

Module 1 | Lesson 3

Least Squares and Maximum Likelihood

Least Squares and Maximum Likelihood

By the end of this video, you will be able to...

- State the connection between the method of least squares and maximum likelihood with Gaussian random variables

Revisiting the Least Squares Criterion

- Recall the least squares criterion

$$\mathcal{L}_{\text{LS}}(x) = (y_1 - x)^2 + (y_2 - x)^2 + \dots + (y_m - x)^2$$

- We've said that the optimal estimate, \hat{x} , is the one that minimizes this 'loss':

$$\hat{x}_{\text{LS}} = \operatorname{argmin}_x \mathcal{L}_{\text{LS}}(x) = \operatorname{argmin}_x (e_1^2 + e_2^2 + \dots + e_m^2)$$

Why squared errors?

Least Squares and Maximum Likelihood



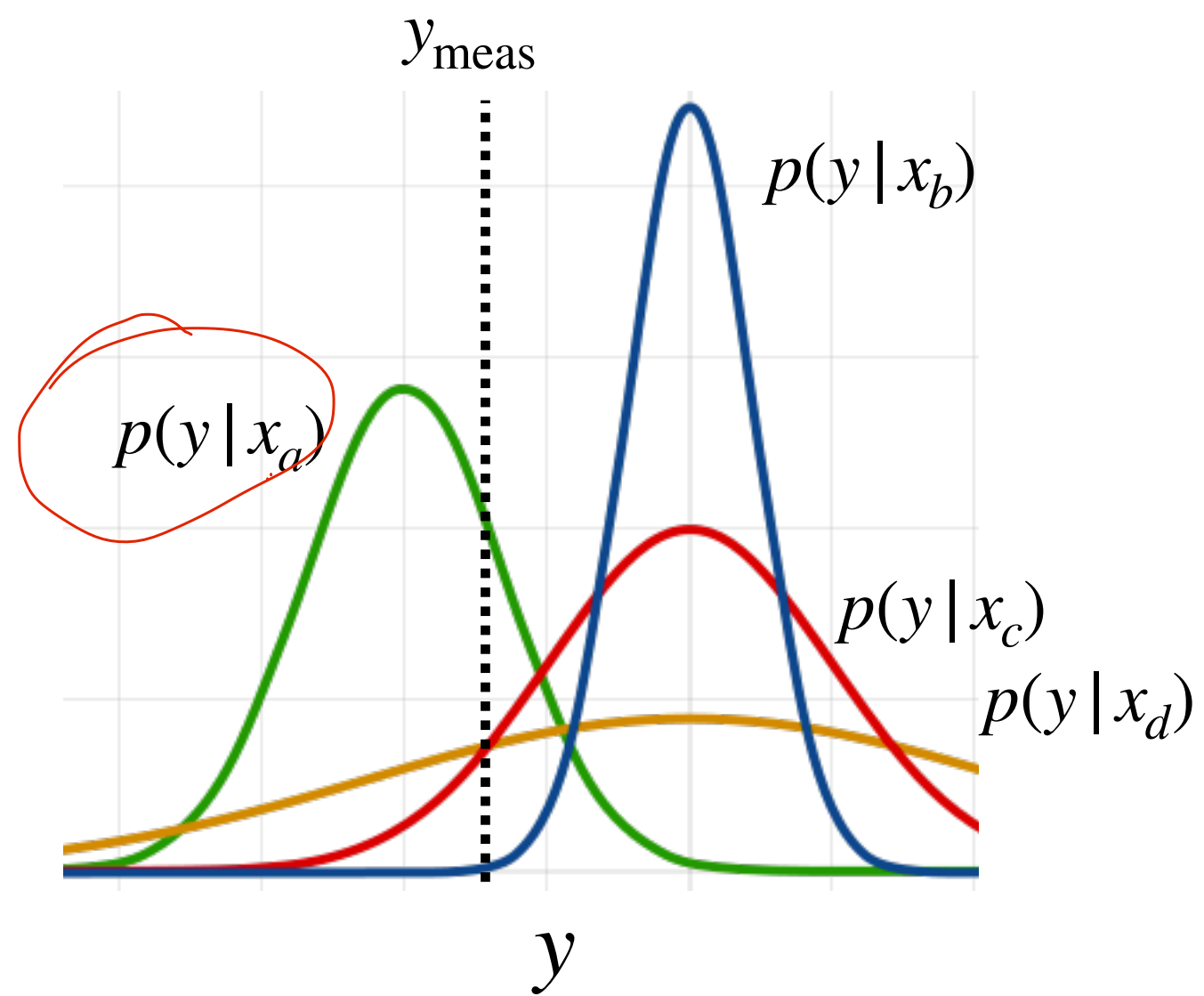
Karl Friedrich Gauss
'Princeps mathematicorum'

Gauss showed the connection between his method of least squares and maximum likelihood with Gaussian measurement models

The Method of Maximum Likelihood

- We can ask which x makes our measurement *most likely*. Or, in other words, which x maximizes the conditional probability of y :

$$\hat{x} = \operatorname{argmax}_x p(y|x)$$



Which x is the most likely given the measurement?

Measurement Model

- Recall our simple measurement model:

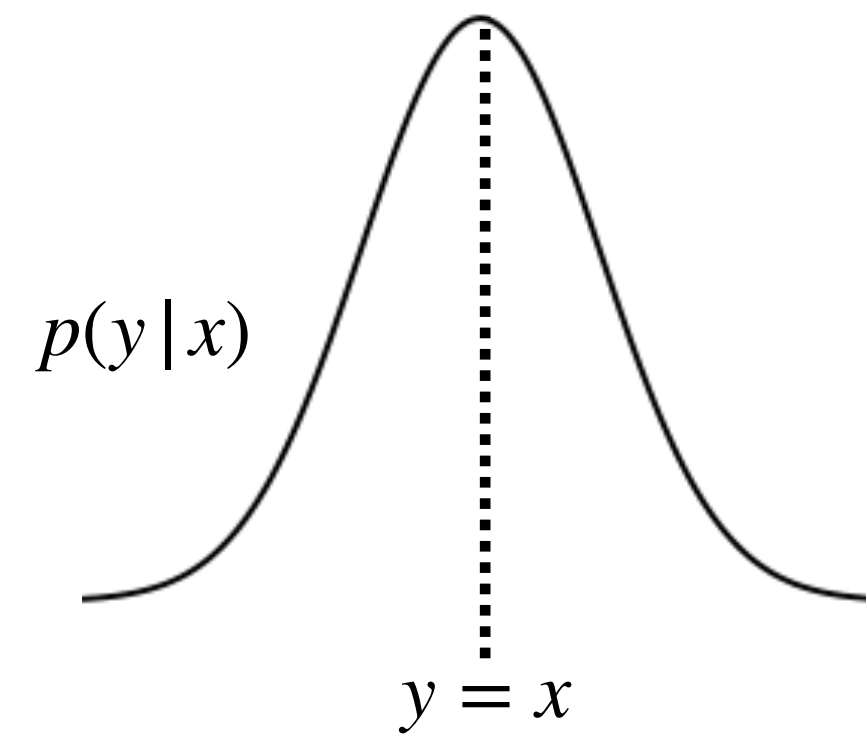
$$y = x + v$$

- We can convert this into a conditional probability on our measurement, by assuming some probability density for v . For example, if

$$v \sim \mathcal{N}(0, \sigma^2)$$

- Then:

$$p(y|x) = \mathcal{N}(x, \sigma^2)$$



Least Squares and Maximum Likelihood

- Probability density function of a Gaussian is:

$$\mathcal{N}(z; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(z-\mu)^2}{2\sigma^2}}$$

- Our conditional measurement likelihood is

$$\begin{aligned} p(y|x) &= \mathcal{N}(y; x, \sigma^2) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(y-x)^2}{2\sigma^2}} \end{aligned}$$

- If we have multiple independent measurements, then:

$$\begin{aligned} p(\mathbf{y}|x) &\propto \mathcal{N}(y_1; x, \sigma^2) \mathcal{N}(y_2; x, \sigma^2) \times \dots \times \mathcal{N}(y_m; x, \sigma^2) \\ &= \frac{1}{\sqrt{(2\pi)^m \sigma^{2m}}} \exp\left(\frac{-\sum_{i=1}^m (y_i - x)^2}{2\sigma^2}\right) \end{aligned}$$

Least Squares and Maximum Likelihood

- The maximal likelihood estimate (MLE) is given by

$$\hat{x}_{\text{MLE}} = \operatorname{argmax}_x p(\mathbf{y} | x)$$

- Instead of trying to optimize the likelihood directly, we can take its logarithm:

$$\begin{aligned}\hat{x}_{\text{MLE}} &= \operatorname{argmax}_x p(\mathbf{y} | x) \\ &= \operatorname{argmax}_x \log p(\mathbf{y} | x)\end{aligned}$$

The logarithm is monotonically increasing!

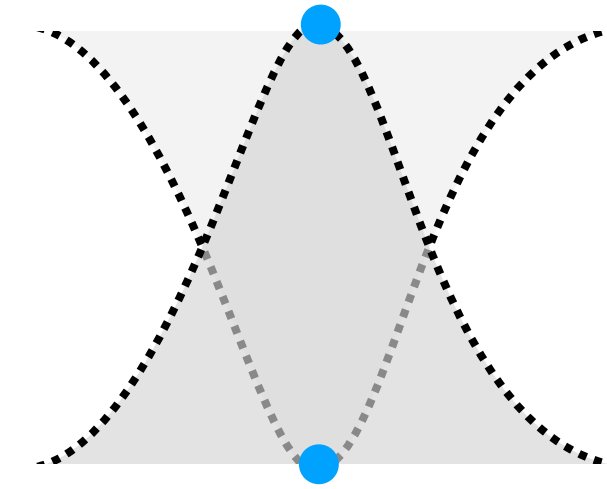
- Resulting in:

$$\log p(\mathbf{y} | x) = -\frac{1}{2\sigma^2} \left((y_1 - x)^2 + \dots + (y_m - x)^2 \right) + C$$

Least Squares and Maximum Likelihood

- Since

$$\operatorname{argmax}_z f(z) = \operatorname{argmin}_z (-f(z))$$



- The maximal likelihood problem can therefore be written as

$$\begin{aligned}\hat{x}_{\text{MLE}} &= \operatorname{argmin}_x -(\log p(\mathbf{y} | x)) \\ &= \operatorname{argmin}_x \frac{1}{2\sigma^2} ((y_1 - x)^2 + \dots + (y_m - x)^2)\end{aligned}$$

Least Squares and Maximum Likelihood

- So:

$$\hat{x}_{\text{MLE}} = \operatorname{argmin}_x \frac{1}{2\sigma^2} \left((y_1 - x)^2 + \dots + (y_m - x)^2 \right)$$

- Finally, if we assume each measurement has a different variance, we can derive

$$\hat{x}_{\text{MLE}} = \operatorname{argmin}_x \frac{1}{2} \left(\frac{(y_1 - x)^2}{\sigma_1^2} + \dots + \frac{(y_m - x)^2}{\sigma_m^2} \right)$$

In both cases,

$$\hat{x}_{\text{MLE}} = \hat{x}_{\text{LS}} = \operatorname{argmin}_x \mathcal{L}_{\text{LS}}(x) = \operatorname{argmin}_x \mathcal{L}_{\text{MLE}}(x)$$

The Central Limit Theorem

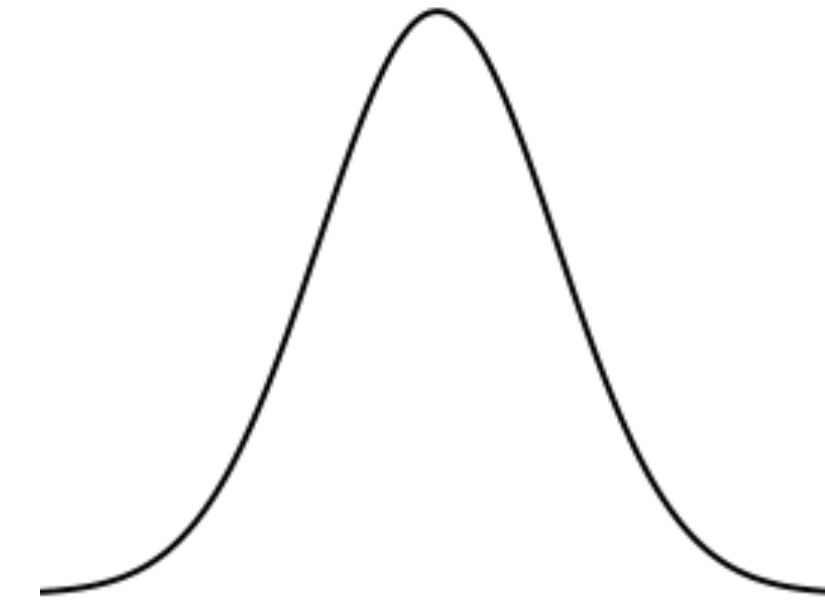
- In realistic systems like self driving cars, there are many sources of 'noise'

Central Limit Theorem: When independent random variables are added, their normalized sum tends towards a normal distribution.

- Why use the method of least squares?
 1. Central Limit Theorem: sum of different errors will tend be 'Gaussian'-ish
 2. Least squares is equivalent to maximum likelihood under Gaussian noise

Least Squares I Some Caveats

- ‘Poor’ measurements (e.g. outliers) have a significant effect on the method of least squares
- It’s important to check that the measurements roughly follow a Gaussian distribution



Under the Gaussian PDF, samples ‘far away’ from the mean are ‘very improbable’

#	Resistance (Ohms)
1	1068
2	988
3	1002
4	996

$$\hat{x} = 1013.5$$

#	Resistance (Ohms)
1	1068
2	988
3	1002
4	996
5 (outlier)	1430

$$\hat{x} = 1096.8$$

Summary | Least Squares and Maximum Likelihood

- LS and WLS produce the same estimates as maximum likelihood assuming Gaussian noise
- Central Limit Theorem states that complex errors will tend towards a Gaussian distribution.
- Least squares estimates are significantly affected by *outliers*