II. Generative Models I

We have some dist. $p_{truth} \equiv p_{data}(x)$ & want to sample "generate" new data $x \sim p_{truth}(x)$

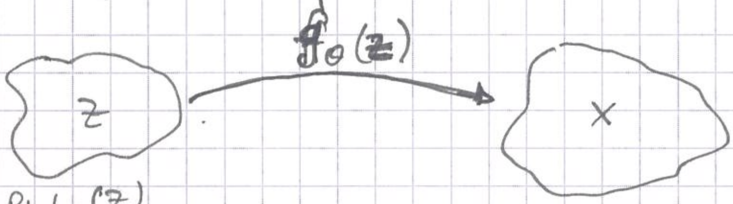
p can be given – explicitly (as function), e.g. $\frac{d\sigma}{\sigma} \sim$ diff x-sec

– implicitly via a set of data points $\{x\} \sim p_{data}(x)$

in HEP:
- event generation
- phase-space integration
- calorimeter (detector)
- unfolding...

$\Rightarrow$ this is a stochastic (random) process (RNG)

$\rightarrow$ need „random" input



$p_{laten}(z)$

$\hookrightarrow$ easy: gauss, uniform

$x = f_\theta(z) \sim p_{model}(x) \overset{!}{\simeq} p_{data}(x)$

Q: How to <u>construct</u> & <u>train</u> $f_\theta(z)$?

$\boxed{\text{1. GANs} \quad \text{2. VAEs} \quad \text{3. NFs} \quad \text{4. Diffusion Model}}$

2.1 How do we know we have a good model?

1) we have p_{model} + p_{truth} explicitly:

$\rightarrow$ e.g. KL (Kullback-Leibler) divergence

not symmetric
not a metric

$$\boxed{KL(p_{model}, p_t) = \int dx \; p_t \log \frac{p_t}{p_\theta} = \left\langle \log \frac{p_{data}}{p_{model}} \right\rangle_{data}}$$

others are $\rightarrow$ Total variation!

or any other f-divergence [1309.3029] $V(p,q) = \frac{1}{2}\int dx \, |p-q|$

or EMD

2) Earth mover distance (EMD) (Wasserstein metric)

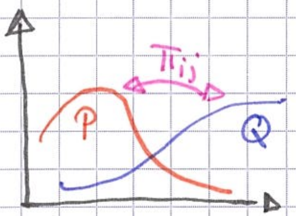
$\hookrightarrow$ next page:

↳ EMD: $\underbrace{\text{works}}_{\text{wassertein metric}}$ for p know, continue & discrete, & only samples available

lets assume you have two set $P = \{(x_i, p(x_i))\}_{i=1}^{N}$

and $Q = \{(y_j, q(y_i))\}_{j=1}^{M}$

⟹ minimum cost to transfom $P \overleftrightarrow{\text{into}} Q$



$$\boxed{\begin{aligned} EMD(p,q) &= \min_{\{\pi_{ij} > 0\}} \sum_{ij} \pi_{ij} \, \|x_i - y_j\|_2 \, \frac{1}{NM} \\ \text{with } \frac{1}{M}\sum_j^M \pi_{ij} &\leq p(x_i) \quad \frac{1}{N}\sum_i^N \pi_{ij} \leq q(x_i) \end{aligned}}$$

$\cancel{\sum_i^N \sum_j^M \pi_{ij} = \min(\sum p, \sum q) = 1}$ ~~if normalized!~~

3) generate event & look

- single events (images)
- histograms/distribution (marginal distributions only)

4) classifier test [1610.06545, 2305.16774]

⟹ train NN classifier on data vs. gen

NP Lemma: optimal classifier yields LL ratio

$$C_{optim} = \frac{p_{data}(x)}{p_{data}(x) + p_{model}(x)}$$

if $C = 0.5$ (confused)

↳ $p_{data} = p_{model}$

+ MMD [1907.03764]

+ ...

2.2 Generative adversarial network (GAN)

Goodfellow et. al [1406.2661]

train 2-NN in competition:

generator G: $Z \to X$

discriminator classifier D: $\begin{cases} 0 & \text{gen} \\ 1 & \text{truth} \end{cases} \to$ D prop of truth!

Loss:

$$\min_D \mathcal{L}_D = -\langle \log D \rangle_{p_{data}} - \langle \log(1-D) \rangle_{p_\theta}$$

$$\min_G \mathcal{L}_G = \langle \log(1-D) \rangle_{p_\theta}$$

$$\simeq \langle \log(1-D(G(z))) \rangle_{z \sim p(z)} \Big\} \begin{matrix} D \to 0 & L \to 0 \\ D \to 1 & L \to -\infty \end{matrix}$$

$\to$ bad minimum

$$\sim -\langle \log D \rangle_{p_\theta} \Big\} \begin{matrix} D \to 0 & L \to \infty \\ D \to 1 & L \to 0 \end{matrix}$$

$\to$ better gradients!

$\to$ $\boxed{\text{show plot!}}$

Training (iteratively):

1. freeze g, train D with batches of <u>real</u> and <u>gen</u>
2. freeze D, train G with samples of g $\to$ iterate

Some problems:

- min-max unstable [1705.09367]
  - $\to$ vanishing gradient $\to$ <u>regularize</u> + gradien penalt [1801.04406]
- metric for sucess (loss <u>only</u> show competion)
- mode collapse

not always convergent!

Ways to improve

- Wasserstein GAN [1701.07875, 1704.00028]
- gradient penalty
- spectral normalization [1802.05957] $\leftarrow$ Lipschitz contraint
- other losses
  - $\to$ • least square [16.11.04076]
  - • Relativistic GAN [1807.00734]
  - • ...

pro:

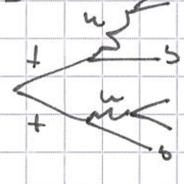
- great, realistic images (StyleGAN I-III)
- main idea is very nice

cons:

- not optimal for distribution
- model is only implicitly given!
  - ↳ not useful for density estimation!

2.2.1 Example I: Event generation [1907.03764] AB, TP, RW

We consider: $pp \to t\bar{t} \to (bq\bar{q}')(\bar{b}\bar{q}q')$



$X = \{\vec{p}_1, \dots \vec{p}_6\}$ using on-shell conditions

→ GAN learn nicely non-sharp distribution

→ But needs help for invariant masses → MMD [0805.2368] Loss

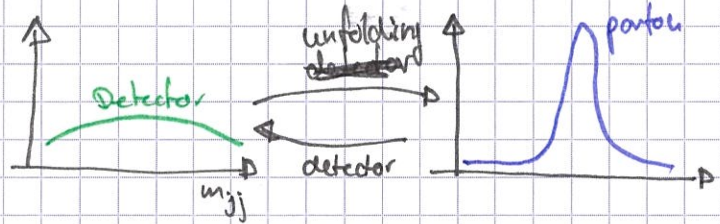
MMD Loss: - kernel based sample test!

↳ show plots!

2.2.2 Example II: Unfolding [1912.00477] AB, TP, RW, MB

What is unfolding?

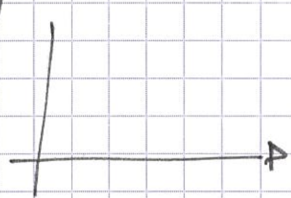
↳ remove detector effects!



Naiv ansatz: Dataset (x_d, x_p)

$$\boxed{x_d \to G \to x_p' \quad D \quad x_p}$$

explain on plots!



↳ problem: - locality information is lost!
- deterministic & detector is stochastic

unfolding should obey this!

↳ Show plots! full and cutted!