Fit to Eat

A Fable About Loss Functions

Laurel Ruhlen

April 22, 2025

Roadmap

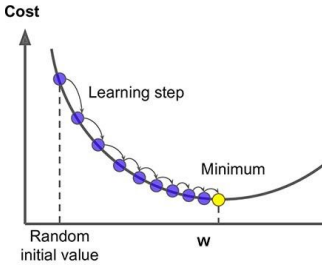
- Quick bio
- What and why: the Hungry King
- Loss functions
  - Quick intro
  - Typical use
  - Typical kludge
  - Custom loss functions
- Tips, tricks, and caveats

Quick Bio

2002 — 2014 — 2023 — Now

Physics → **DS & MLE** → **EM**

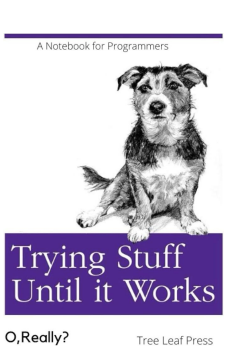
χ^2



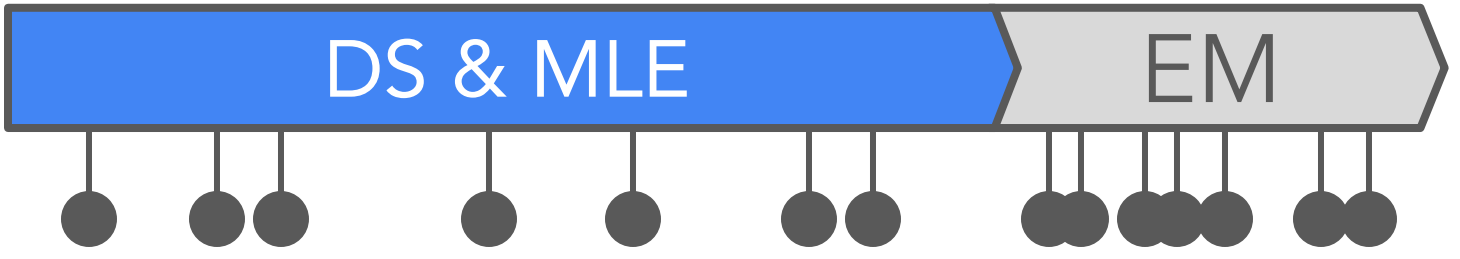
$e^{i\pi} = -1$

jupyter

AWS

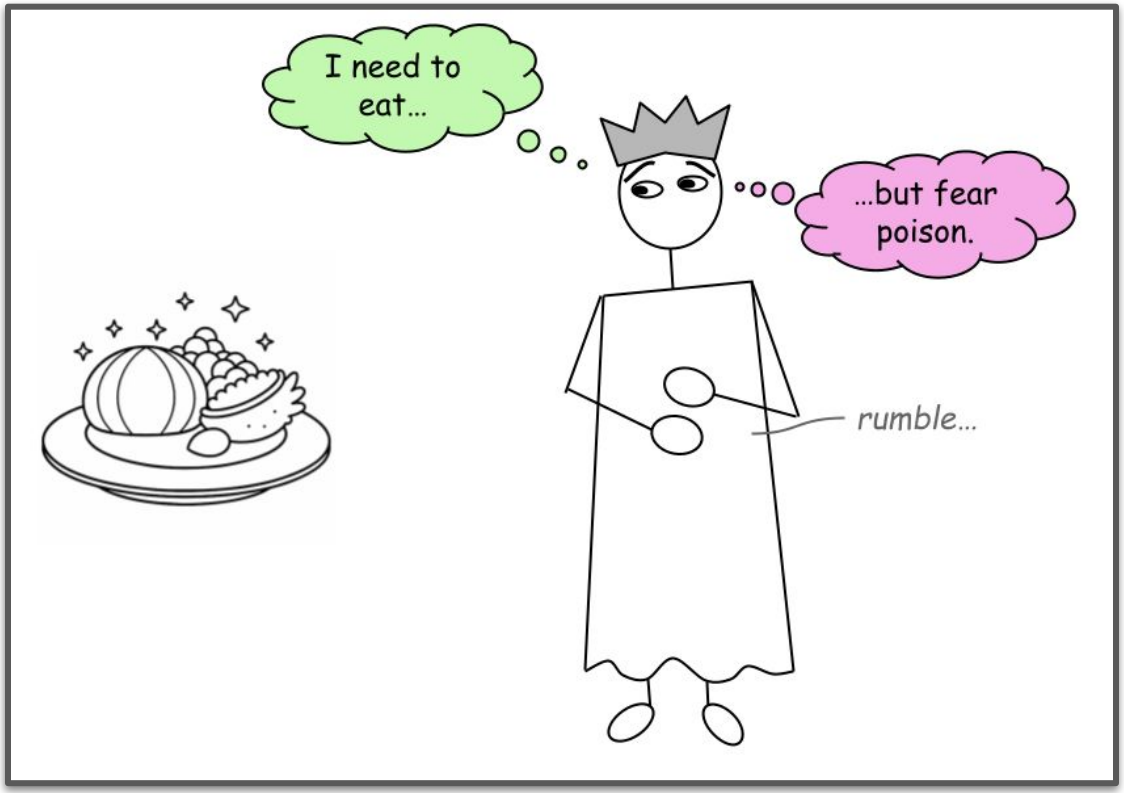


There's a Pattern



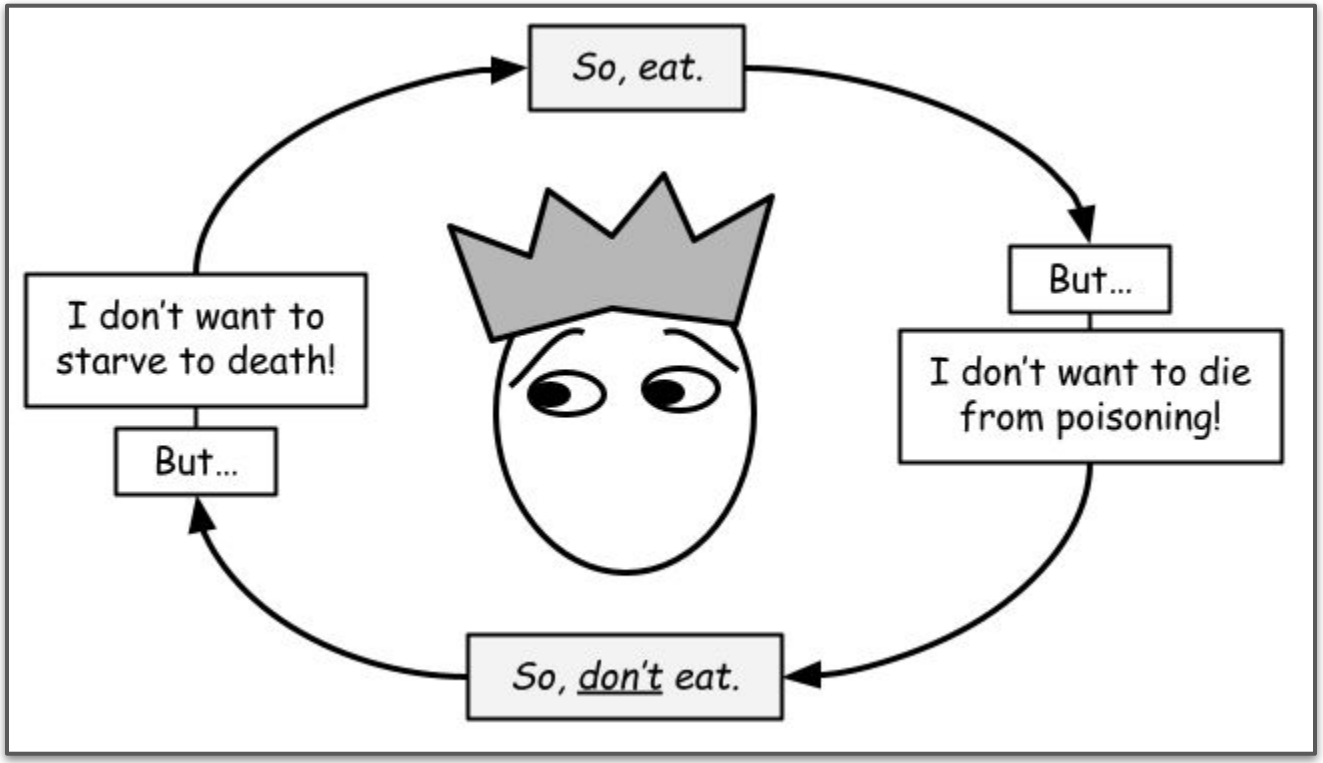
Not a fluke

Pattern: The “Hungry King”

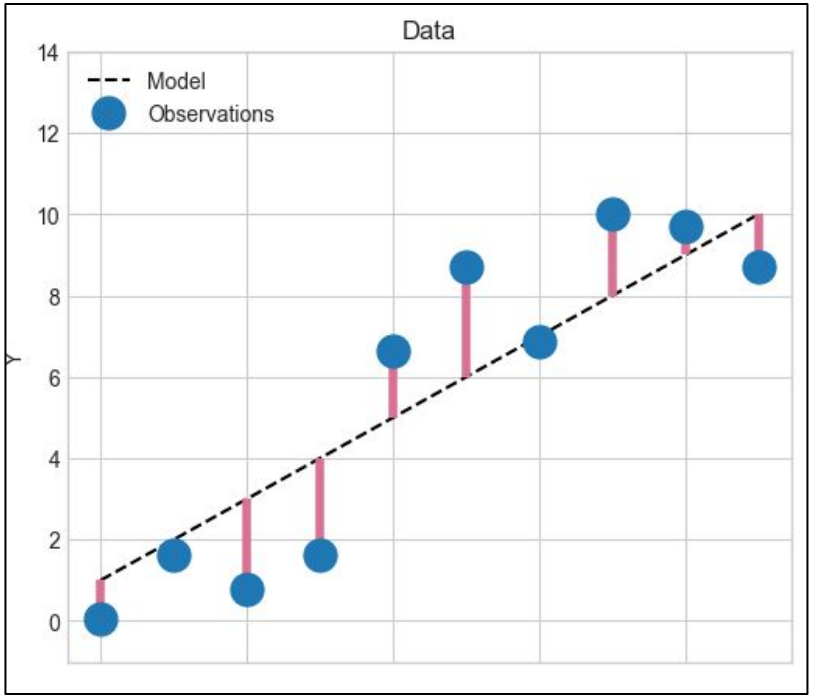


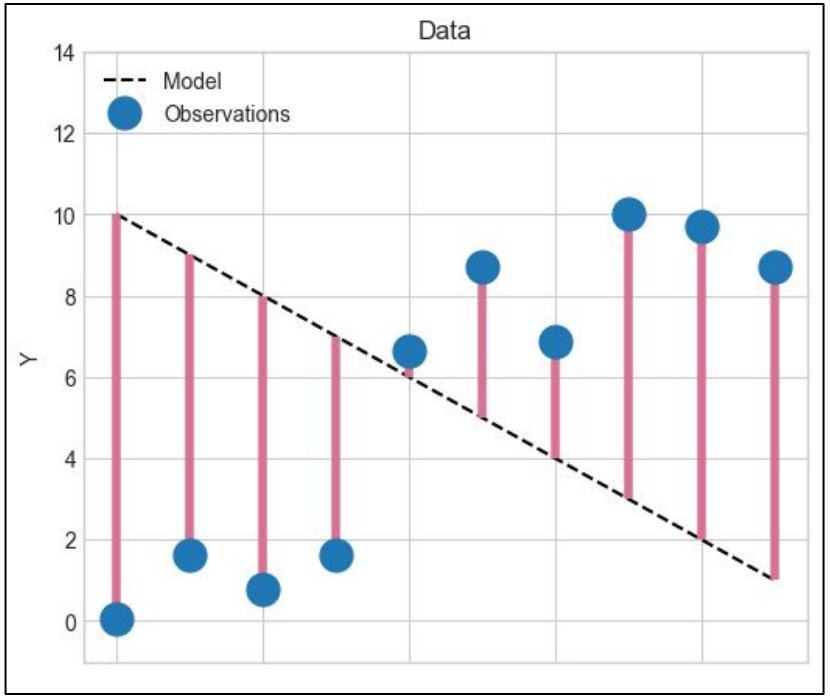
The Hungry King's Many Faces

Identifying a Hungry King

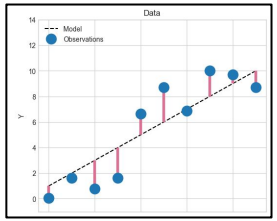


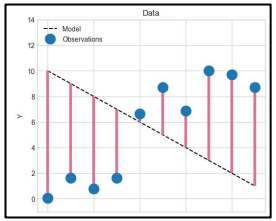
Loss Functions: Defining Good and Bad Fits

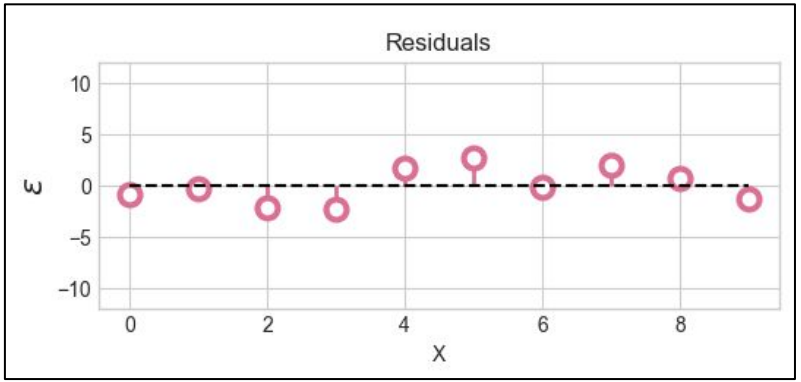


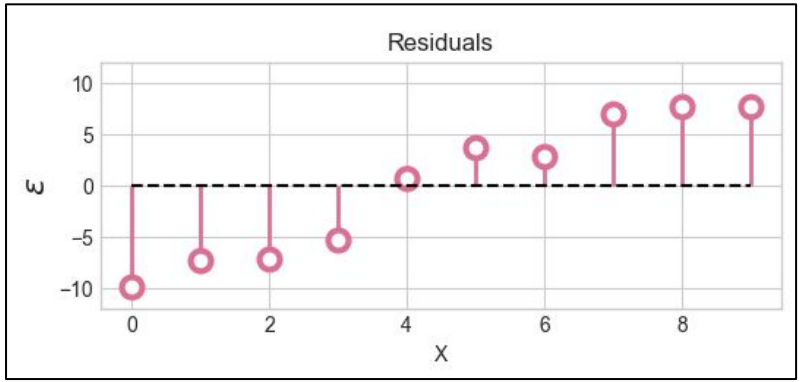


Loss Functions: Sum the errors?





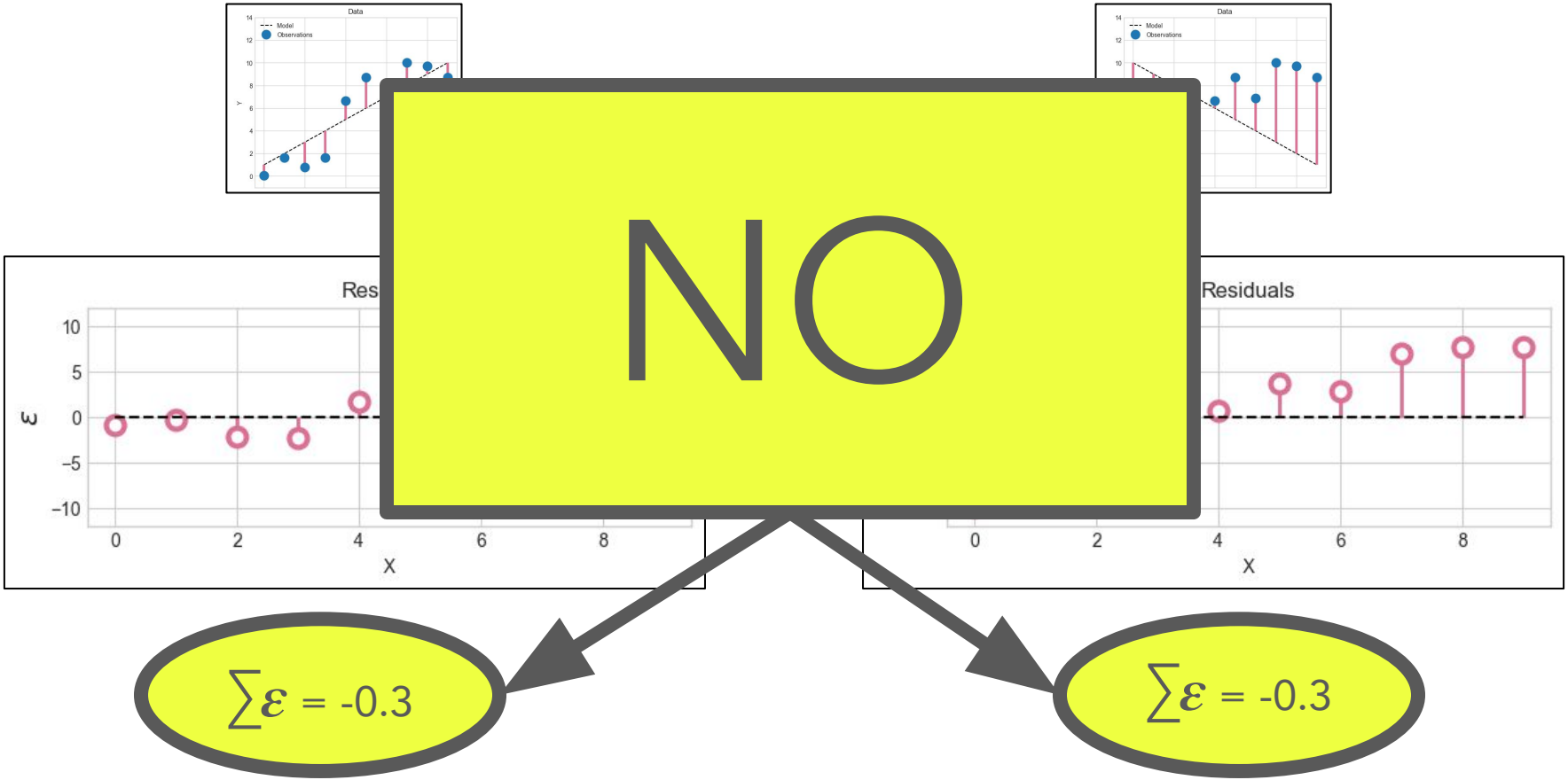




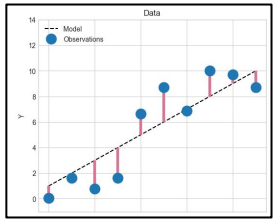
$\sum \varepsilon$ = -0.3

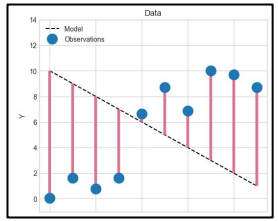
$\sum \varepsilon$ = -0.3

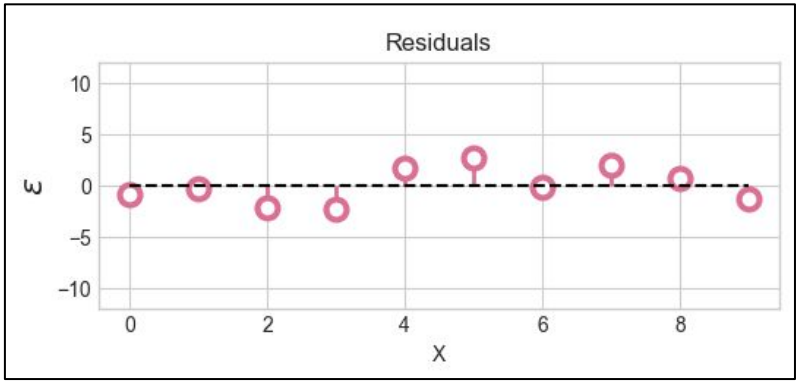
Loss Functions: Sum the errors?

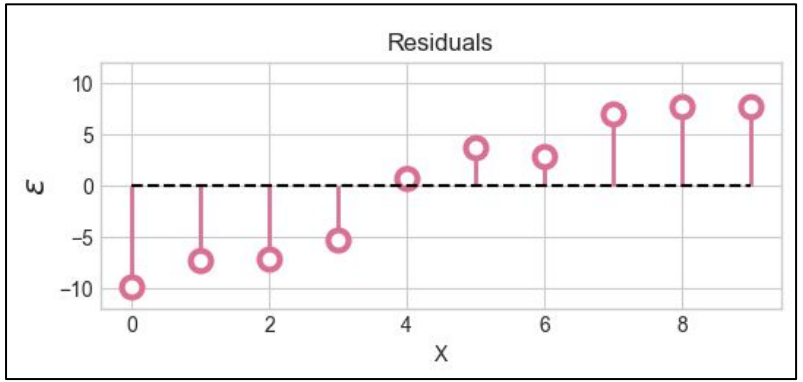


Loss Functions: Sum the *size* of the errors?





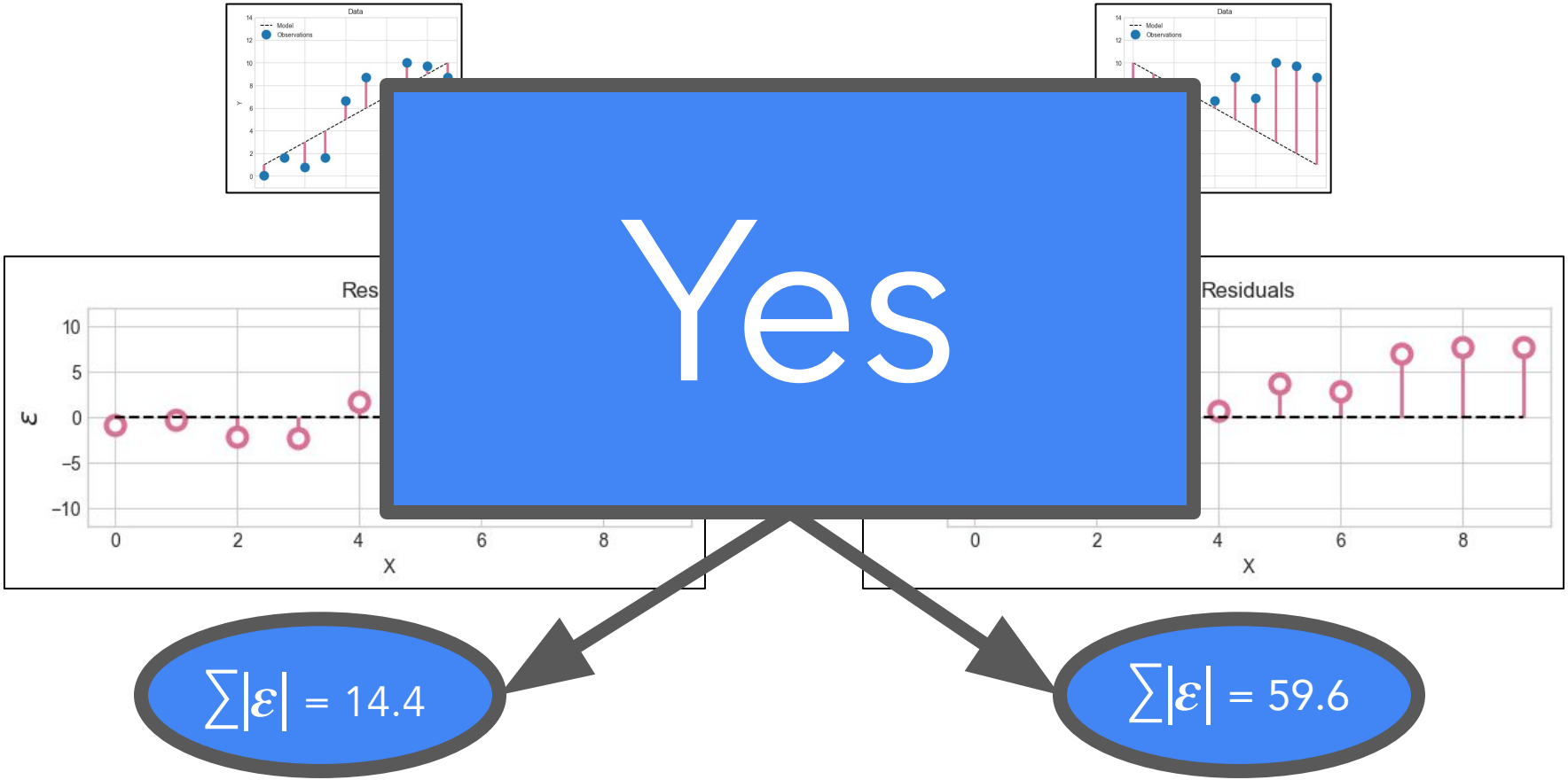




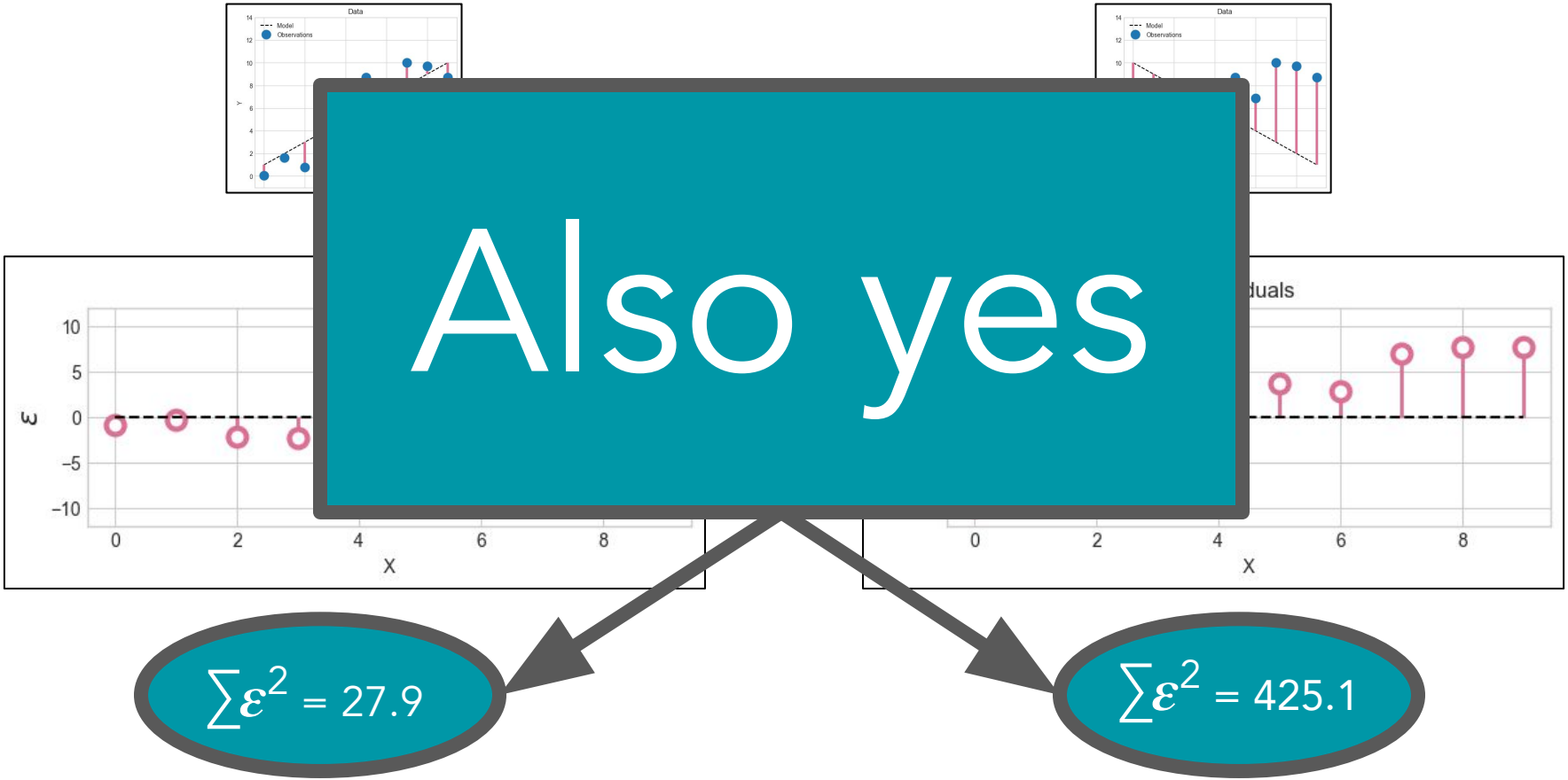
$\sum|\varepsilon| = 14.4$

$\sum|\varepsilon| = 59.6$

Loss Functions: Sum the *size* of the errors?



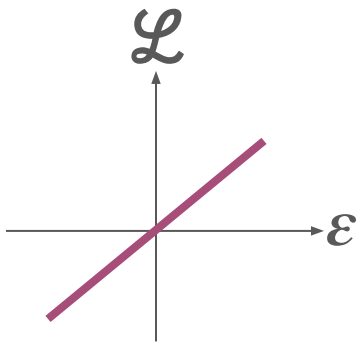
Loss Functions: Sum the *squares* of the errors?



Loss Function Requirements

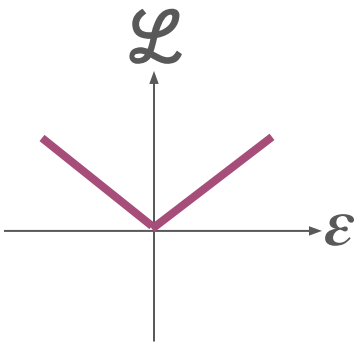
No

$$\mathcal{L} = \sum \varepsilon$$

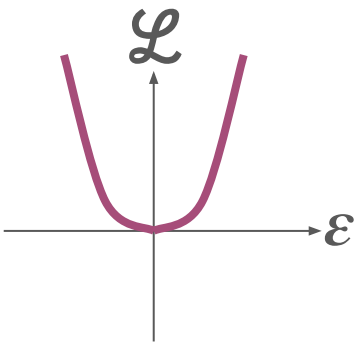


Yes

$$\mathcal{L} = \sum |\varepsilon|$$



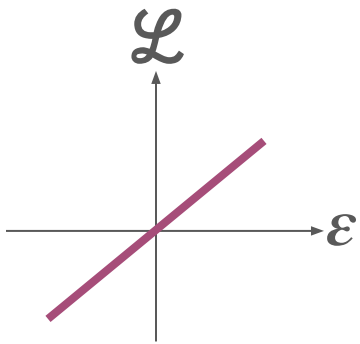
$$\mathcal{L} = \sum \varepsilon^2$$



Loss Function Requirements

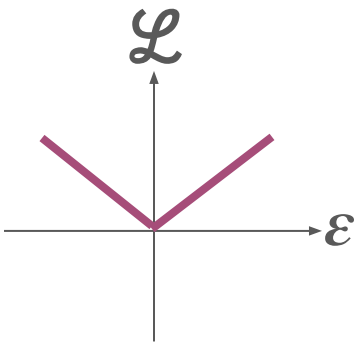
Odd

$$\mathcal{L} = \sum \varepsilon$$

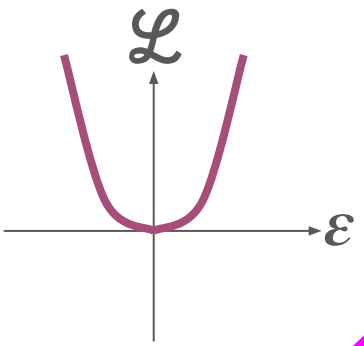


Even

$$\mathcal{L} = \sum |\varepsilon|$$



$$\mathcal{L} = \sum \varepsilon^2$$



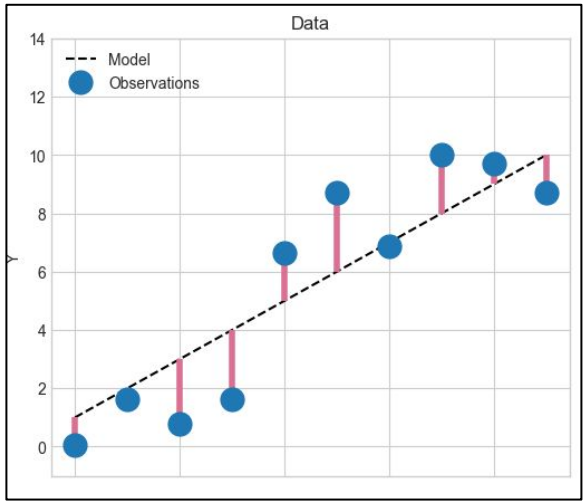
Loss Function → Model Fit

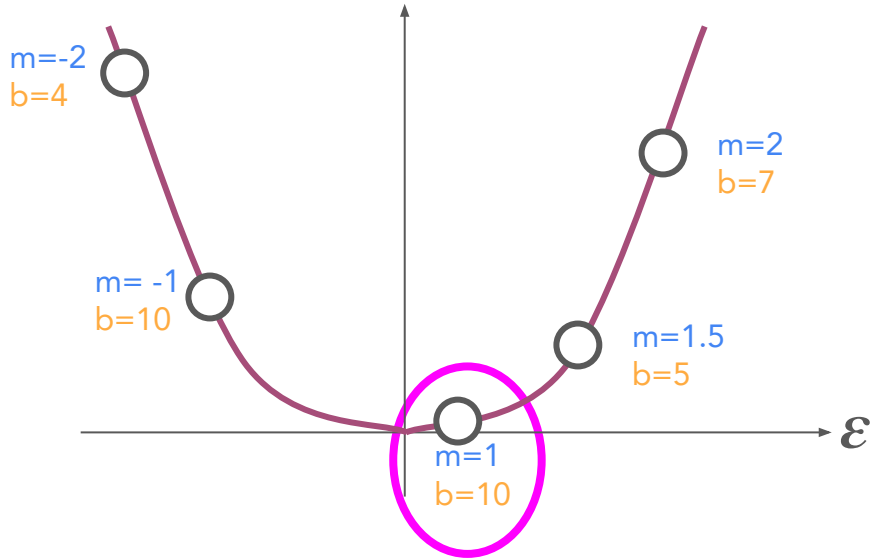
Model:

$y = mx + b$

Loss Function:

$$\mathcal{L} = \sum \varepsilon^2$$



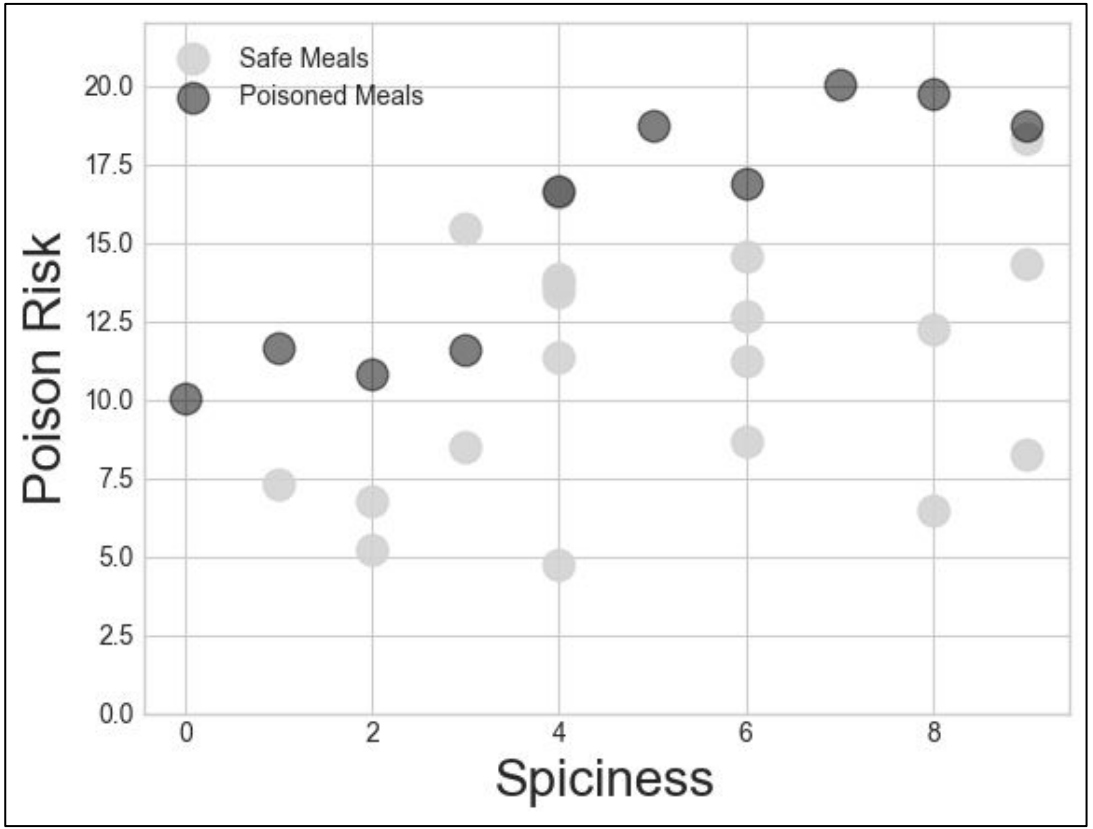


Return of the Hungry King

When can We stop feeling hungry and scared?

Soon!
We've found a relationship between a meal's spiciness and risk of being poisoned.

Show Us now!!

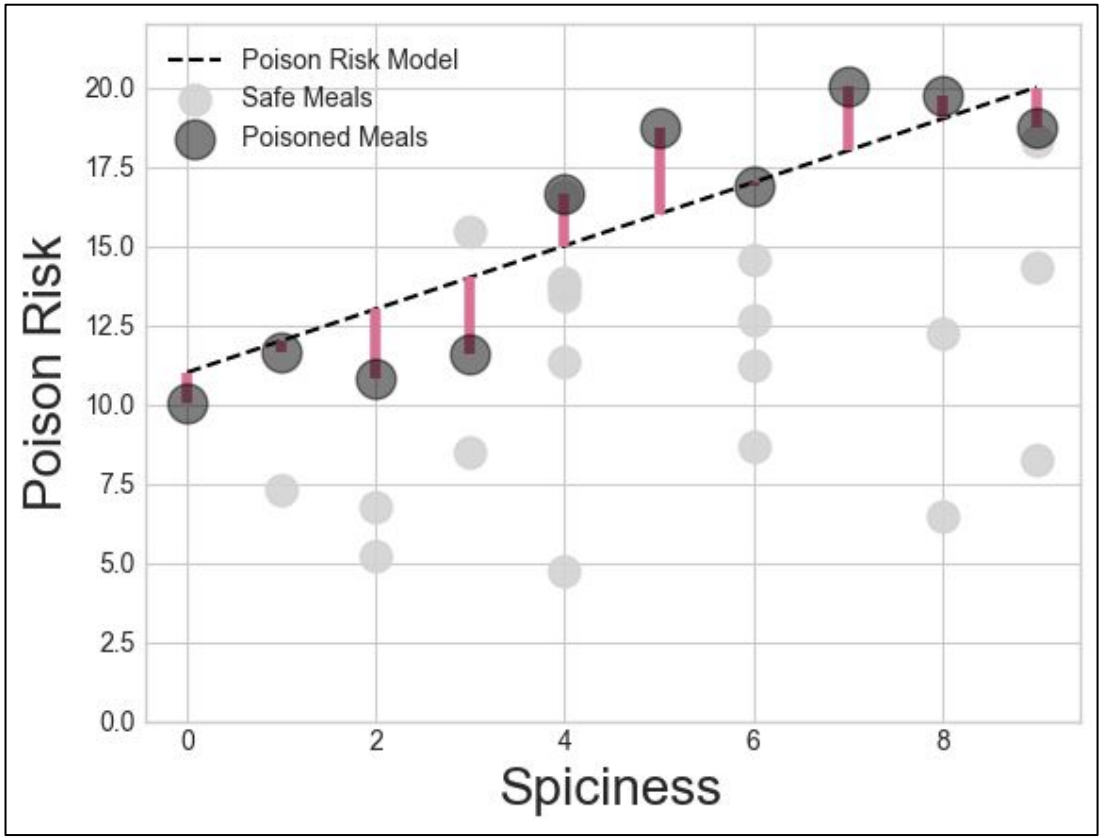


Typical Loss Function Use

Minimal loss!

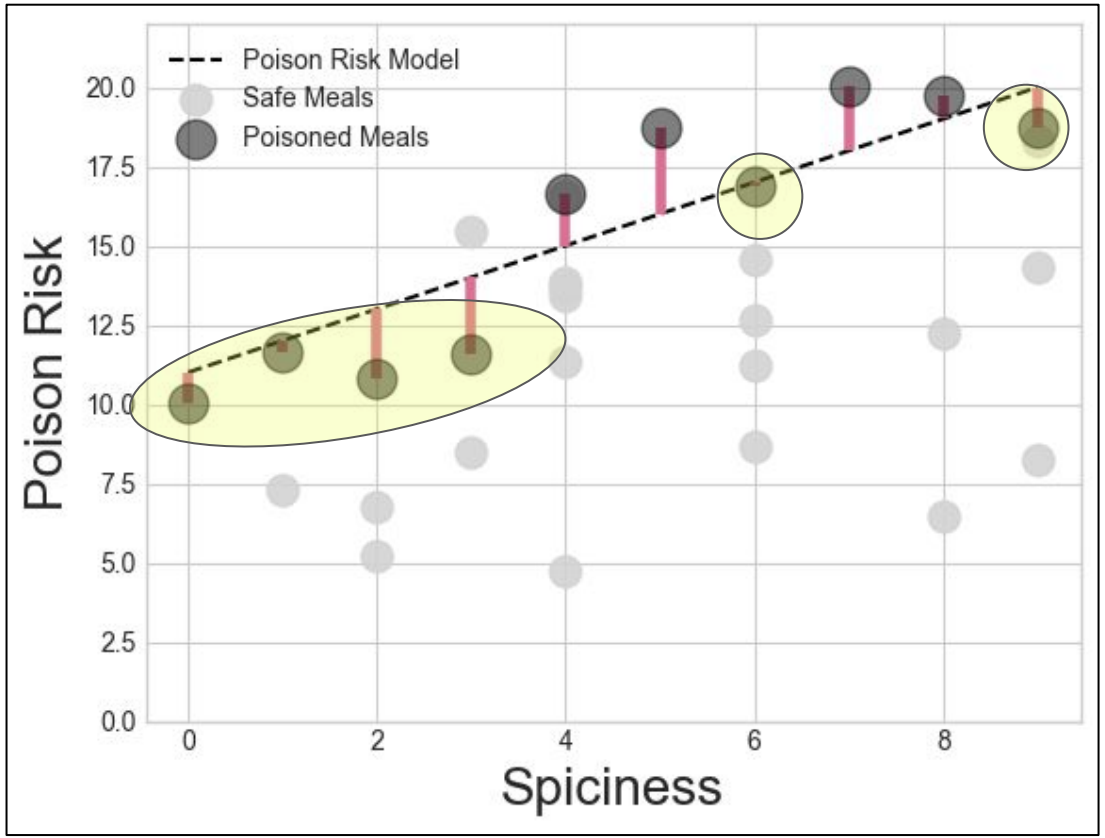
$$\mathcal{L} = \sum \varepsilon^2 = 27.9$$

...



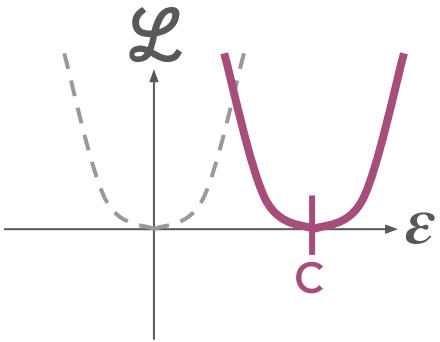
Typical Loss Function Problems

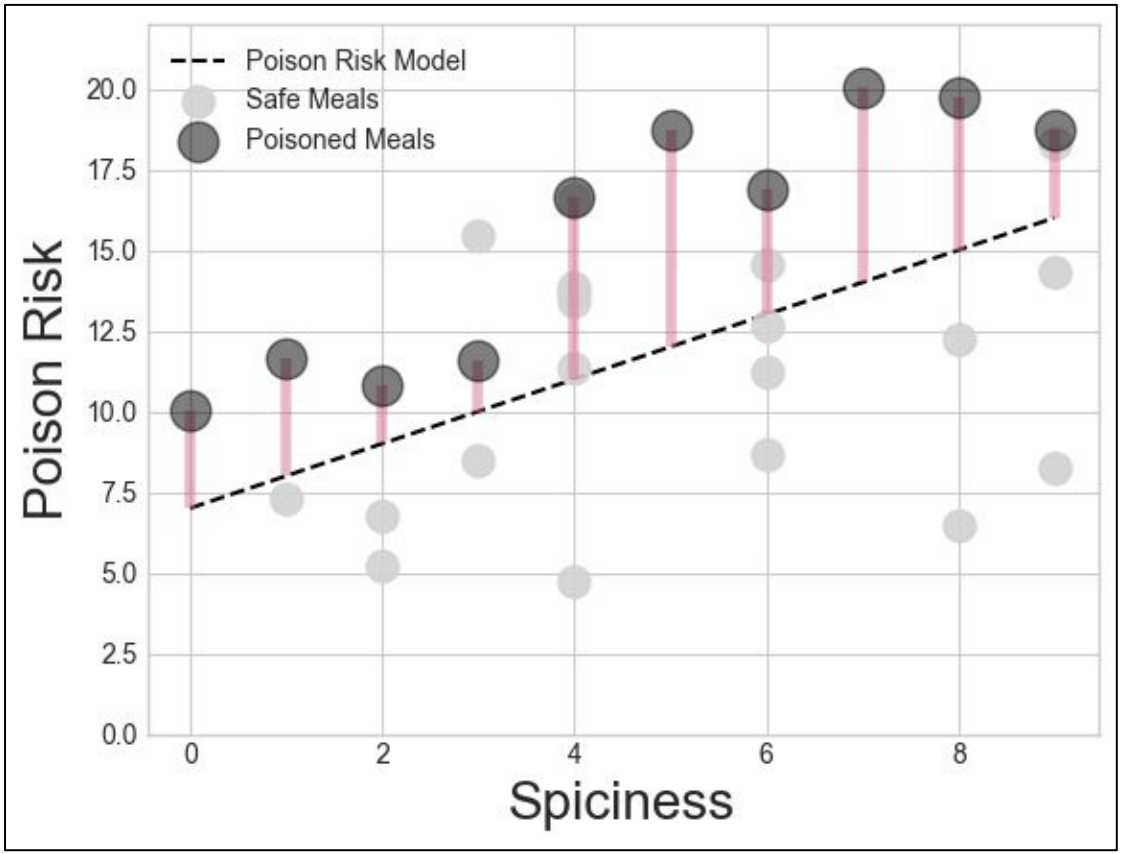
Off with your head!



Typical Loss Function Kludges

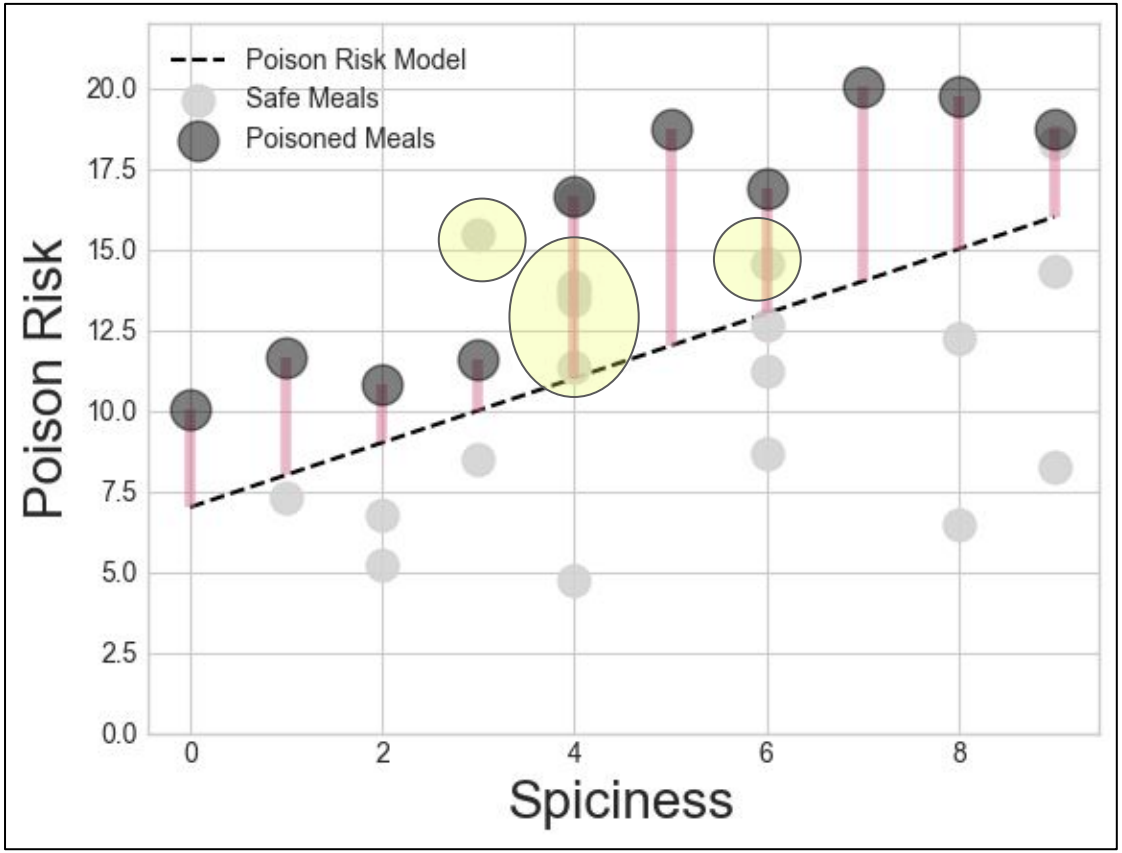
$$\mathcal{L} = \sum(\varepsilon - c)^2$$



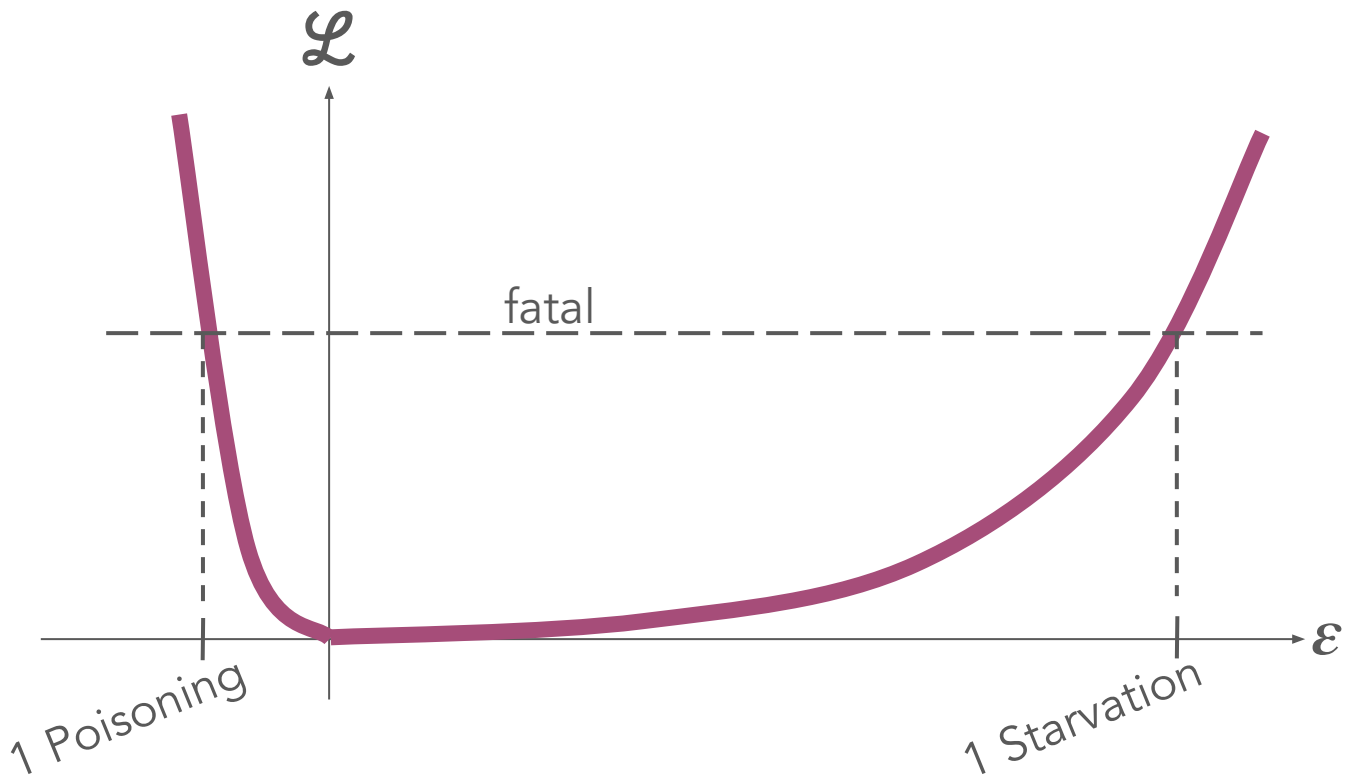


Typical Loss Function Kludges

Are you trying to starve Us?!

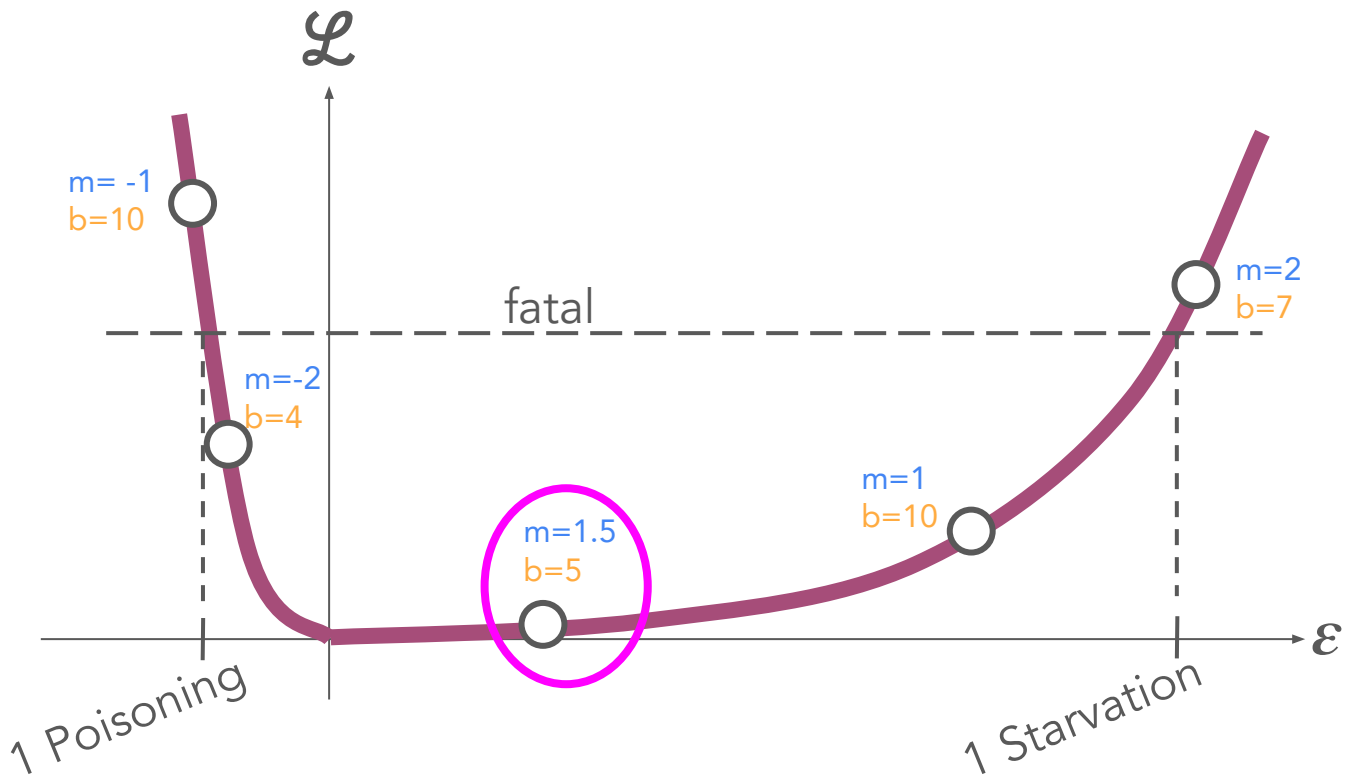


Custom Loss Functions



Interview his Majesty to quantify his fears.

Custom Loss Functions

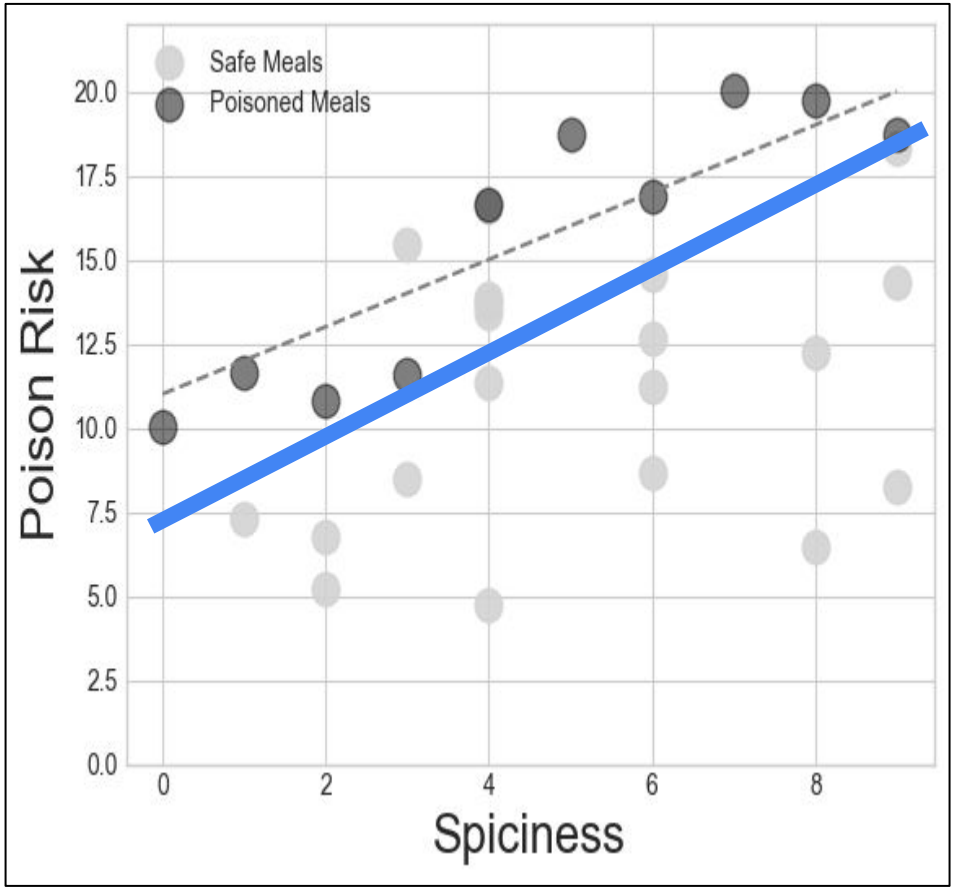


Custom Loss Functions

It's mathematically impossible to do better, unless your risk tolerances change.

Alas!
We must just accept
Our mortality, then.

Or only eat gruel We prepare Ourselves



Custom Loss Functions: Tips, Tricks, and Caveats

- A Sometimes Solution
- Not included in `scikit-learn`
  - Needs bespoke code
  - More complex
- Loss function must be:
  - Continuous
  - Concave
  - Differentiable (with respect to model parameters)
    - Ex: Autoregressive models = tricky
- Fits can be slow
- Don't stress about getting the loss function *exactly* right.
- *Customers/stakeholders build the loss function.*

What Just Happened

- Defined the “Hungry King” dilemma
- Reviewed loss functions’ influence on model fit & performance
- How to translate end users’ latent requirements into better model fits
- Tips, tricks, and caveats

Thank you!
Questions?