New topic: Generative Modeling & Density Estimation

- Important & powerful class of ML techniques
- Generally <u>unsupervised</u> — learning from data, no labels
- Both aim to learn underlying $p_{data}(x)$ from samples
  probability density

with NNs, x could be very high dimensional! (hundreds, thousands, even millions of dims)

- GM: given samples $x_i \sim p_{data}(x)$
  learn to generate more samples following p_{data}.

  general approach: take $z \sim$ simple, known dist'n
  e.g. $N(0,1)^{d'}$

  learn map $z \xrightarrow{G} x$
  s.t. $x \sim p_{data}(x)$.

  could have $d' << d$
  "manifold hypothesis"

  data lives in $\mathbb{R}^d$ but secretly lives on a submanifold of much lower dimensions (data is controlled by a much lower dim'l latent space)

  GM is "noise 2 real"

- DE: given samples $x_i \sim p_{data}(x)$
  learn estimator $\hat{p}_{data}$ for p_{data} itself
  (so can report $p_{data}(x_i)$ also know it for new samples)

~~New approach: learn map $x \to z$ s.t. $z \sim N(0,1)^{d}$~~

- Note: GM & DE similar but can do one w/out other!
  - Can learn GM w/ pdata implicitly (GAN, VAE)
  - Can learn DE w/out being able to sample (KDE)

- Many pot'l applications of GM & DE!
  - countless industry applications (deepfakes, chatbots, ...)
  - in physics & astronomy
    - fast simulation: surrogate modeling ←

      generate samples — expensive
      train GM on samples
      generate more — fast

    - upsampling

      e.g. turning star particles into stars

    - simulation based inference

      ↓

      learn $p(x|\theta)$ or $p(\theta|x)$
      from samples (x, θ), perform MLE → θ
      or sample from posterior

      ↓

      powerful if x, θ very high dim'l — full phase space beyond summary statistics

      fast if $p(x|\theta)$ slow to evaluate but forward model fast to run (faster than MCMC)

      see my paper tonight using this for pulsar timing arrays! uses LSTMs too!

    - phase space integration / importance sampling
      know $p(x)$ exactly but can't sample from it
      fit model for $p_{\theta}(x)$ to $p(x)$, sample from model to integrate over $p(x)$

- anomaly detection
  - low $p(x)$ - rare?
  - learn $p(x|y)$ for $y \in CR$
    interpolate y into SR
    get background model, fully data driven
    for x

useful beyond anomaly detection of course!

- .... probably more applications out there!

---

4 classes of NN-based GMs:

- GANs
- VAEs
- Flows (also DE)
- Diffusion (also DE)

Plan is to introduce class to each one, with examples
