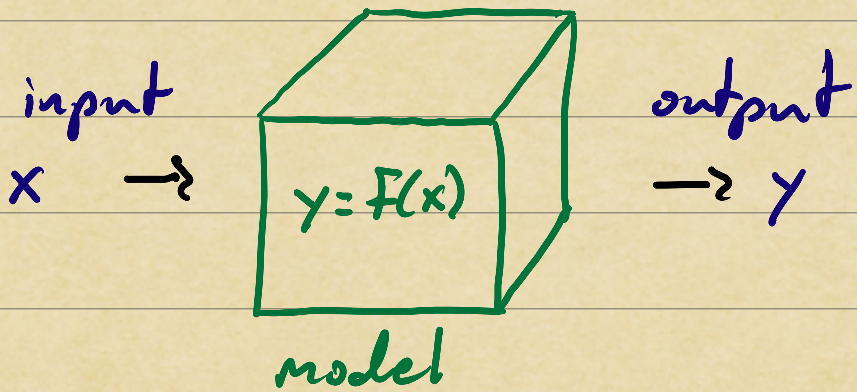
3 Foundational Ideas

Aim: use data to develop mathematical models of complex phenomena.



"Good" Models

Explanatory Power
Predictive Accuracy
Capacity to Generalize
Computability/Solvability
Interpretable/Actionable

Newton's Laws
General Relativity
Quantum Mechanics
⋮

Question: What is new in our day & age?

Big Computers + Big Data

To leverage these resources toward our aim of constructing "good" models, we need math!

The following three ideas will guide our mathematical journey as we seek to build good models from data.

Idea 1. Model Regression

Q: How well does a model fit data?

Suppose that we have "input-output" data

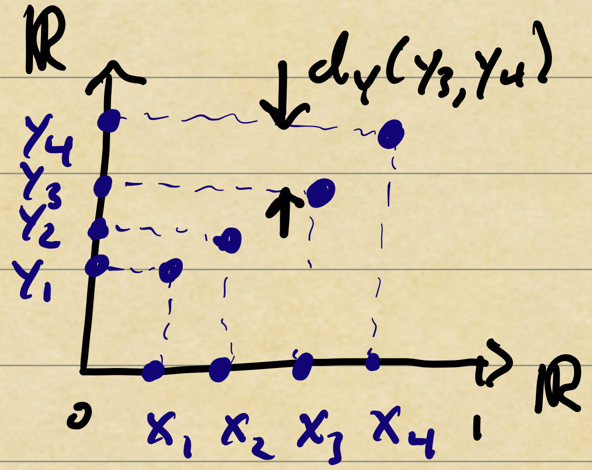
$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n) \in X \times Y,$$

where Y is a metric space with metric

$$d_Y : Y \times Y \to \mathbb{R}.$$

"distance on Y"

Example. "Scatter Plot" regression



$X = [0,1], \quad Y = \mathbb{R}$

$d_Y(y_i, y_j) = |y_i - y_j|$

Q: How do we find a model

$$F : X \to Y$$

that "fits" the data well?

The simplest idea is to look for an F that best reproduces the data:

$$E_j^F = d_Y(y_j, F(x_j)) \qquad j = 1, \dots, n,$$

should be as small as possible.

Q: Where should we look for F?

Idea 2. Model Accuracy

Q: How well does a model predict new data?

Suppose that we have found a model

s.t. $E_i^F = d_Y(y_i, F(x_i)) = 0, \quad i = 1, \dots, n,$

i.e., the model fits the data perfectly.

If the data is generated by a map

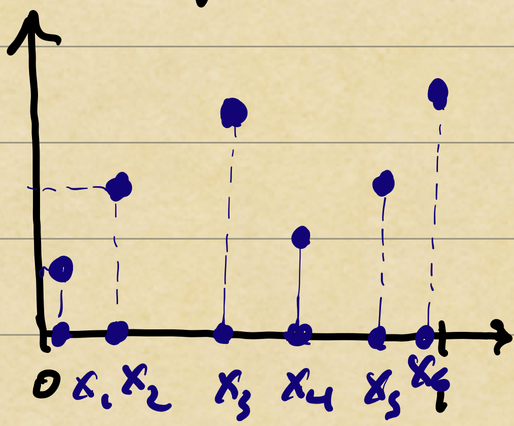
$$G : X \to Y,$$

what can we say about the error

$$E_F(x) = d_Y(F(x), G(x))$$

associated with a new input $x \in X$?

Example. Polynomial Interpolation



$$f(x) = a + bx + cx^2$$

$$\text{s.t. } |y_i - f(x_i)| = 0, \quad i = 1, \dots, n$$

Q: Under what circumstances can $E_f(x)$ be large/small between data?

$\Rightarrow$ If the ground truth G is truly quadratic, then we can recover G exactly on $[0,1]$ as long as we have data at $n \geq 3$ distinct points.

pf] We have $G = \alpha + \beta x + \gamma x^2$, and $y_1 = G(x_1), \dots, y_n = G(x_n)$ implies

$$\begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

For any 3 distinct points x_i, x_j, x_k, the square Vandermonde system

$$\begin{bmatrix} 1 & x_i & x_i^2 \\ 1 & x_j & x_j^2 \\ 1 & x_k & x_k^2 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} = \begin{bmatrix} y_i \\ y_j \\ y_k \end{bmatrix}$$

is invertible and therefore α, β, γ (and, consequently, G) are determined uniquely by the data.

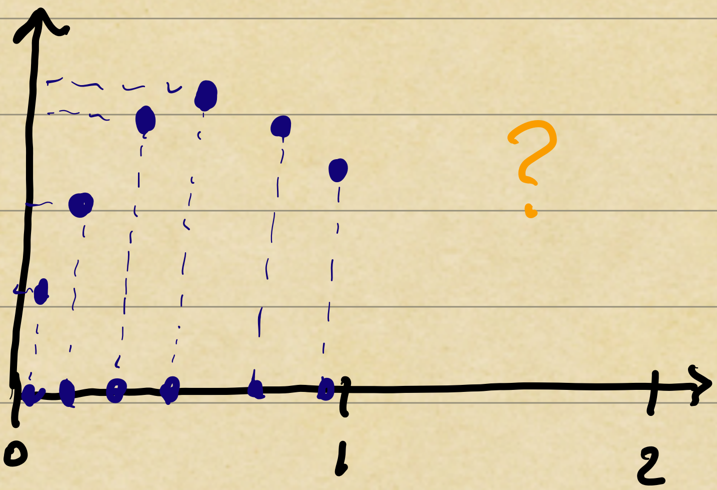
$\Rightarrow$ If G is merely continuous on $[0,1]$ and we have finite data, then the error between data points may be arbitrarily bad! Try to construct an example yourself...

$\Rightarrow$ In general, model accuracy will require knowledge of G (e.g., $G \in \mathbb{P}_2$) and structure in the data (e.g., data at $n \geq 3$ distinct points).

Idea 3. Model Generalization

Q: Can model predict inputs that are "far" from those previously seen?

Example. Extrapolating from Data



$$F(x) = a_0 + a_1 x + \cdots + a_n x^n$$
$$+ \cdots +$$

$$E_F(x) = 0 \quad x \in [0,1]$$

Q: How accurate is F in $[1,2]$?

$\Rightarrow$ Even for very smooth G and exact data on $[0,1]$, problem is typically ill-posed and accuracy may deteriorate rapidly away from $[0,1]$.