A couple notes on MATLAB (& HW)

If you want to see output as you go:

disp()

sprintf() ← include formatting options

fprintf()

goes to command line

If you do not see enough digits: format ~~long~~ short

Relative error

⇒ gives you approximate # of digits of accuracy

⇒ to automate exactly, try using round()

rounds to nearest integer

Models & data

→ models have parameters

e.g. $N(t+dt) = N(t)\eta$ ← probability an atom survives over dt timestep

e.g. $A + B \xrightarrow{\alpha} C$ ← reaction rate

→ does model capture/summarize the data?

⇝ what choice of parameters optimizes this?

Ex: maximum likelihood estimate for exponential decay

Recall: our model is

$$N(t+dt) = N(t)\eta \quad \leftarrow \; = 1 - \delta$$

$\boxed{Q}$ How do we estimate true η?

How do we know it is "good"?

⇒ we want η which maximizes the likelihood that, if you used it in the model, you would get the data you observe

$$\underbrace{p(\text{data}|\eta)}_{\text{joint prob function}} = \prod_{t=0}^{T-1} \underbrace{\binom{N(t)}{N(t+1)}}_{\text{choice of atoms that survive}} \left(\underbrace{\eta^{N(t+1)}}_{\text{probability they survive}} \cdot \underbrace{(1-\eta)^{N(t)-N(t+1)}}_{\text{probability the rest decay}} \right)$$

$$= L(\eta)$$

$\Rightarrow$ want

$$\left.\begin{aligned} \frac{d}{d\eta}(L(\eta)) &= 0 \\ \frac{d^2}{d\eta^2}(L(\eta)) &< 0 \end{aligned}\right] \text{ defines a maximum point}$$

notice:

$$\frac{d}{d\eta}(\log(L(\eta))) = \frac{1}{L(\eta)} \frac{dL}{d\eta} \Rightarrow \text{same root}$$

$$\log(L(\eta)) = \sum_{t=0}^{T-1} \log\binom{N(t)}{N(t+1)} + \log\left(\eta^{N(t+1)}\right) + \log\left((\eta-1)^{N(t)-N(t+1)}\right)$$

$$= \sum_{t=0}^{T-1} \log\binom{N(t)}{N(t+1)} + N(t+1)\log\eta + (N(t)-N(t+1))\log(\eta-1)$$

$$\frac{d}{d\eta}(\log(L(\eta))) = \sum_{t=0}^{T-1} \frac{N(t+1)}{\eta} + \frac{N(t)-N(t+1)}{\eta-1} = 0$$

$$\Rightarrow \quad \sum N(t+1)(\eta) - \sum N(t+1) = \sum (N(t+1)-N(t))(\eta)$$

$$\Rightarrow \quad \sum N(t)\cdot\eta = \sum N(t+1)$$

$$\Rightarrow \quad \boxed{\eta^* = \frac{\sum_{t=0}^{T-1} N(t+1)}{\sum_{t=0}^{T-1} N(t)}} \rightsquigarrow \delta^* = 1-\eta^*$$

$$= \frac{\sum N(t)-N(t+1)}{\sum N(t)}$$

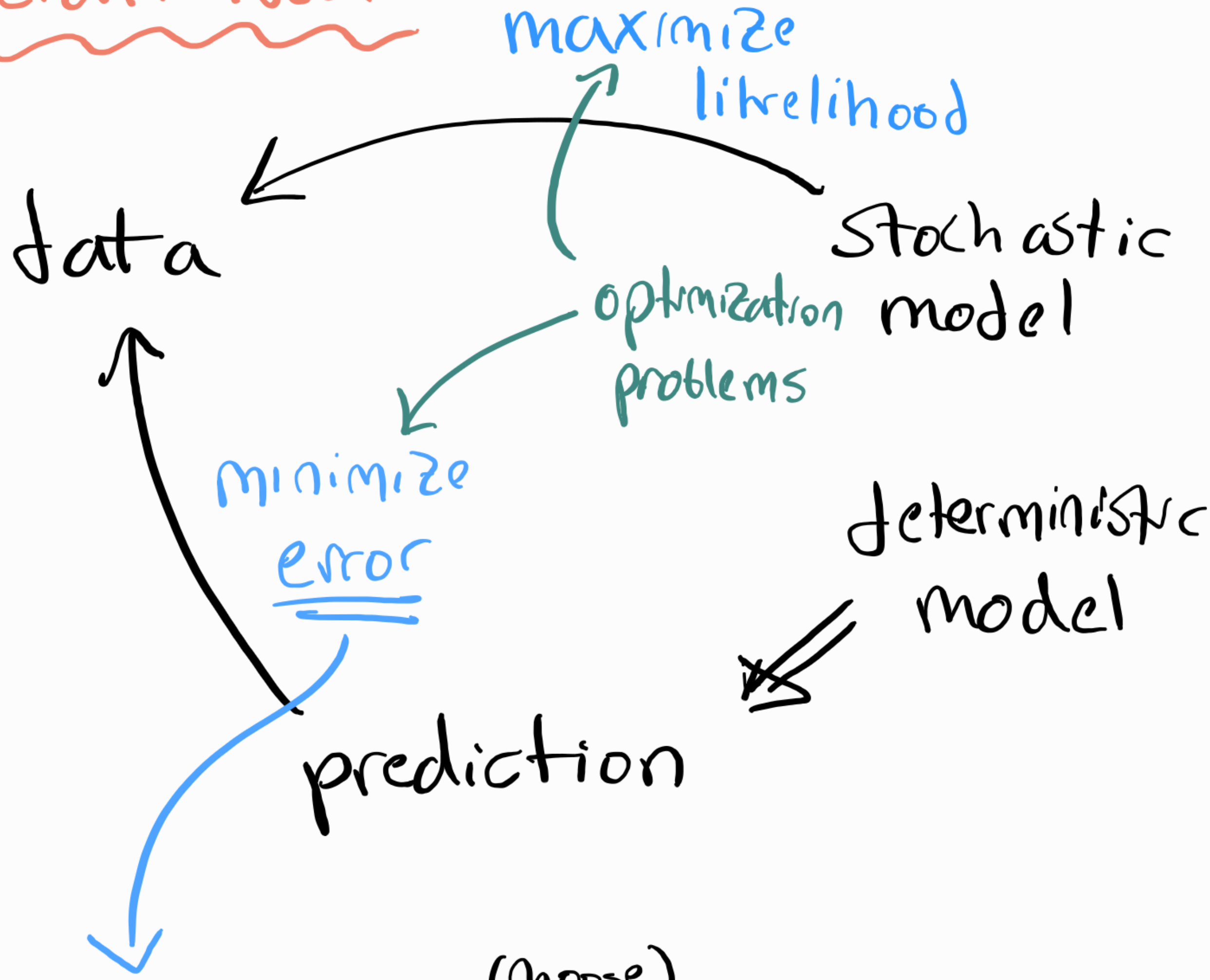
maximum likelihood estimate

as in previous lecture

$$= \frac{N(0)-N(T)}{\sum_{t=0}^{T-1} N(t)}$$

note: robust to experiment length

Overall idea



(choose)
Q How do we define error (functions)?

→ just like when choosing a model, choosing an error measure requires careful thought (context, experimental error, etc)

Ex: Summarizing data (0-dim regression)

suppose you have the (hourly) incomes of 7 people

data: \$10, \$12.50, \$12.50, \$113.00, \$18.80, \$273.00, \$156.00

⇝ from basic statistics, we know

average $= \frac{\sum I_n}{7} = \$85.11$

median $= \$18.80$

minimax $= \$141.50$

mode $= \$12.50$

if you had to summarize this data with one income, what would it be?

$\boxed{Q}$ Pros & Cons?

Lp error: $E_p^p(s) = \sum_{n=1}^{7} |I_n - s|^p$

"summary" of data

⇓

want s_p which minimizes error $E_p^p(s)$

$$\leadsto \frac{dE_p^p}{ds} = \sum p(I_n - s)|I_n - s|^{p-2}$$

↪ want $\frac{dE_p^p}{ds} = 0$ & $\frac{d^2E_p^p}{ds^2} > 0$ [minimizer]

concave up

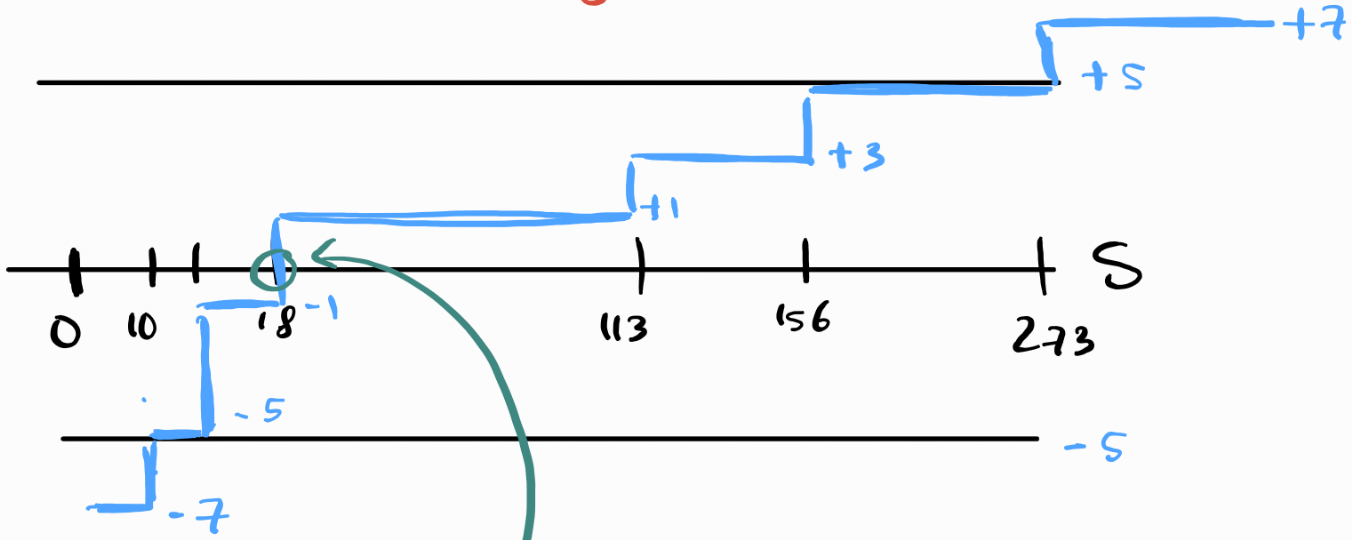
L_1 error: $E_1'(s) = \sum_{n=1}^{7} |I_n - s|$

$$\Longrightarrow \frac{dE_1'}{ds} = \sum_{n=1}^{7} \frac{I_n - s}{|I_n - s|} = \sum_{n=1}^{7} \text{sign}(I_n - s)$$

$= 1$ if $I_n > s$

$= -1$ if $I_n < s$

$\boxed{Q}$ What does $\frac{dE}{ds}$ look like?



Notice: insensitive to magnitude of outliers

$$\frac{dE}{ds} = 0 \text{ at } s = \$18.80$$

$\Longrightarrow \frac{dE_1'}{ds}$ gives us exactly the median

$\rightsquigarrow \boxed{Q}$ Why?

Notice

$$E_1'(s) = \sum_{n=1}^{7} |I_n - s| = \left[|10 - s| + |273 - s|\right]$$
$$+ \left[|12.5 - s| + |156 - s|\right]$$
$$+ \left[|12.5 - s| + |113 - s|\right]$$
$$+ |18.8 - s|$$

$s \in [10, 273]$

$s \in [12.5, 156]$

$s \in [12.5, 113]$

$s = 18.8$

L_2 error: $E_2^2(s) = \sum_{n=1}^{7} (I_n - s)^2$

We know this is concave up

$$\frac{dE_2^2}{ds} = \sum 2(I_n - s) = 0$$

$$\Rightarrow \sum_{n=1}^{N} I_n - \sum_{n=1}^{N} s = 0$$

$= N \times s$

$$\Rightarrow s = \frac{1}{N} \sum_{n=1}^{N} I_n$$

$\Leftarrow$ average

notice: each data point carries equal weight

L_∞ error: $E_p^p(s) = \sum |I_n - s|^p$

$$\rightsquigarrow E_p(s) = \sqrt[p]{\sum |I_n - s|^p}$$

Let m be s.t. $|I_m - s| = \max_n |I_n - s|$

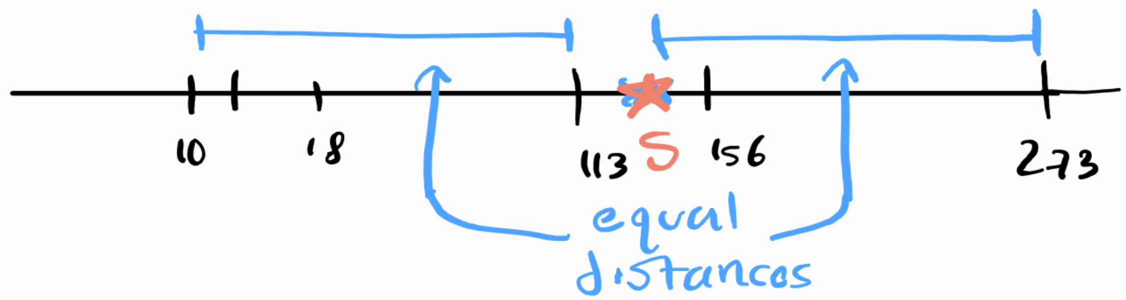
Then,

$$E_p(s) = \left(|I_m - s|^p \sum \underbrace{\left(\frac{|I_n - s|}{|I_m - s|} \right)^p}_{\le 1} \right)^{1/p}$$

as $p \to \infty$,

$$E_p(s) \to \left(|I_m - s|^p \times (1 + 0 + 0 + \dots) \right)^{1/p}$$

$$E_\infty(s) = |I_m - s| = \max_n |I_n - s|$$



notice: sensitive to outliers

L_0 error:

as $p \to 0$, $E_p^p(s) \to \sum_n \lim_{p\to 0} |I_n - s|^p = \begin{cases} 1 & I_n = s \\ 0 & \text{else} \end{cases}$

$\Rightarrow E_0(s) =$ # points matching s

= mode of data = \$12.50 in this case

notice: not useful in near continuous data, only in discrete cases

Ex: summarizing a relation (1-dim regression)

Suppose we have N-data points

$\left\{A, \frac{dA}{dt}\right\}_j$ for reactions $\begin{cases} A \xrightarrow{\tilde{a}} B \\ C \xrightarrow{\gamma} A \end{cases}$

($\frac{dA}{dt} = y$)

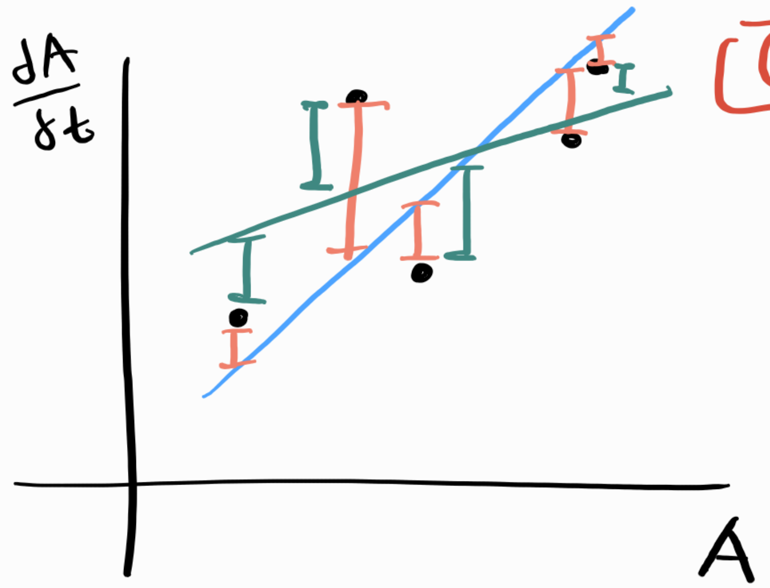
$\Rightarrow$ our model: $\frac{dA}{dt} = -\tilde{a}A + \gamma C = aA + b = y$

What (a,b) choices best fit our data?

L_p error:

$$E_p^p = \sum_{j=1}^{N} |y(A_j) - y_j|^p$$

$$\Rightarrow E_p^p(a,b) = \sum_{j=1}^{N} |\underbrace{aA_j + b - y_j}_{\text{vertical distance between prediction \& data}}|^p$$



[Q] How might we interpret different E_p^p error minimizers?

$E_\infty \to$ line such that largest deviation is minimal

$E_0 \to$ line which passes through the most points (sort of)

$E_1, E_2 \to E_2$ penalizes larger deviations more than E_1, "equal" contributions

L_2-error:

$$E_2^2(a,b) = \sum_{j=1}^{N} (aA_j + b - y_j)^2$$

want both $\partial_a E_2^2(a,b) = 0$ and $\partial_b E_2^2(a,b) = 0$

$$\partial_a E_2^2(a,b) = \sum 2A_j(aA_j + b - y_j) = 0$$

$$\Rightarrow 2a\sum A_j^2 + 2b\sum A_j = 2\sum A_j y_j$$

linear!

$$\partial_b E_2^2(a,b) = \sum_{j=1}^{N} 2(aA_j + b - y_j) = 0$$

$$\Rightarrow 2a\sum A_j + 2bN = 2\sum y_j$$

linear:

linear system

$$\begin{bmatrix} \sum A_j^2 & \sum A_j \\ \sum A_j & N \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum A_j y_j \\ \sum y_j \end{bmatrix} \quad (*)$$

$\Longrightarrow$ solve for a, b to get L_2 minimizer

L_1-error:

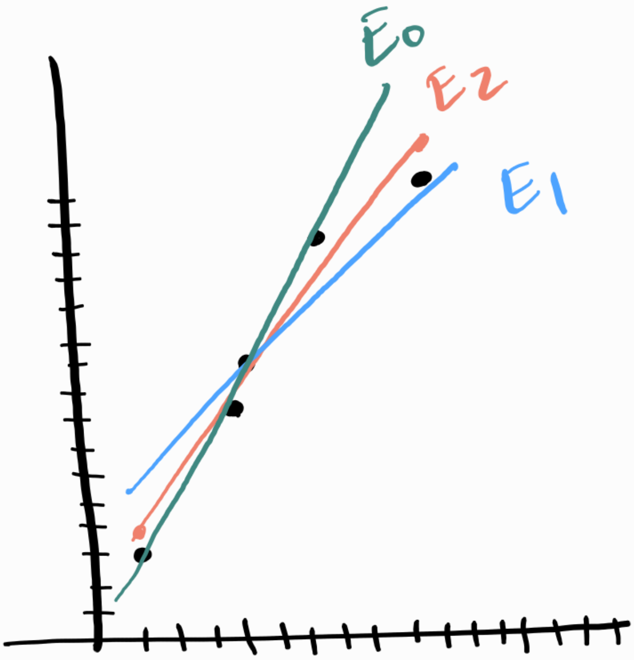
$$E_1(a,b) = \sum_{j=1}^{N} |aA_j + b - y_j|$$

$$\Downarrow \begin{cases} \partial_a E_1 = \sum_{j=1}^{N} \operatorname{sign}(aA_j + b - y_j) \cdot A_j \\ \partial_b E_2 = \sum_{j=1}^{N} \operatorname{sign}(aA_j + b - y_j) \end{cases}$$

example

$y = 2A + 1 \leftarrow E_0$ minimizer

$(A, y) = (1,3), (3,8), (4,9), (6,13), (9,15)$



E_2 minimizer

solving (*) gives

$$a = \frac{46}{31} \approx 1.484$$

$$b = \frac{86}{31} \approx 2.774$$

E_1 minimizer

$$a = \frac{6}{5} = 1.2$$

$$b = \frac{21}{5} = 4.2$$

Quiz

Suppose you survey your friends about their favorite number, getting data:

$$\{3, 9, 3, 13, 2736, 0, \pi, 0.08\}$$

(a) What is the minimizer of L_1 error?

(b) What is the minimizer of L_2 error?

(c) What is the minimizer of L_4 error?

[approximately. ok to use computer]

(d) What is the minimizer of L_∞ error?

(e) Which minimizer do you believe best summarizes the data, in your opinion? Why?

Show your work. Explain your steps.

(a) median $3.071 = \frac{3+\pi}{2}$

(b) mean 345.9

(c) ≈ 942.176

(d) minimax 1368

(e) subjective