Generative Neural Networks for the Sciences

Assumption: "Reality" works according to some hidden mechanisms

$$X \sim p^*(X) \quad \text{unknown true probability, "true generating process"}$$

Goal: find approximation $X \sim p(X)$ (or $\hat{p}(X)$)

we have data: training set $TS = \{ X_i \sim p^*(X) \}_{i=1}^{N}$

$\Rightarrow$ $\boxed{\text{use TS to learn } p(x) \approx p^*(X)}$

two aspects of generative modeling:

- "inference": given some data instance X_i, calculate value of $\hat{p}(X_i)$
- "generation": create synthetic data $X \sim \hat{p}(X)$ which is indistinguishable from real data $X \sim p^*(X)$

- downstream benefits of generative modeling:

- powerful tool for humans (e.g. chat Bot as personal teacher)
- help produce insight: identify important factors influencing the outcome
  ex: "symbolic regression": find (learn) analytic formulas to explain reality (vanilla neural networks are black boxes)
- use $p(x)$ for decision making: ex: is treatment A better than B for patient X?
  "digital twin", "precision medicine"

R. Feynman: "What I cannot create, I do not understand."