Title: Computational Physics - Lecture 10

Date: Oct 12, 2018 01:00 PM

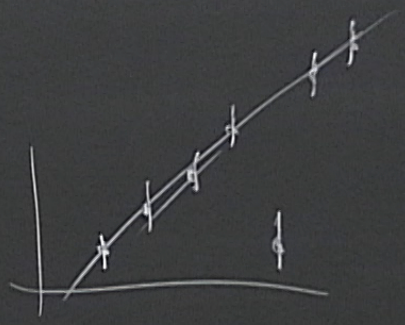
URL: http://pirsa.org/18100051

Abstract:

1008.4686 s3

## Outliers

- by hand
- by heuristic (sigma clipping)
- statistics!



## Generative models

$y_i = b x_i + m + e_i$

- need to model how our data can be bad.

Data core from good "foreground" or bad "background" distributions

- backgrounds Gaussian

① Add New params

$$q_i = \begin{cases} 0 & \text{if data point } i \text{ is bad} \\ 1 & \text{good} \end{cases}$$

$$P(y_i \mid m, b, g_i, Y, V) = P_{fg}(y_i \mid m, b)^{g_i} \times P_{bg}(y_i \mid Y, V)^{(1-g_i)}$$

Marginalize out $g_i$:

$$\int P(y_i \mid m, b, g_i, Y, V)\, \underbrace{P(g_i)}_{\text{probability that any } g_i = 1}\, dg_i$$

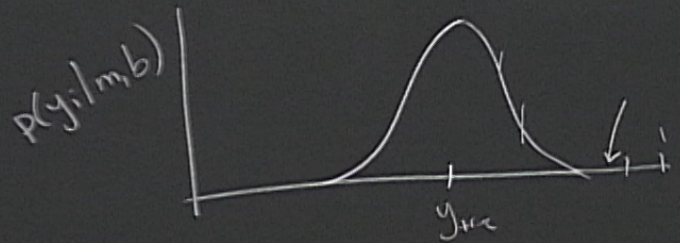
$$= P(y_i \mid \ldots g_i = 0 \ldots) P(g_i = 0) + P(y_i \mid \ldots g_i = 1 \ldots) P(g_i = 1)$$

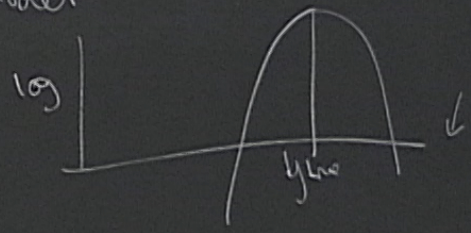
$$P(y_i \mid m, b, Y, V, P_{bad}) = P_{bad} \times P_{bg}(y_i \mid Y, V) + (1 - P_{bad}) \times P_{fg}(y_i \mid m, b)$$

"Mixture model" or "foreground-background" model

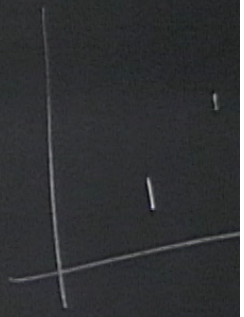
$$P(y_i \mid m, b, Y, V, P_{bad}) = P_{bad} \times P_{bg}(y_i \mid Y, V) + (1 - P_{bad}) \times P_{fg}(y_i \mid m, b)$$

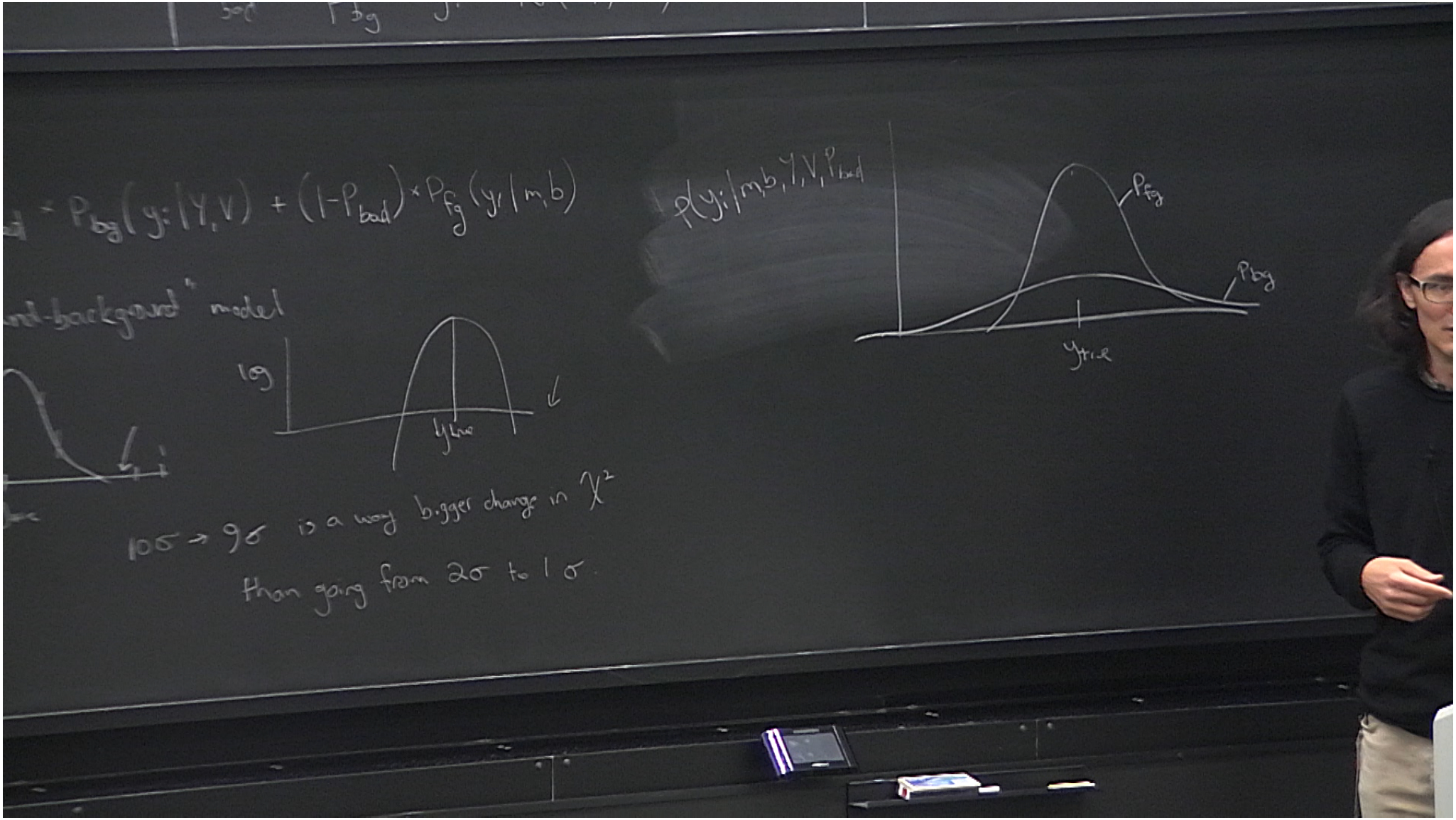
"Mixture model" or "foreground-background" model

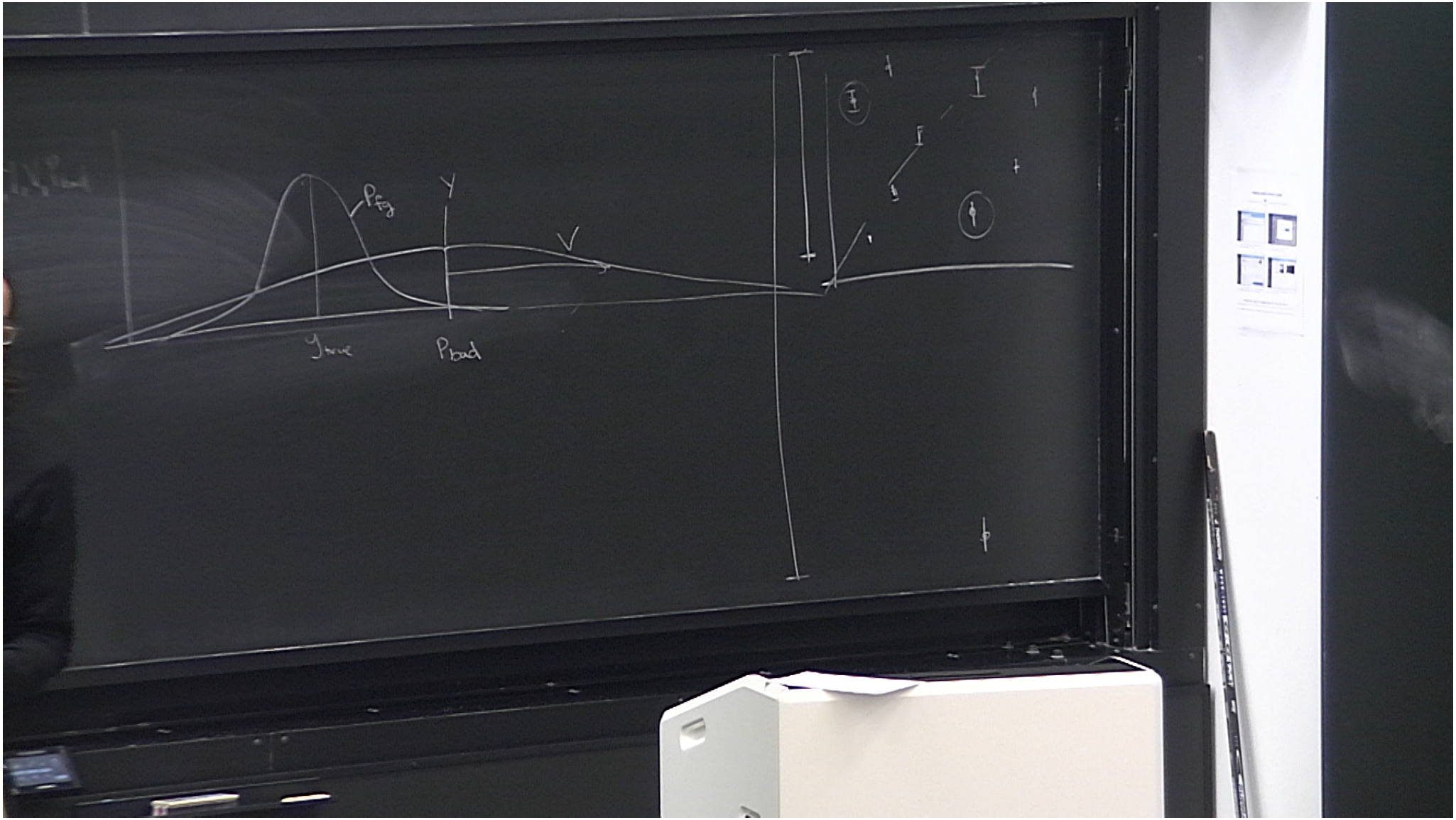



$10\sigma \to 9\sigma$ is a way bigger change in $\chi^2$ than going from $2\sigma$ to $1\sigma$.




$P_{bg}(y_i|Y,V) + (1-P_{bad}) \times P_{fg}(y_i|m,b)$
$p(y_i|m,b,Y,V,P_{bad})$
$P_{fg}$
$P_{bg}$
$y_{true}$
"background" model
log
$y_{true}$
$10\sigma \rightarrow 9\sigma$ is a way bigger change in $\chi^2$
than going from $2\sigma$ to $1\sigma$.


Pfg
Y
V
ytrue
Pbad

Chrome File Edit View History Bookmarks People Window Help — Fri 1:47 PM Dustin Lang

lab2-scaffold-1 | Home | lab2-scaffold

localhost:8888/notebooks/lab2-scaffold.ipynb

jupyter lab2-scaffold Last Checkpoint: an hour ago (autosaved) Logout

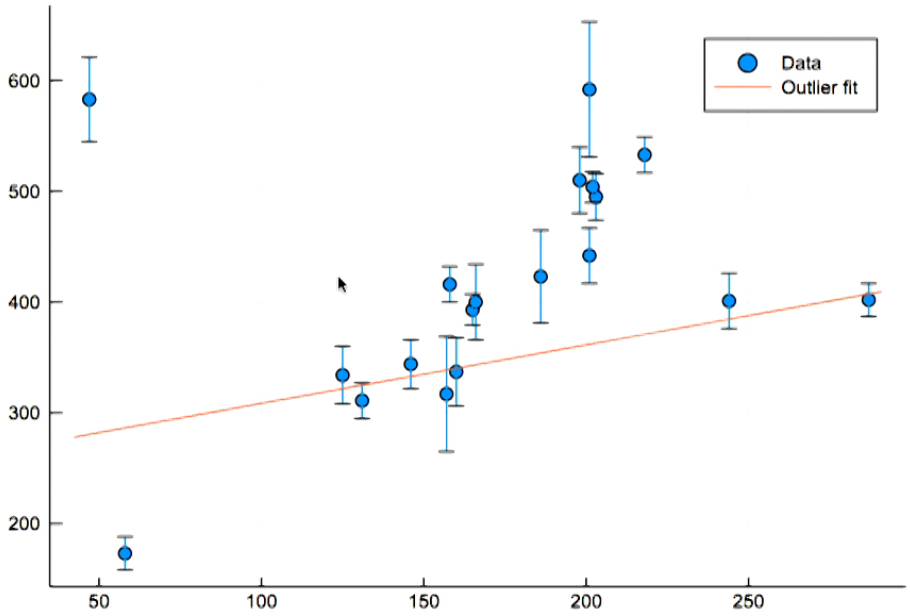
File Edit View Insert Cell Kernel Widgets Help Trusted Julia 1.0.0

Let's pull out the b,m terms and plot them:

In [21]:

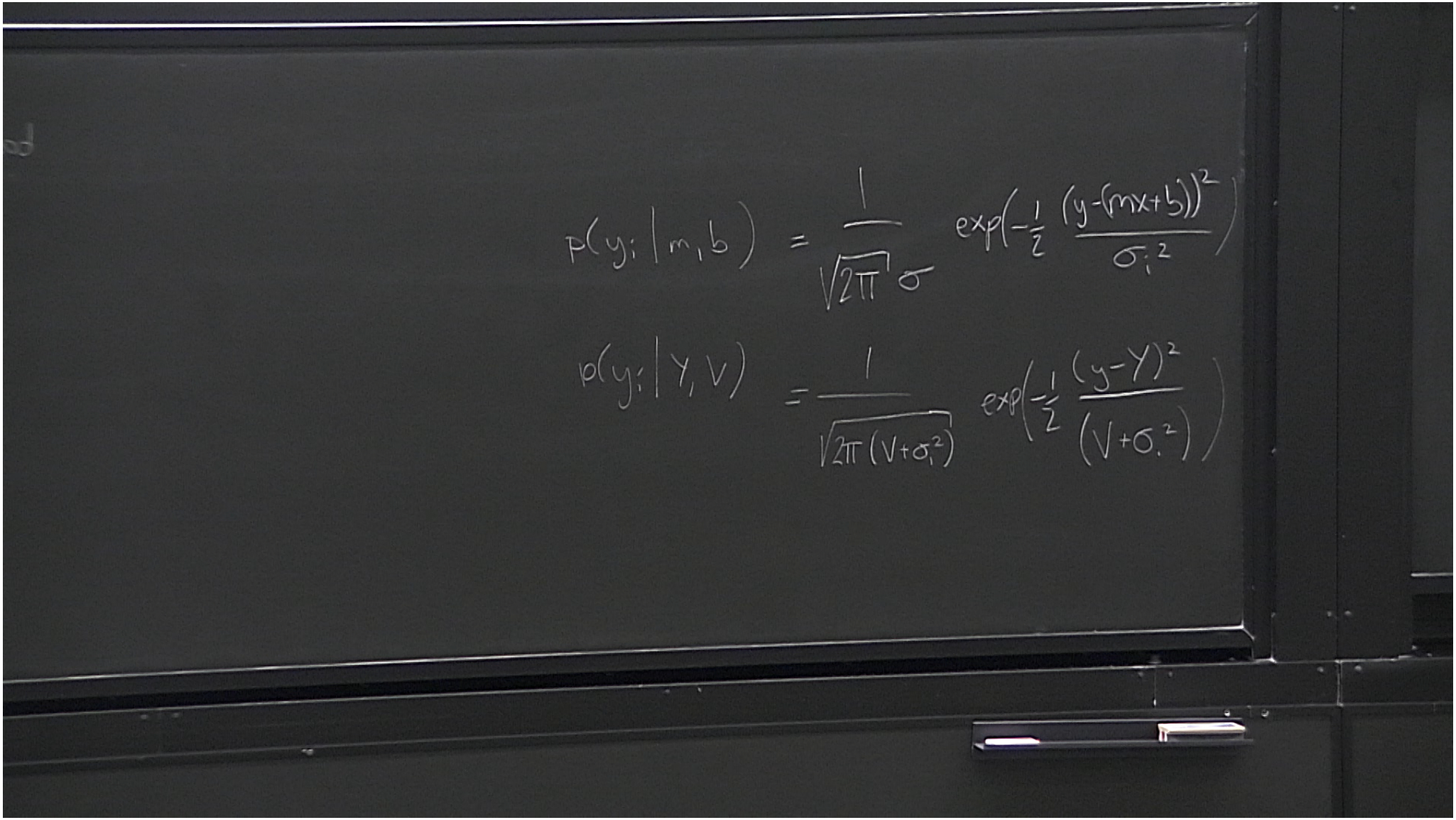
```
1 bx,mx = r.minimizer[1:2]
2 plot(x, y, yerr=yerr, seriestype=:scatter, label="Data")
3 plot!(x->bx+mx*x, label="Outlier fit")
```

Out[21]:




Does this make you happy?

Remember how I said the optimizer finds a local minimum? Can you try initializing it in a part of parameter space that looks about right?


$p(y_i | m, b) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2} \frac{(y-(mx+b))^2}{\sigma_i^2}\right)$
$p(y_i | Y, V) = \frac{1}{\sqrt{2\pi(V+\sigma_i^2)}} \exp\left(-\frac{1}{2} \frac{(y-Y)^2}{(V+\sigma_i^2)}\right)$

fg bg - log - likelihood

p-fg = @.

p-bg