CS2810 Day 1 Jan 18 (sec 3)

Data Models
- what does it take to build a good data model?

Admin

Linearity

Gauss Jordan Elimination

Modelling question:
What is the probability that at least one person in class right now has covid and is contagious?

Assume:

- 100 students in this class
- prob contracting covid is uniform per person
- prob students have covid are each independent of all others
- one who contracts covid is contagious for 7 days

assuming p, the prob the one student is contagious with covid, is same as prevelence of contagious folks with covid in Mass

notes:
population of mass 7000k people
20k people in Mass test positive for covid in a day

$$\overbrace{1 - \underbrace{(\overbrace{1-p}^{\text{Prob 1 person not contagious}})^{100}}_{\text{Prob 100 not contagious}}}^{\text{Prob > 0 people in class have covid}} = 86\%$$

$$p = \frac{20}{7000}^{*} \cdot 7 = 2\%$$

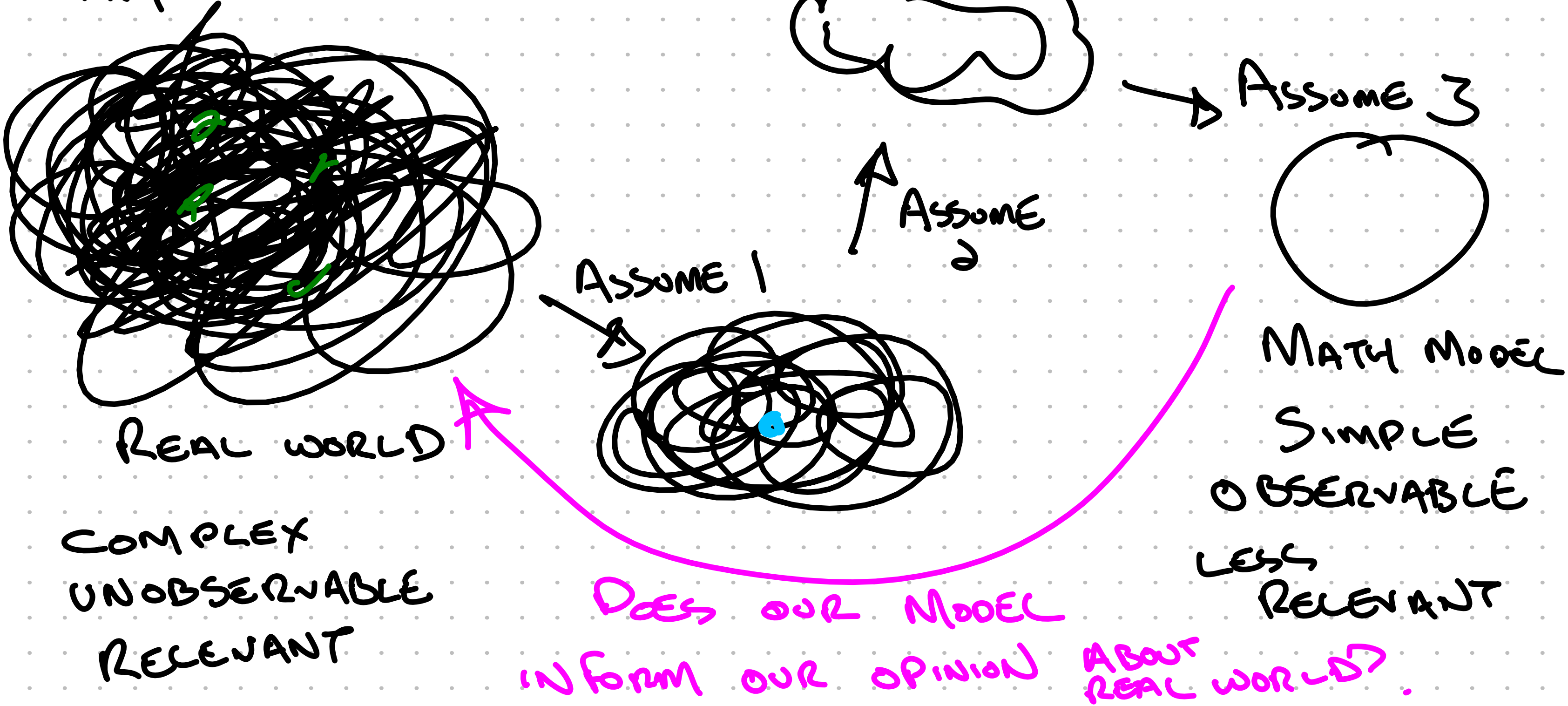
ICA A:
Critique this model in a group
- do you believe in this model's conclusion?
- provide 2-3 of the strongest critiques of this model

(prob of at least 1 person contagious is too low)
prevelence in MA doesn't account for non-tested covid cases
population @ NU might be more risk-taking
populatoin @ NU exposed to urban covid rates

(prob of at least 1 person contagious is too high)
population @ NU not represented by population @ MA
    - vax / testing requirement, quarantine protocol

(model wrong ... not necessarily too high or low)
population @ NU is more insular ("in their own bubble")
recent travel to NU not represented by MA population
    - use each individual's prevelence from where they've spent the past 7 days
each student does not contract covid uniformly, or independent

Math of Data Model



Does our Model inform our opinion about real world?

What does it take to build a "good" data model of the real world?

- a breadth of math models to choose from in one's mental library
    (we can choose the one which most appropriate for our real world target)

    - we'll share with you many models from:
        - linear algebra
        - prob and statistics

- the ability to be rigorous / creative in making and evaluating assumptions

    - we'll build a few mini applications and discuss assumption validity

- a keen sense of which aspects of the application we seek to model most accurately

    - experience

Linearity (intuitive definition: use this one to understand meaning)

A function is linear if
- scaling, applied before or after the function, has the same effect
- addition, applied before or after the function, has the same effect

$$\alpha \in \mathbb{R} \qquad x \in \text{DOMAIN}(f)$$

$$f(\alpha x) = \alpha f(x)$$

SCALE BEFORE FNC

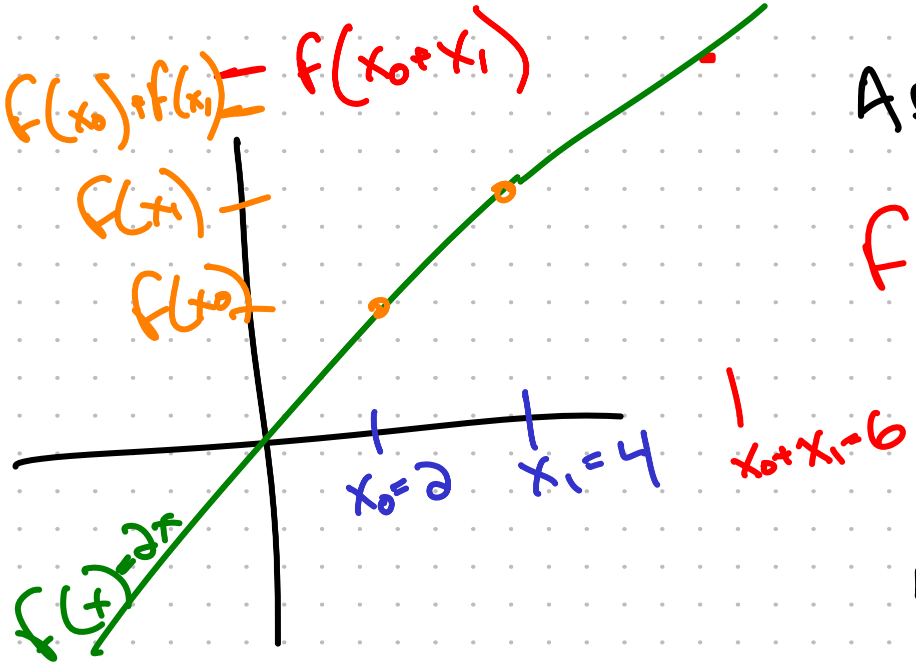
SCALE AFTER FNC

$$x, y \in \text{DOMAIN}(f)$$

$$f(x+y) = f(x) + f(y)$$

ADD BEFORE FNC

ADD AFTER FNC



ADD BEFORE APPLY FNC

$$f(x_0+x_1) = f(2+4)$$

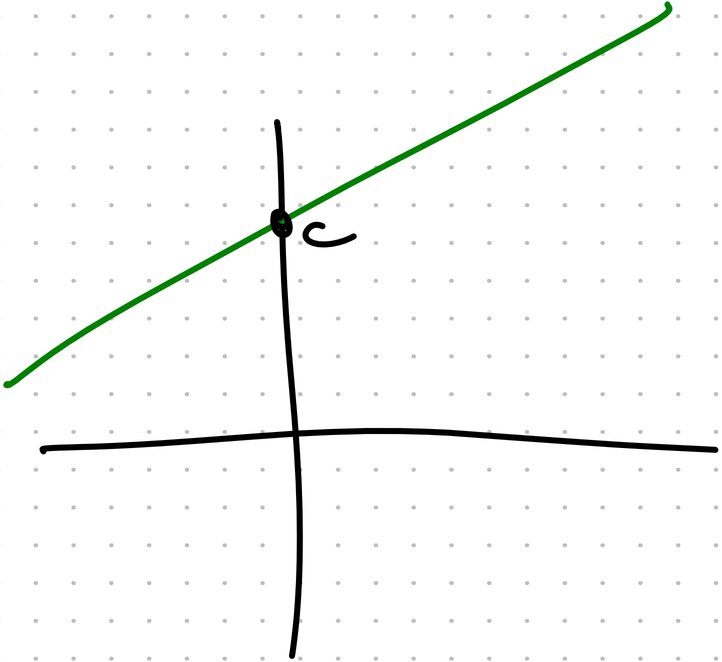
$$= f(6)$$

$$= 12$$

ADD AFTER APPLY FNC

$$f(x_0) + f(x_1)$$

$$= f(2) + f(4) = 4+8 = 12$$



$f(x)=2x+1 \rightarrow$ NOT LINEAR

$f(0)=1 \qquad f(1)=3$

$f(0\underset{\uparrow}{+}1)=3$

ADD BEFORE

$f(0)\underset{\uparrow}{+}f(1)=1+3=4$

ADD AFTER

$$f(0)+f(1)\neq f(0+1)$$

Linearity (working definition: use this one to show a function is linear or not)

INPUTS TO F

A FUNCTION F IS LINEAR IF

FOR ALL $\alpha, \beta \in \mathbb{R}$ AND $x, y \in \text{DOMAIN}(f)$

SCALARS

$$f(\alpha x + \beta y) = \alpha f(x) + \beta f(y)$$

$f(x) = 10x$ IS IT LINEAR?

CHOOSE $\alpha, \beta \in \mathbb{R}$ AND $x, y \in \text{DOMAIN}(f)$

$f(\alpha x + \beta y)$

$$= 10(\alpha x + \beta y)$$

$$= \alpha \cdot 10 \cdot x + \beta \cdot 10 \cdot y$$

$$= \alpha \cdot f(x) + \beta \cdot f(y)$$

NON-LINEAR EXAMPLE

$f(x) = x^2$

$\alpha = 1 \quad \beta = 2 \quad x = 3 \quad y = 4$

$$f(\alpha x + \beta y) = f(1\cdot 2 + 3\cdot 4)$$
$$= f(2 + 12)$$
$$= 196$$

$$\alpha f(x) + \beta f(y)$$
$$= 1 f(3) + 2 \cdot f(4)$$
$$= 9 + 2\cdot 16$$

SO THERE EXISTS $\alpha, \beta \in \mathbb{R}$, $x, y \in \text{DOMAIN}(f)$ w/ $f(\alpha x + \beta y) \neq \alpha f(x) + \beta f(y)$

why all the fuss about linearity?

many real wrold things are linear, many that aren't (on first glance) can be recast as linear

- with this class of equalities it is possible to either ...
  - find all the solutions of them
  - if there isn't a solution to a linear system ... we can find one which as close as possible

For example:
finding the line which passes through all green points is not possible.

because its a set of linear equalities, we can find the line which best satisfies the equalities