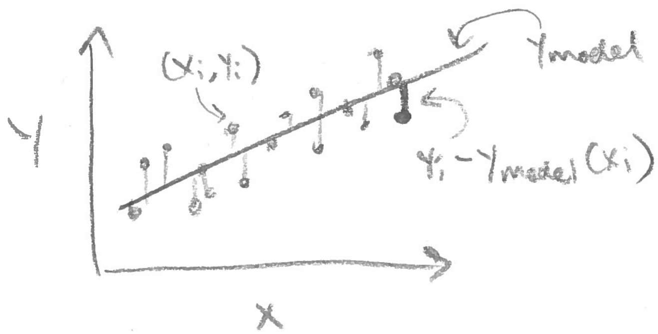
How do we determine $\underbrace{P(\underbrace{a,b}_{\text{parameters}}\,|\,\underbrace{D}_{\text{"data"}})}_{\text{"given"}}$?

Let's assume that each data point y_i is drawn from a Gaussian distribution centered on $y_{model}(x_i)$ and with a width of σ.



That means that for each datapoint, the PDF is

$$P(y_i) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{y_i - y_{model}(x_i)}{\sigma}\right)^2}$$

and we will further assume that our datapoints are all independent of each other.

Therefore, the probability of our combined dataset is the product of the independent probability PDFs

$$P(D|a,b) = P(y_0)\,P(y_1)\,P(y_2)\ldots$$
$$= \prod_{i=1}^{N} P(y_i)$$

This "likelihood" $P(D|a,b)$ is the probability of our data given the parameters of our model. It's a function we can calculate directly for any values of a and b.

Bayes Theorem tells us

$$\underbrace{P(a,b|D)}_{\text{posterior probability}} \propto \underbrace{P(D|a,b)}_{\text{likelihood}}\ \underbrace{P(a,b)}_{\text{prior probability}}$$

If we assume flat priors (we know nothing a priori about relative probabilities of a and b values), then $P(a,b)$ is a constant, and

$$P(a,b|D) \propto P(D|a,b)$$

[This is the statement that Bayesian and frequentist approaches are the same in the limit of flat priors.]

We can now actually write out the whole expression:

$$P(a,b|D) \propto \prod_{i=1}^{N} p(y_i|a,b)$$

$$\propto \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{y_i - y_{model}(x_i)}{\sigma}\right)^2}$$

$$\propto \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^N \prod_{i=1}^{N} e^{-\frac{1}{2}\left(\frac{y_i - y_{model}(x_i)}{\sigma}\right)^2}$$

which, gosh, looks like a mess. But, what if we took the logarithm?

$$\ln P(a,b|D) = N \ln\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right) - \frac{1}{2}\sum_{i=1}^{N}\left(\frac{y_i - y_{model}}{\sigma}\right)^2$$

where we can define the handy quantity of $\chi^2 = \sum_{i=1}^{N}\left(\frac{y_i - y_{model}(x_i)}{\sigma}\right)^2$ and so this becomes simply

$$\ln P(a,b|D) = N \ln\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right) - \frac{1}{2}\chi^2$$

If we want to maximize the probability $P(a,b|D)$, then we would want to minimize the χ^2! This is why χ^2-minimization is such a thing! If our assumptions hold about the Gaussian probability distributions for each of our y_i, then minimizing χ^2 is really telling us the maximum likelihood estimate of our parameter values!