems feasible. Rob Dimeo has shown me how to do that using his GET_PEAK_POS routine. It really works well until the peaks are either close together or the second peak has a very small amplitude. I see the latter case when my images contain objects that are 'all edge'. That is when the interior of the objects (plankton in this case) have transparent interiors. The fit-the-biggest-peaks-then-remove works pretty well in this case.

Thanks again,
Ben