* R & Python

Lec 1

{MAT 342W}

2-1-21

Github is like your CV for data science or software.

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

Model	Phenomenon
model airplane	real airplane
Street map	actual roads
"early to bed, early to rise makes a man healthy, wealthy and wise"	human health, human wealth and human wisdom

"All models are wrong but some are useful" - George Box, 1984

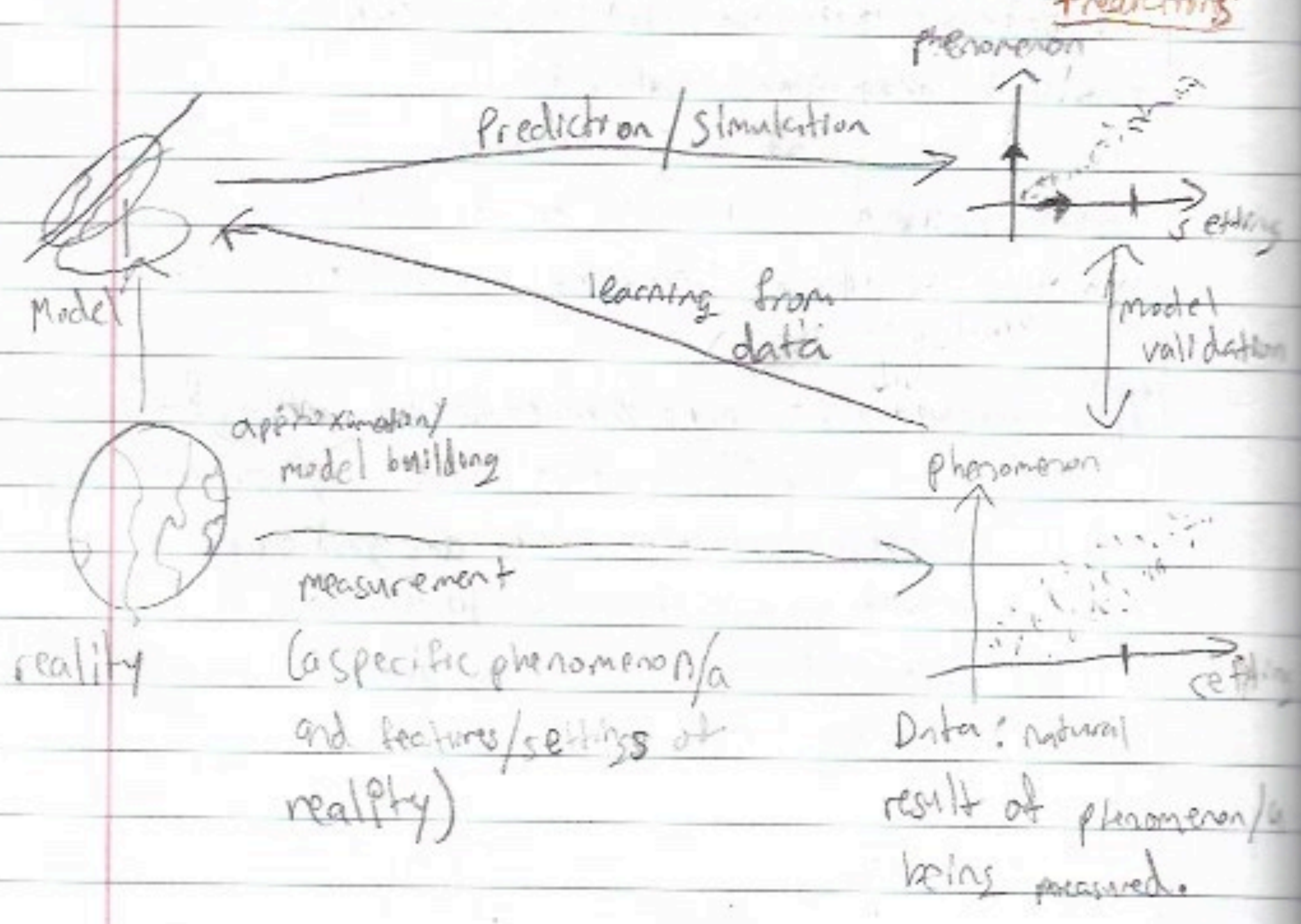
wrong: by defn approximations which are not reality

useful: are good enough to be used for a practical purpose

Models are generally used for two goals:

*(1) Prediction: can the model tell us what will happen in a certain phenomenon in a certain setting.

(2) Explanation: how does reality really work? What causes phenomena to manifest? (the purview of other courses)



Presteps to modeling:

(1) Identify a phenomenon/a you wish to predict/explain. This is your target of the modeling procedure.

(2) Figure out a way to measure it.

(3) Measure features/settings of the system/reality.

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

Phenomena: human health, wealth and wisdom (3)
Features/settings: bedtime, waketime (2)

This model is ambiguous! We don't know how to measure the settings and phenomena. In order to make this model unambiguous, we need to establish "metrics". Metrics are well-defined ways to numerically gauge phenomena/settings.

Model

Features/Phenomena	Metric	Symbol
bed time	average daily bedtime b/w ages 18-60 measured in hrs past 5pm	b
waketime	average waketime " " measured in hrs past 4am	w
health	Longevity/lifespan, QOL metric	l
wealth	net worth at time of death	n
wisdom	take a test about situations and what you would do in situations and have a panel of old people provide answers.	s

→

Mathematical Model

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

model — two settings (inputs) — three ~~phenomena~~ predictions (outputs)

* we will only build models with ONE output.

since the inputs/outputs are numerical, f is called a "mathematical model"

Mathematical models are not physical. They are themselves ideas and abstractions. But they are extremely useful! We've been building them for ~4000 yr.

i.e.: $a = F/m$, $E = mC^2$

For the purposes of this class, we will assume the universe is mathematical:

Assume: a phenomenon denoted y, can be expressed as:

$$y = t\underbrace{(z_1, z_2, \ldots, z_k)}$$

y: phenomenon, response, outcome, endpoint, dependent variable

$z_1, z_2, \ldots, z_k$: casual inputs: the true drivers of the phenomenon. In reality we don't know them.

Let's examine the phenomenon y = pays back loan on time

$y \in \{0, 1\} = \mathcal{Y}$ output space

did not pay back on time

paid back on time

(convention: 1 is the "positive" event or the thing you want to happen).

Models w/ output spaces of cardinality 2 are called "binary classification models".

The casual inputs are features or characteristics of the individual person.
We don't know the casual model why people pay / don't pay back loans. We are going to make one up just as an illustration.