Math 390.4 1/29/19 Lec 1

— Syllabus

"Models" are approximations / abstractions to reality / absolute truth / systems / phenomena.

Model	Reality
Model Airplane	Airplane
Street Map	Road system
Wind tunnel	fast-moving air
"easy to heal, easy to rise..."	relative success of people

"All Models are wrong but some are useful" - George Box (famous statistician)

They are by definition, approximations

$3.141593 \neq \pi$

If you wrote =, you would be wrong in an absolute sense.

They serve a useful function

$3.141593 \approx \pi$

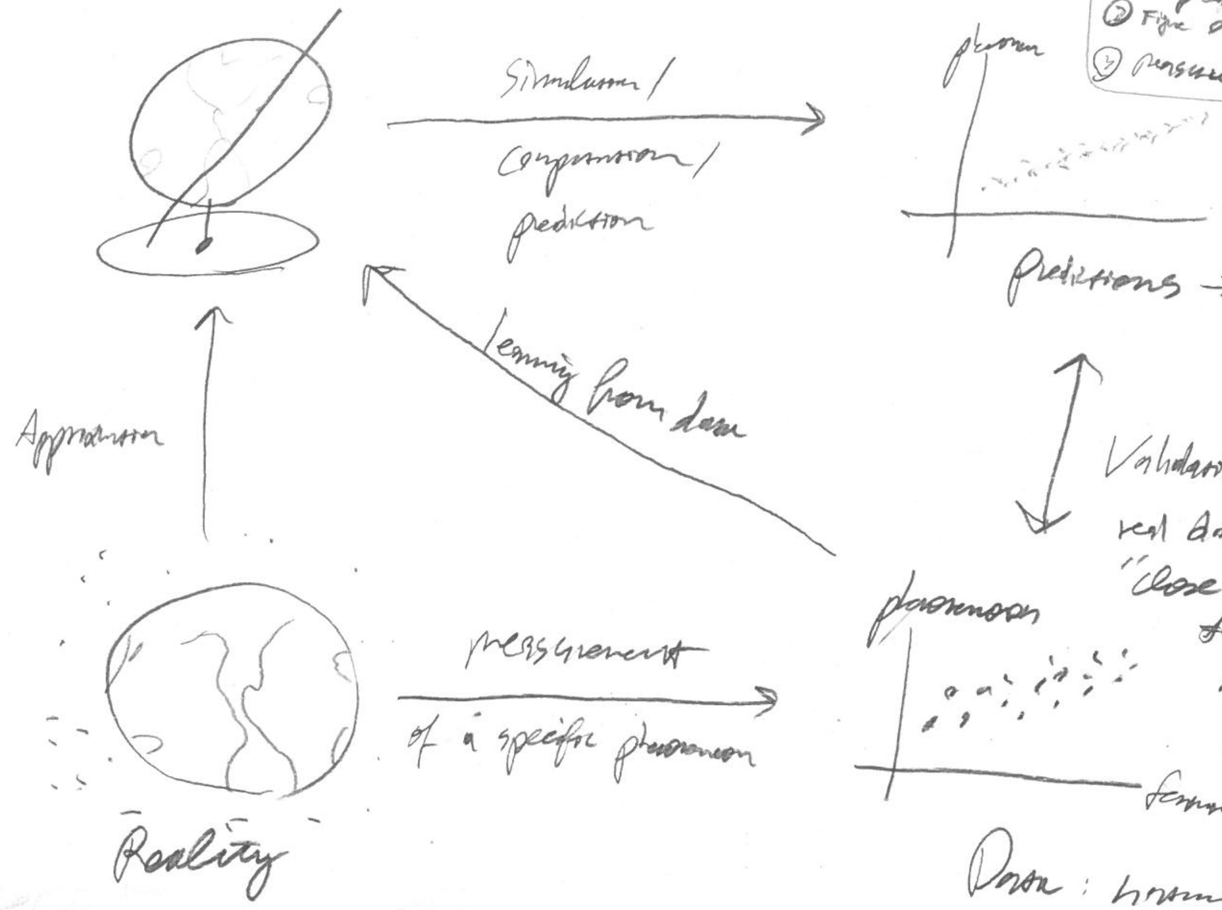
you can build bridges with that # of decimal points

Models generally useful for two goals:

1. Prediction: can the model tell us what will happen to a certain phenomenon in reality?
2. Explanation: How does reality really work? (Science)

A schematic of Models for Phenomena

First Steps
① Identify phenomena you wish to predict and/or explain
② Figure out a way to measure it
③ Measure features of the system



Predictions → look like data but are not actual data

Validation: comparing real data to predictions. If "close" ⇒ model is good for prediction! If not, maybe we can rebuild the model?

Data: numerical results of phenomenon which is measured

Consider the aphorism:

"Early to bed, early to rise, makes a man healthy, wealthy and wise"

Is this a model? Yes. What is it a model of? Health, wealth, wisdom (the phenomena)

Does it produce predictions? Explanations? Yes, Yes. What are the features of the system? bedtime, waketime

Lets try to predict. If someone goes to bed early, wakes up early, how wise will he be?? Somewhat wise!

This is a very ambiguous model! How can we make this unambiguous? Use #'s! If you use #'s, everyone knows exactly what you mean!

We need to establish "metrics" — a means to capture a phenomenon or feature of the system <u>numerically</u>

features	metric	symbol
"early to bed"	avg. bedtime in 24hr time	(b)
" " rise"	" " wakeup in 24hr time	(w)

good metrics?

1. Do they capture the feature or phenomenon adequately well? Yes
2. Are they easily readable? Yes
3. Do they have good resolution? 22:28 vs 22:32 Yes!
4. Monotonic? No. 23:58 vs 0:02

   Can we fix? Yes... measure # hrs after 5PM

   6:58 vs 7:02
5. Is this a #? No. Still not a #

   6:58 ⇒ 6.9667
   7:02 ⇒ 7.0333
6. Do they capture the phenomenon <u>very</u> well?

No... maybe measure multiple metrics? Which ones?

phenomenon	metric	symbol
health	longevity	l
wealth	Net Worth at Age 60	n
wisdom	some sort of test??	s

Creating models is hard work!!!

$$\begin{bmatrix} l \\ h \\ s \end{bmatrix} = f\left(\begin{bmatrix} b \\ w \end{bmatrix}\right)$$

phenomena are the outputs of the function

the relationship between forces and phenomena

forces of the system are the inputs to the function

Is this really an = sign??? Wait...

with properly defined inputs and outputs (by models), f is called a mathematical model.

Mathematical models are "ideas", abstractions, not concrete objects like a model airplane or a map. They are very useful! We've been building them for $\gtrsim 2000$ yr.

$a = \frac{F}{m}$; $E = mc^2$

If you believe in these models as reality... you are assuming the universe is explainable mathematically. We will assume this implicitly in this course.

Assume any result for a phenomenon y can be explained as:

$$y = t(z_1, z_2, \ldots, z_t)$$

y ← phenomenon, response, outcome, endpoint, dep. var.

$z_1, z_2, \ldots, z_t$ ← causal input information

t ← the true function that combines inputs using mathematical operators and constants

In this class, y is 1-d.

Note: the $=$ sign implies the phenomenon is perfectly explicable. Determinism vs. Random debate.

Let's examine the phenomenon y = pay back loan or not

$y \in \begin{cases} \text{pay back loan AKA creditworthy,} \\ \text{not} \ldots\ldots\ldots \text{non-} \ldots \end{cases}$

Mathematical models require #s

$y \in \{0, 1\} = \mathcal{Y}$ → output space, support of response var. By convention 1 is the "positive" class

The inputs are characteristics about a person.

What are causal inputs?

nobody define this usual e.o.s.

This is a very theoretical exercise.... Consider 3 z's:

z_1: has suff. funds to pay back loan at the time $\in \{0,1\}$

z_2: unforeseen emergency? $\in \{0,1\}$

z_3: criminal intentions $\in \{0,1\}$

$$y = t(z_1, z_2, z_3) = z_1 (1 - z_2)(1 - z_3)$$

Problems in practice:

1) Do you have z_1, z_2, z_3? No... they are inputs assessed in the future.

2) Do you know t? No... could be very complicated

Next best thing? Obtain information that approximates the information in the z's and try to combine them together to approx/model y. We call these the x's and we denote p as their dimensionality. For ex.

x_1: salary at time of loan application $\in \mathbb{R}^+$

x_2: miss loan payment or credit payment previously $\in \{0,1\}$

x_3: criminal charge in past $\in \{0,1\}$

x_j's are called features, characteristics, attributes, variables, ind. variables, regressors. Business-as-usual: use what you have or what is easy to obtain.