
Subject: Re: Segregating data in bimodal distribution
Posted by [Jeremy Bailin](#) on Wed, 03 Aug 2011 15:37:36 GMT
[View Forum Message](#) <> [Reply to Message](#)

On 8/3/11 8:35 AM, Eric Hudson wrote:

> Hi,
>
> Is anyone aware of an IDL implemented algorithm for segregating data
> in a bimodal distribution into two groups?
>
> My data is such that I could do it manually (make a histogram, decide
> on a threshold between the two peaks in the histogram, then pull out
> the data above and below that into two separate groups). There isn't
> a true gap between the two peaks, but they are pretty well separated.
> The part which is non-obvious to me is to how to programmatically
> choose the threshold value. And since I have to do this on many data
> sets, where the threshold is going to be different for each, I prefer
> to not do it manually.
>
> Thanks,
> Eric
>
> PS In searching I found something called the KMM algorithm which
> seems like it would work, but I haven't found code for it.

Are the peaks well-represented by a known function (e.g. Gaussian)? If so, you could fit a bimodal Gaussian/whatever to the distribution and use the parameters of the fit to determine when the total is dominated by one or the other peak.

-Jeremy.

Subject: Re: Segregating data in bimodal distribution
Posted by [ben.bighair](#) on Sun, 07 Aug 2011 14:01:22 GMT
[View Forum Message](#) <> [Reply to Message](#)

Hi,

On 8/3/11 11:37 AM, Jeremy Bailin wrote:

> On 8/3/11 8:35 AM, Eric Hudson wrote:
>> Hi,
>>
>> Is anyone aware of an IDL implemented algorithm for segregating data
>> in a bimodal distribution into two groups?
>>
>> My data is such that I could do it manually (make a histogram, decide

>> on a threshold between the two peaks in the histogram, then pull out
>> the data above and below that into two separate groups). There isn't
>> a true gap between the two peaks, but they are pretty well separated.
>> The part which is non-obvious to me is to how to programmatically
>> choose the threshold value. And since I have to do this on many data
>> sets, where the threshold is going to be different for each, I prefer
>> to not do it manually.
>>
>> Thanks,
>> Eric
>>
>> PS In searching I found something called the KMM algorithm which
>> seems like it would work, but I haven't found code for it.
>
> Are the peaks well-represented by a known function (e.g. Gaussian)? If
> so, you could fit a bimodal Gaussian/whatever to the distribution and
> use the parameters of the fit to determine when the total is dominated
> by one or the other peak.

A while back I translated some MatLab code to do this sort of thing. I
never got it to run very fast but it seemed to do pretty well. If I
rightly recall, I think it performed well when the peaks overlapped a lot.

You can find a copy of it here...

http://dl.dropbox.com/u/8433654/mb_mixg.pro

Note there are some obscure references and an example routine...

```
IDL> .compile mb_mixg
IDL> example
Threshold Selected =      132.47748
```

Cheers,
Ben

Subject: Re: Segregating data in bimodal distribution
Posted by [Eric Hudson](#) on Wed, 10 Aug 2011 11:59:45 GMT
[View Forum Message](#) <> [Reply to Message](#)

>
> A while back I translated some MatLab code to do this sort of thing. I
> never got it to run very fast but it seemed to do pretty well. If I
> rightly recall, I think it performed well when the peaks overlapped a lot.
>
> You can find a copy of it here...
>

> http://dl.dropbox.com/u/8433654/mb_mixg.pro
>
> Note there are some obscure references and an example routine...

Wow, that is exactly what I needed. The example even looks like my data.

Thanks so much,
Eric

Subject: Re: Segregating data in bimodal distribution
Posted by [wulf.hendrik](#) on Wed, 22 Feb 2017 14:30:34 GMT
[View Forum Message](#) <> [Reply to Message](#)

On Sunday, August 7, 2011 at 4:01:22 PM UTC+2, ben.bighair wrote:

> Hi,
>
>
> On 8/3/11 11:37 AM, Jeremy Bailin wrote:
>> On 8/3/11 8:35 AM, Eric Hudson wrote:
>>> Hi,
>>>
>>> Is anyone aware of an IDL implemented algorithm for segregating data
>>> in a bimodal distribution into two groups?
>>>
>>> My data is such that I could do it manually (make a histogram, decide
>>> on a threshold between the two peaks in the histogram, then pull out
>>> the data above and below that into two separate groups). There isn't
>>> a true gap between the two peaks, but they are pretty well separated.
>>> The part which is non-obvious to me is to how to programmatically
>>> choose the threshold value. And since I have to do this on many data
>>> sets, where the threshold is going to be different for each, I prefer
>>> to not do it manually.
>>>
>>> Thanks,
>>> Eric
>>>
>>> PS In searching I found something called the KMM algorithm which
>>> seems like it would work, but I haven't found code for it.
>>
>> Are the peaks well-represented by a known function (e.g. Gaussian)? If
>> so, you could fit a bimodal Gaussian/whatever to the distribution and
>> use the parameters of the fit to determine when the total is dominated
>> by one or the other peak.
>
> A while back I translated some MatLab code to do this sort of thing. I
> never got it to run very fast but it seemed to do pretty well. If I

> rightly recall, I think it performed well when the peaks overlapped a lot.
>
> You can find a copy of it here...
>
> http://dl.dropbox.com/u/8433654/mb_mixg.pro
>
> Note there are some obscure references and an example routine...
>
> IDL> .compile mb_mixg
> IDL> example
> Threshold Selected = 132.47748
>
> Cheers,
> Ben

Hi Ben,

do you still have the original Matlab code?

best,
Hendrik
