01/08/2020

"Models" are "approximations" or "abstractions" to "reality", "truth", "phenomenal"

| Model | Reality |
|---|---|
| Model airplane | Airplane |
| Map | Roads in a city |
| Wind tunnel | Fast moving air in the sky |
| "Early to bed, early to rise, makes a man wealthy" | Success of a person |

"All models are wrong but some are useful."
[George Box & Drapers]


$c \neq 3.14d$ by definition approximate

useful → you can use it for some purpose

$c \approx 3.14d$.

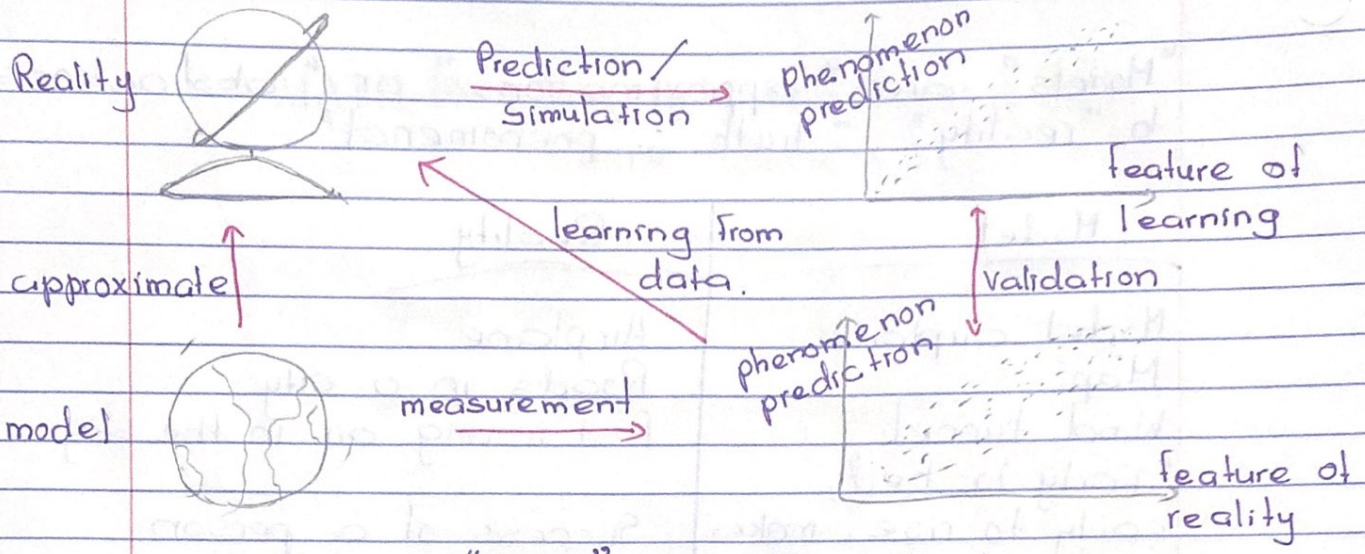
Two main goals

① Prediction :- Can the model predict a future event of the phenomenon under examination.

② Explanation :- How does the phenomenon work? (science).

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

→ Ambiguous (unclear, has more meaning)




"Data": Natural result of measuring a phenomenon.

What is this modelling?

output { health, wealth, wisdom } three different phenomenon

To make this model concrete, we need numerical definitions, we need "metrics"; These define how to measure both phenomenons and features of reality.

What are features?

input { Bedtime, "wake time" }

① Metrics for bedtime (b)
- Avg. bedtime in # of seconds and tenths from 17:00 from age 18-65

② Wake time (w)

③ Health measured by longevity (l)

④ Wealth measured by networth +65 (m)

⑤ Wisdom measured by a test (s)

Metric Evaluation

① Does if capture the feature / phenomenon?
Yes

② Is it easily readable and unambiguous?
Yes (e.g. s.568)

③ Good resolution?
Yes

④ Is it monotonic?

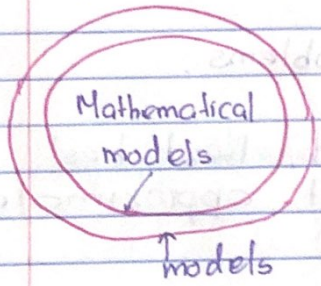
We want to estimate f where

$$\begin{bmatrix} l_{\star} \\ m \\ s \end{bmatrix} = f(b, w)$$

outputs: $l_{\star}$, m, s — inputs: b, w

Prediction of phenomenons

"Mathematical models" are idea and observations, not physical.




Mathematical models are at least 4,000 yrs old. Examples

$a = F/M = f(m, F)$

$F = mc^2$

deterministic — true function (unknown); that combines $z_1 \ldots z_k$

$$y = f(z_1, z_2, \ldots, z_k)$$

true causal inputs

The phenomenon, response, outcome, endpoint, dependent variable (one dimentional)

Phenomenon is pay back mortgage ($y=1$) or not pay back mortgage ($y=0$).

$y \in [0, 1] = y$ (output space)

Binary

"Positive class"

What are the causal inputs?

$z_1, z_2, \ldots, z_k$

$z_1$ = Has the money $\in [0, 1]$ at payback time

$z_2$ = Unforces emergency $\in [0, 1]$

$z_3$ = Criminal intense $\in [0, 1]$

$f(z_1, z_2, z_3) = z_1 (1 - z_2)(1 - z_3)$

## Fundamental modelling problems

You don't know the $z$'s or $f$. Next best thing is obtain measurements that approximate the $z$'s.

Call these measurements $x$'s

$x_1 \rightarrow$ Credit score $\in \mathbb{R}$

$x_2 \rightarrow$ Salary based on tax return $\in \mathbb{R}_+$

$x_3 \rightarrow$ Miss loan previously $\in [0, 1]$

$x_4 \rightarrow$ Crime in past $\in [0, 1]$