On Unifying Deep Generative Models

Herman Dong

February 20, 2020

Outlines

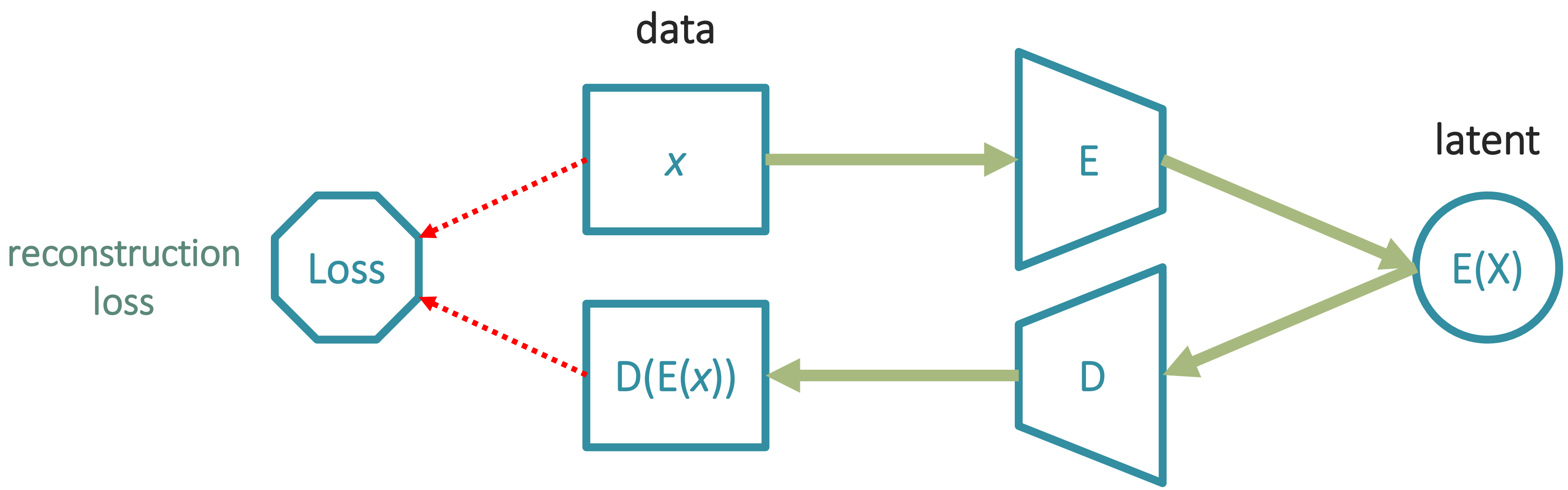
- Brief introduction on some notable DGMs (AE, VAE, GAN, AAE, InfoGAN, ADA)
- Schematic graphical model representation
- Reformulating different DGMs
- Discussions

Some notable DGMs

A brief introduction

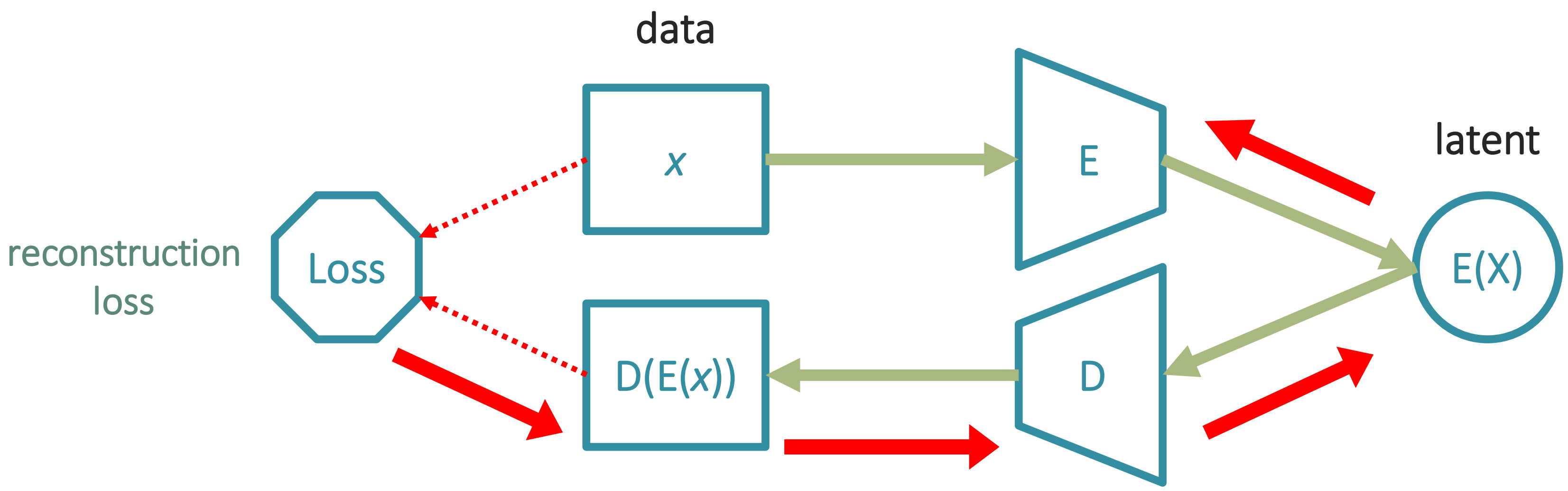
AE (Autoencoder)

- minimize **reconstruction loss**



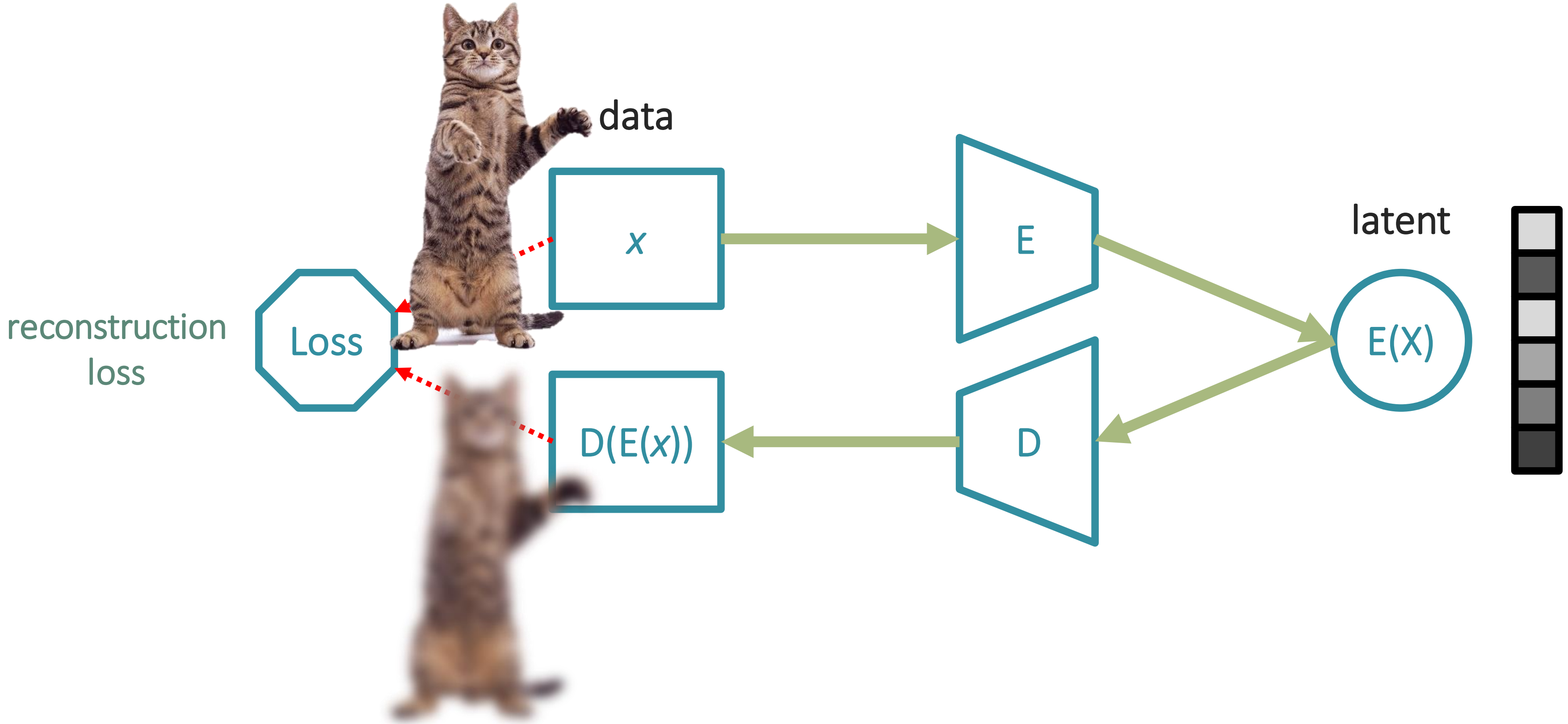
AE (Autoencoder)

- minimize **reconstruction loss**



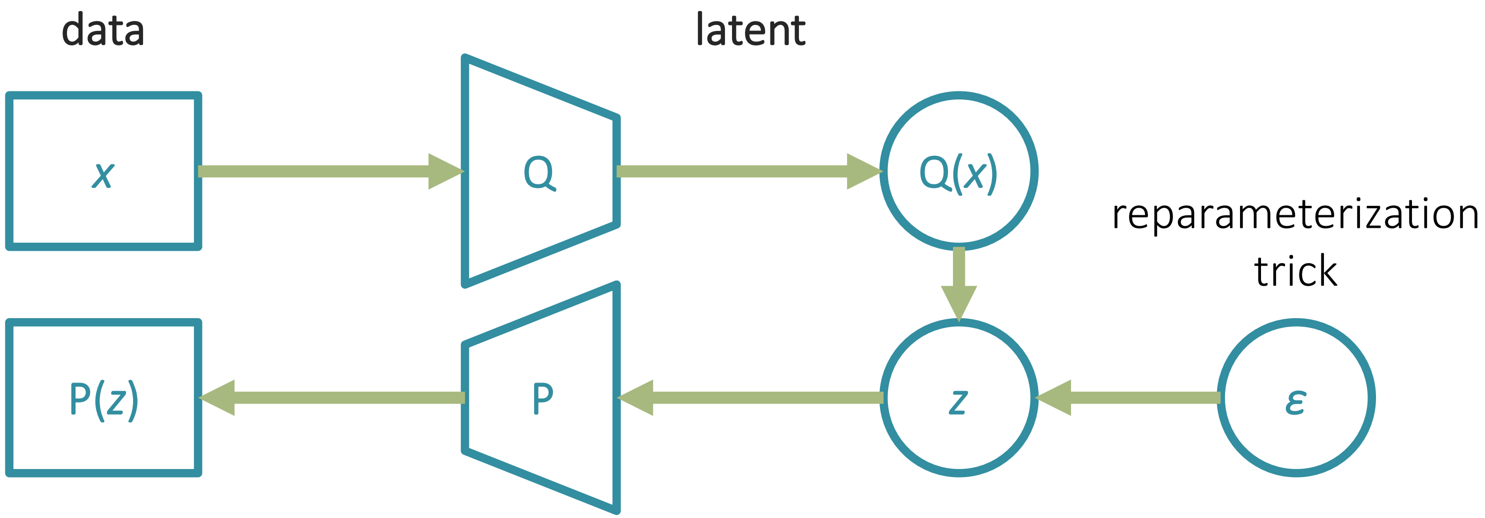
AE (Autoencoder)

- minimize **reconstruction** loss



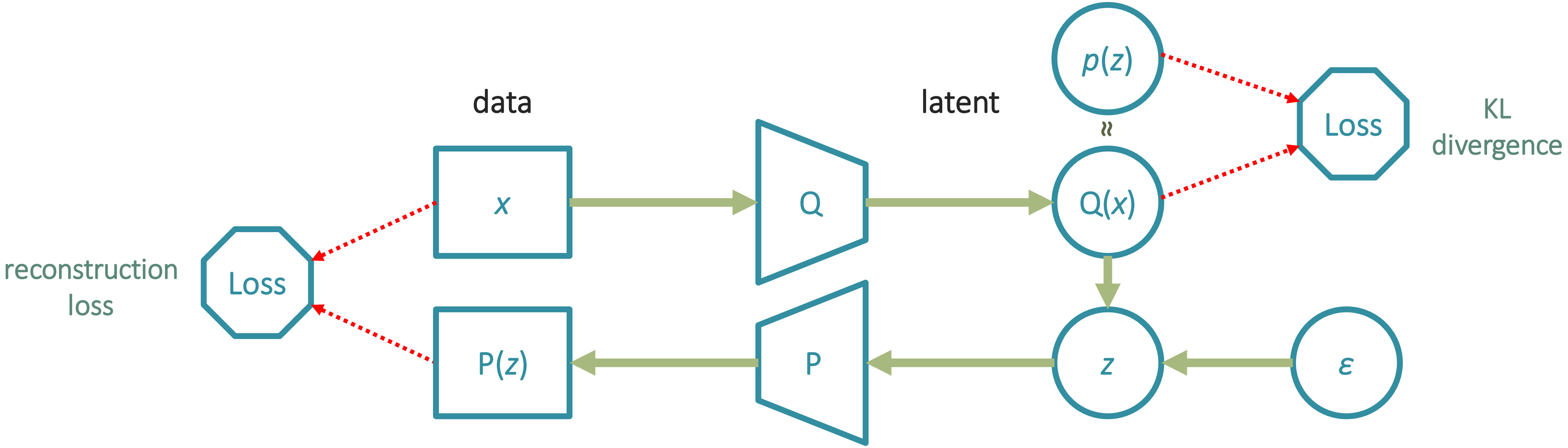
VAE (Variational Autoencoder)

- minimize **reconstruction loss**
- minimize **divergence** between encoded latent distribution and prior distribution



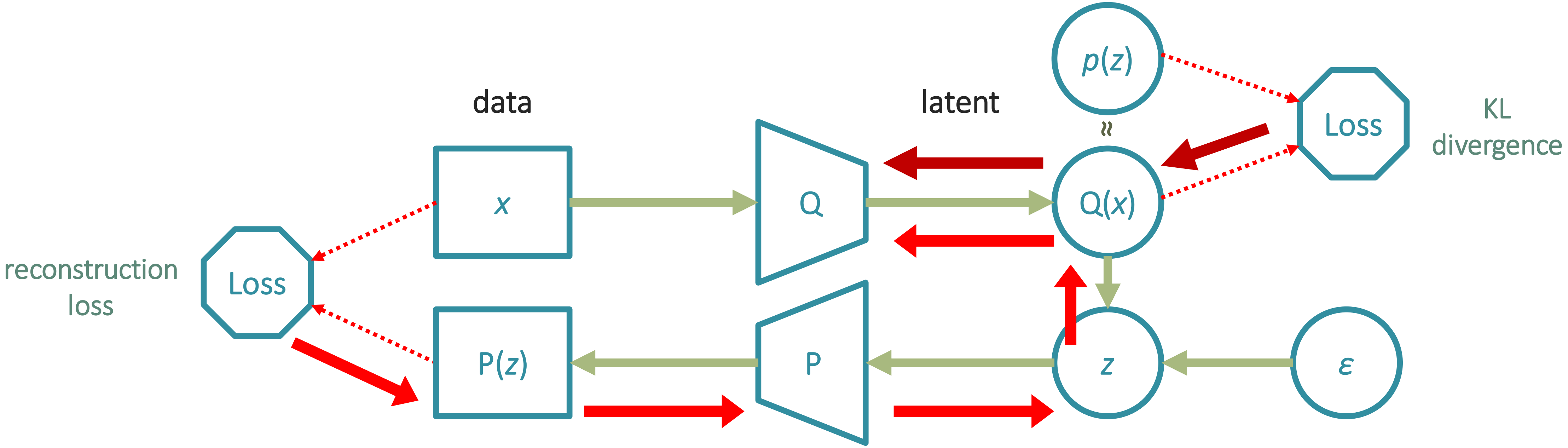
VAE (Variational Autoencoder)

- minimize **reconstruction loss**
- minimize **divergence** between encoded latent distribution and prior distribution



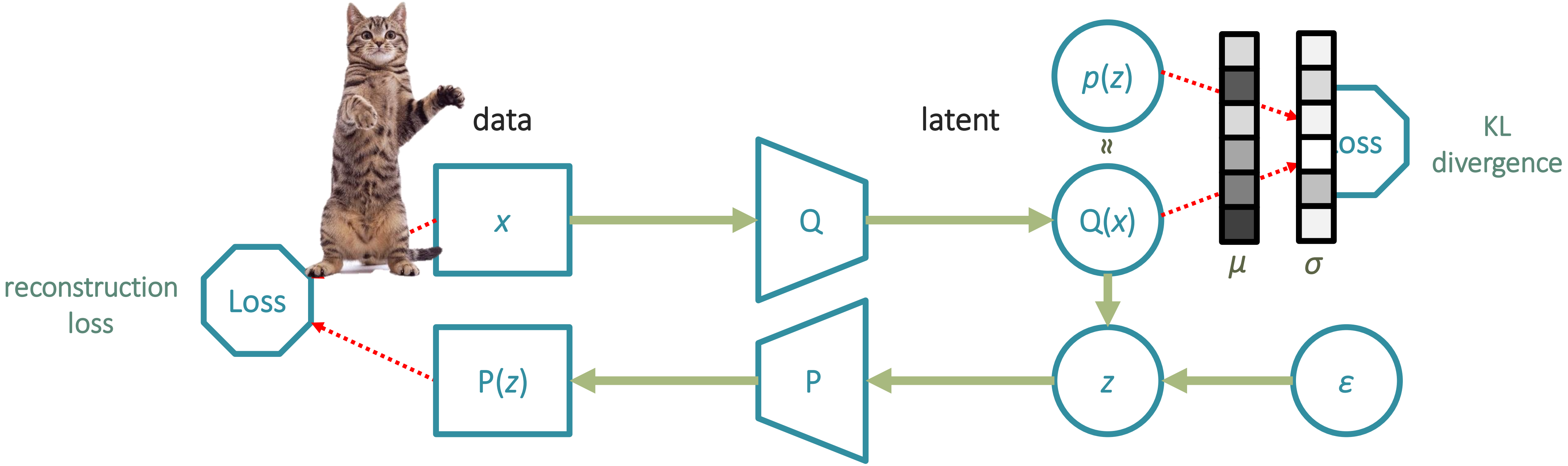
VAE (Variational Autoencoder)

- minimize **reconstruction loss**
- minimize **divergence** between encoded latent distribution and prior distribution



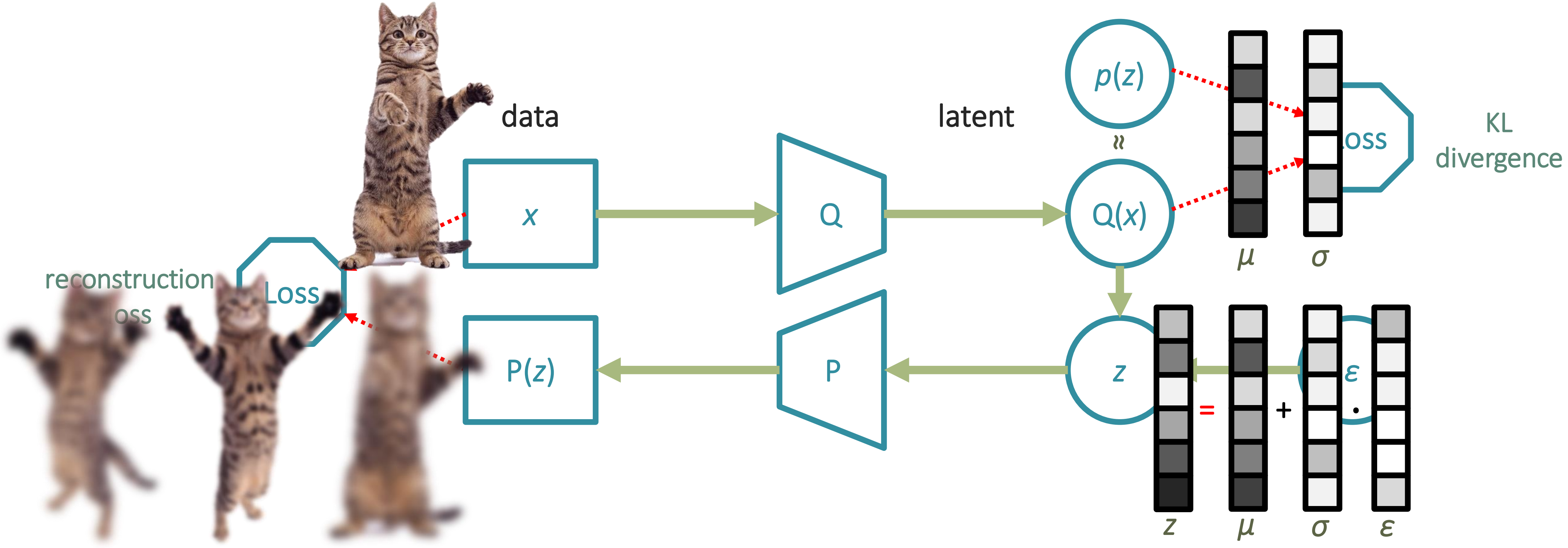
VAE (Variational Autoencoder)

- minimize **reconstruction loss**
- minimize **divergence** between encoded latent distribution and prior distribution



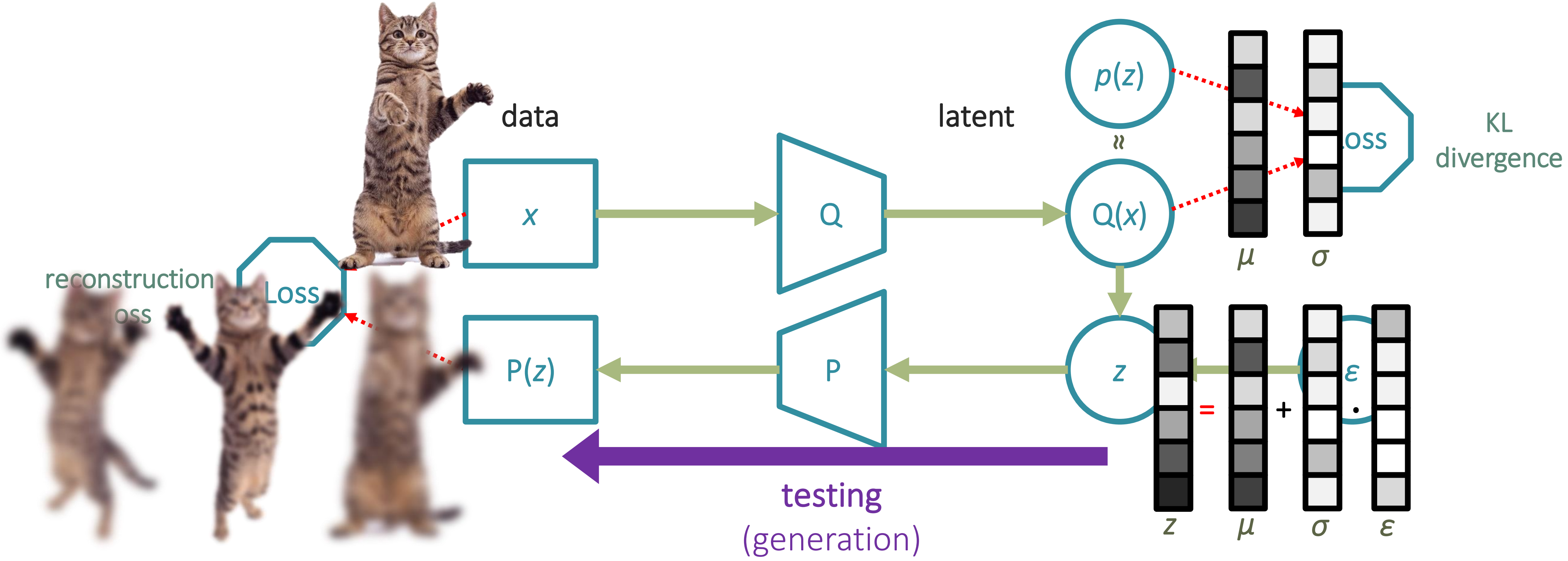
VAE (Variational Autoencoder)

- minimize **reconstruction loss**
- minimize **divergence** between encoded latent distribution and prior distribution



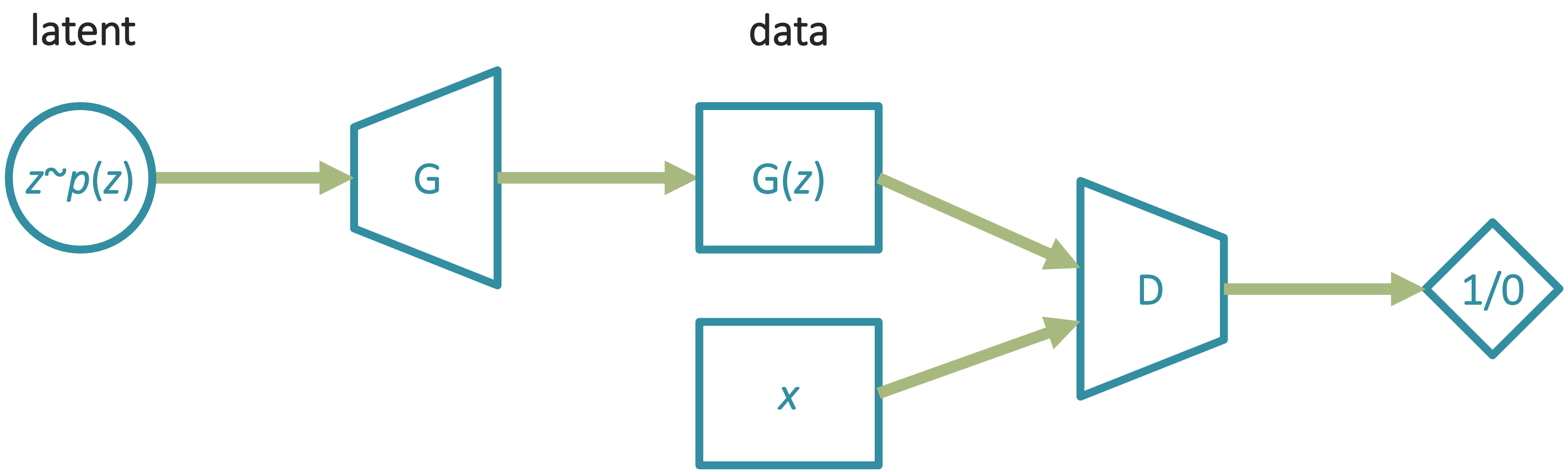
VAE (Variational Autoencoder)

- minimize **reconstruction loss**
- minimize **divergence** between encoded latent distribution and prior distribution



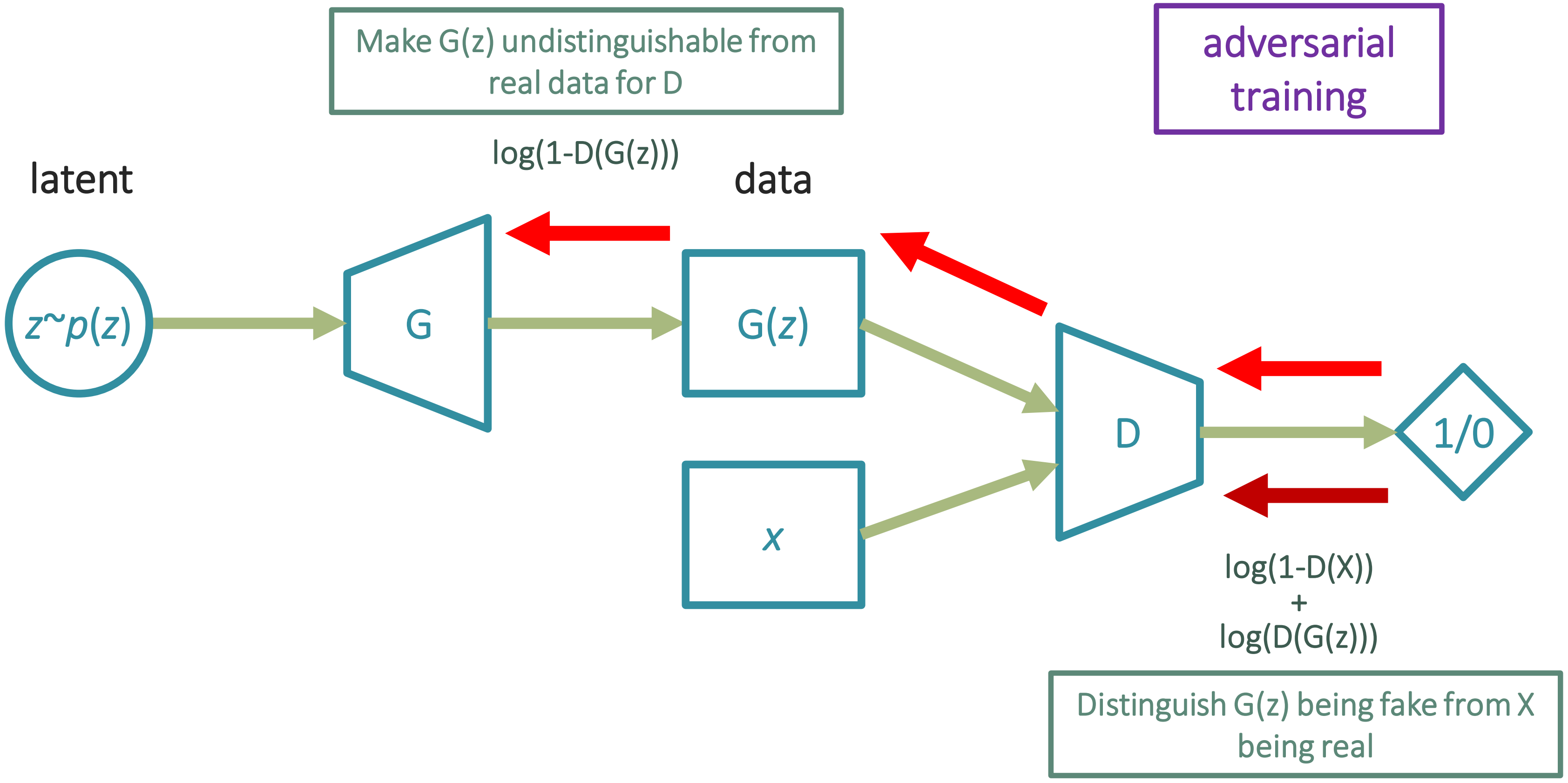
GAN (Generative Adversarial Network)

- minimize **divergence** between the distribution of real data and generated samples



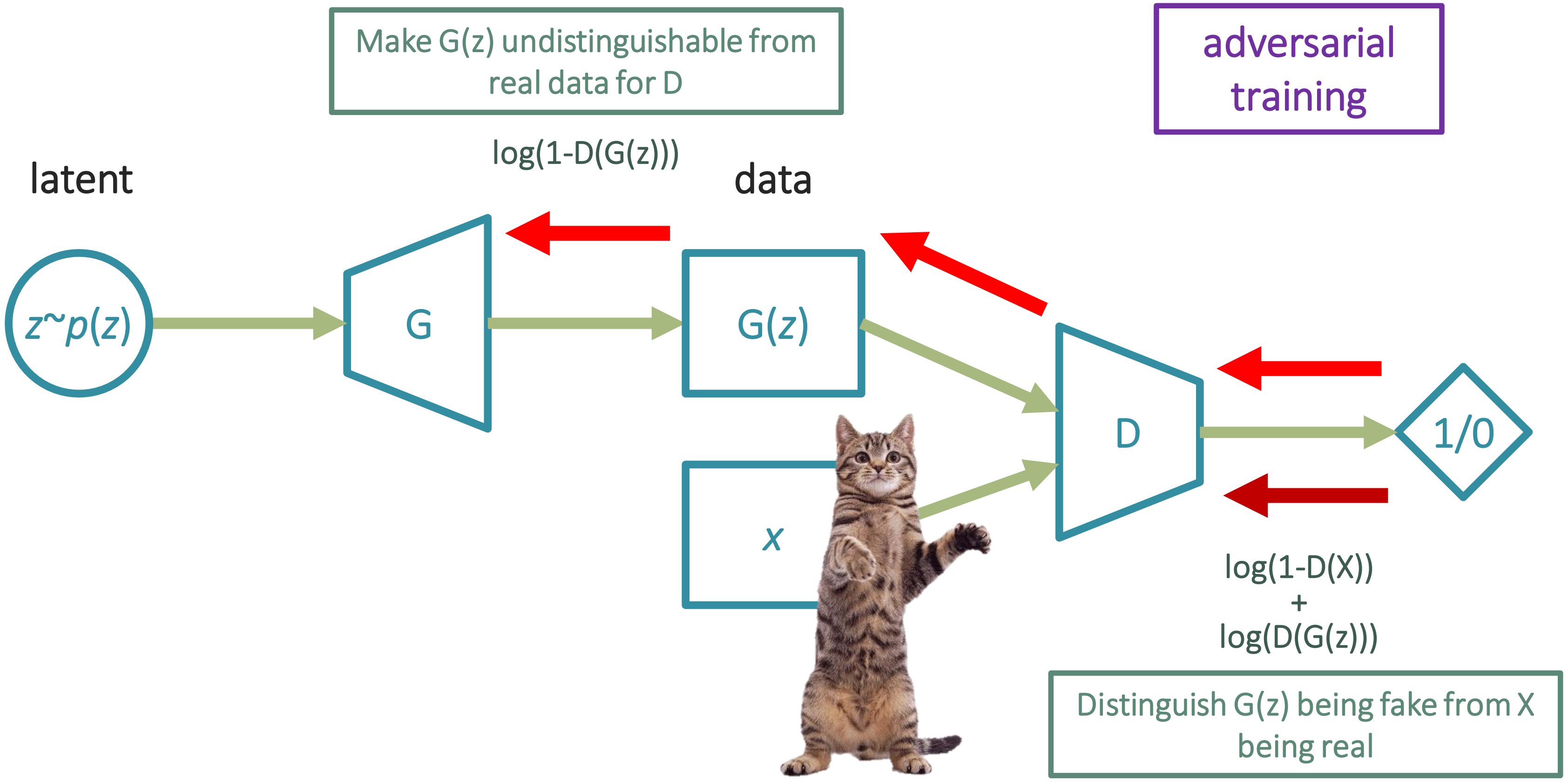
GAN (Generative Adversarial Network)

- minimize **divergence** between the distribution of real data and generated samples



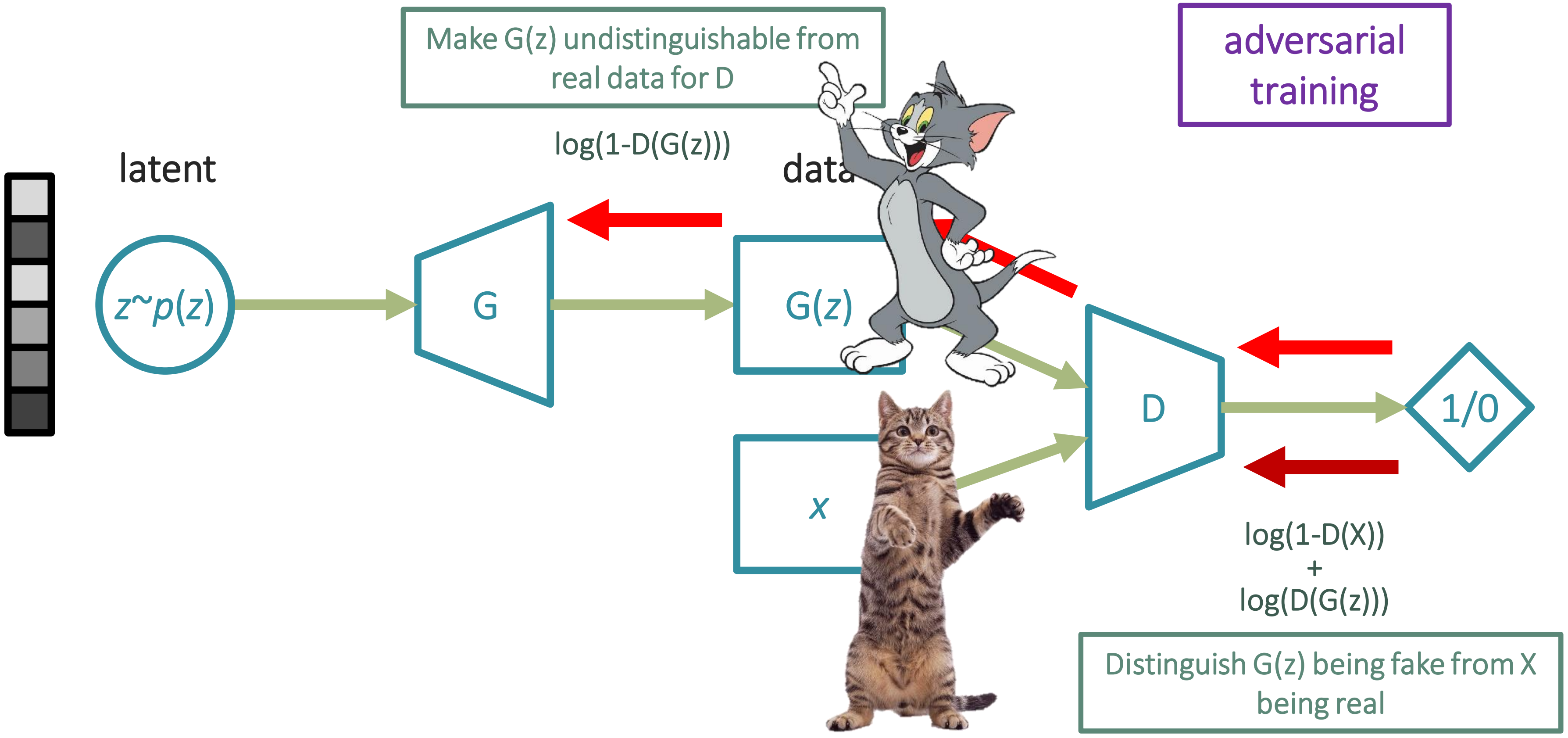
GAN (Generative Adversarial Network)

- minimize **divergence** between the distribution of real data and generated samples



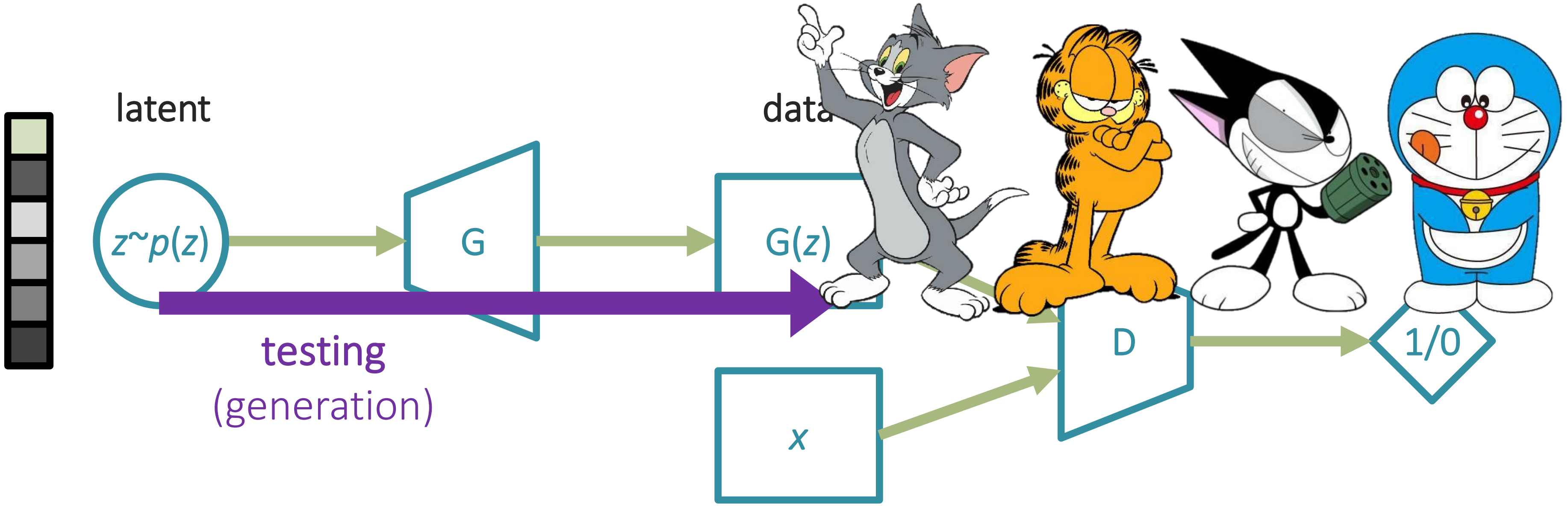
GAN (Generative Adversarial Network)

- minimize **divergence** between the distribution of real data and generated samples



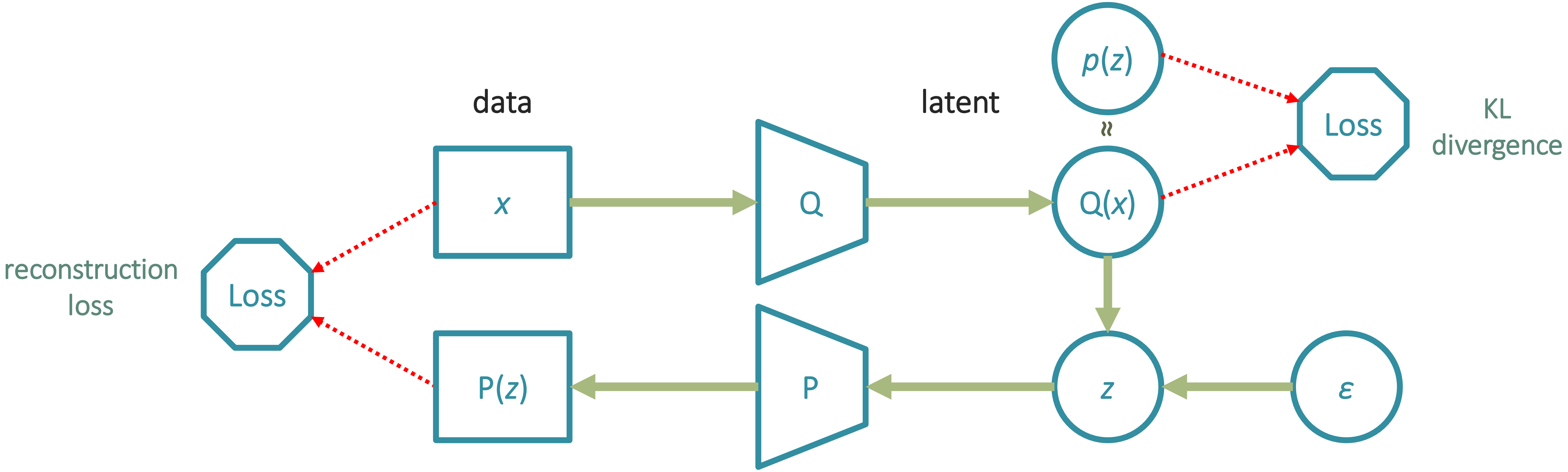
GAN (Generative Adversarial Network)

- minimize **divergence** between the distribution of real data and generated samples



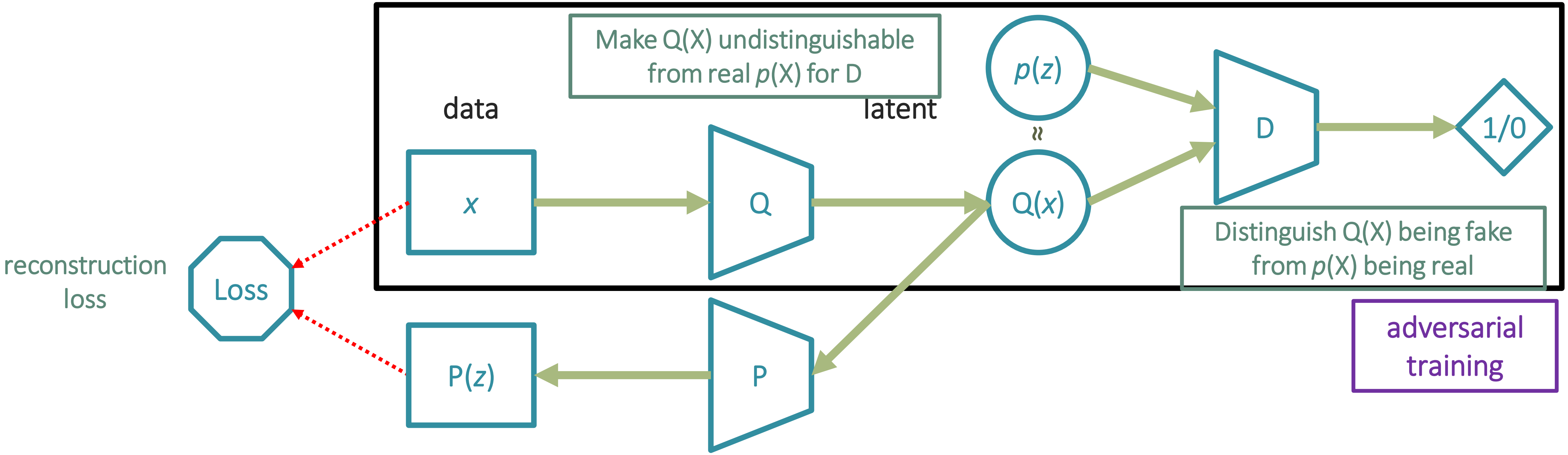
AAE (Adversarial Autoencoder)

- minimize **reconstruction loss**
- minimize **divergence** between encoded latent distribution and prior distribution



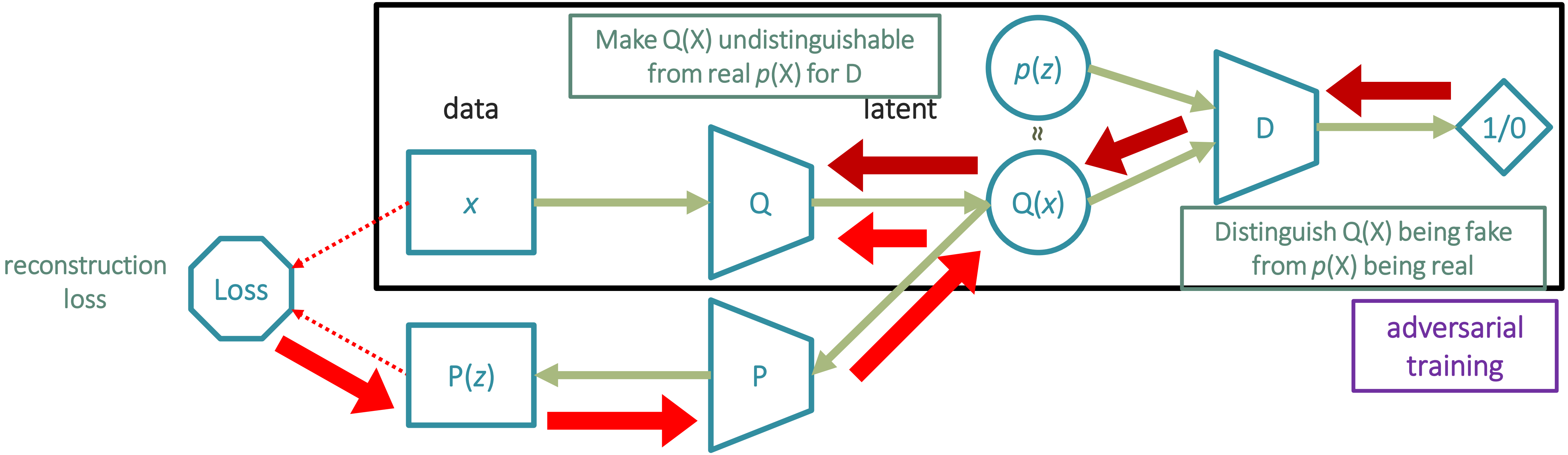
AAE (Adversarial Autoencoder)

- minimize **reconstruction loss**
- minimize **divergence** between encoded latent distribution and prior distribution



AAE (Adversarial Autoencoder)

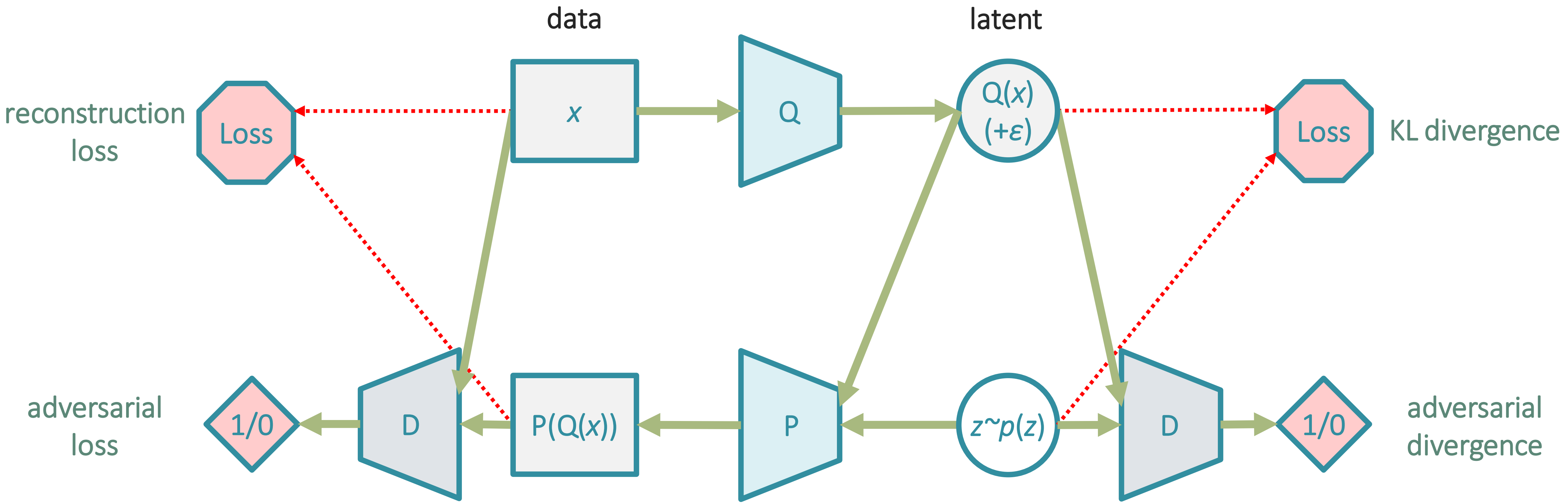
- minimize **reconstruction loss**
- minimize **divergence** between encoded latent distribution and prior distribution



What's going on?

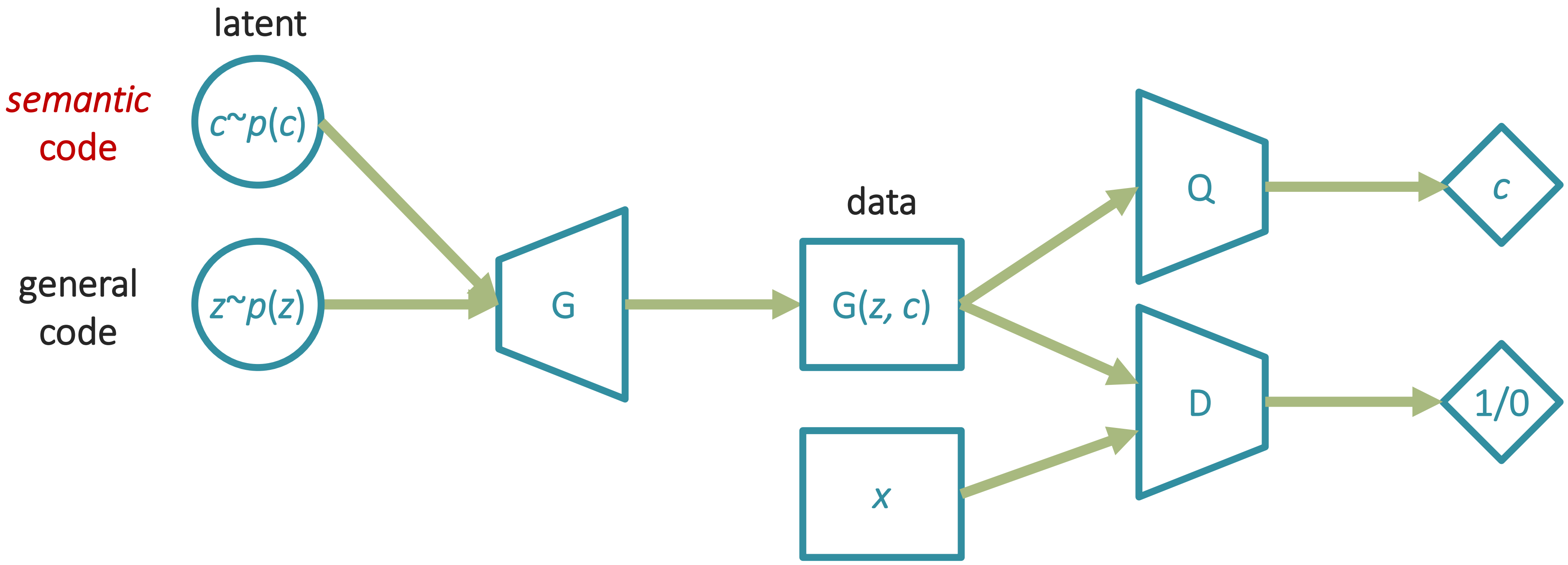
Observations

- Latent and data spaces are sort of "symmetric"
- Mappings $\mathrm{X} \rightarrow z$ and $z \rightarrow \mathrm{X}$ are sort of "symmetric"



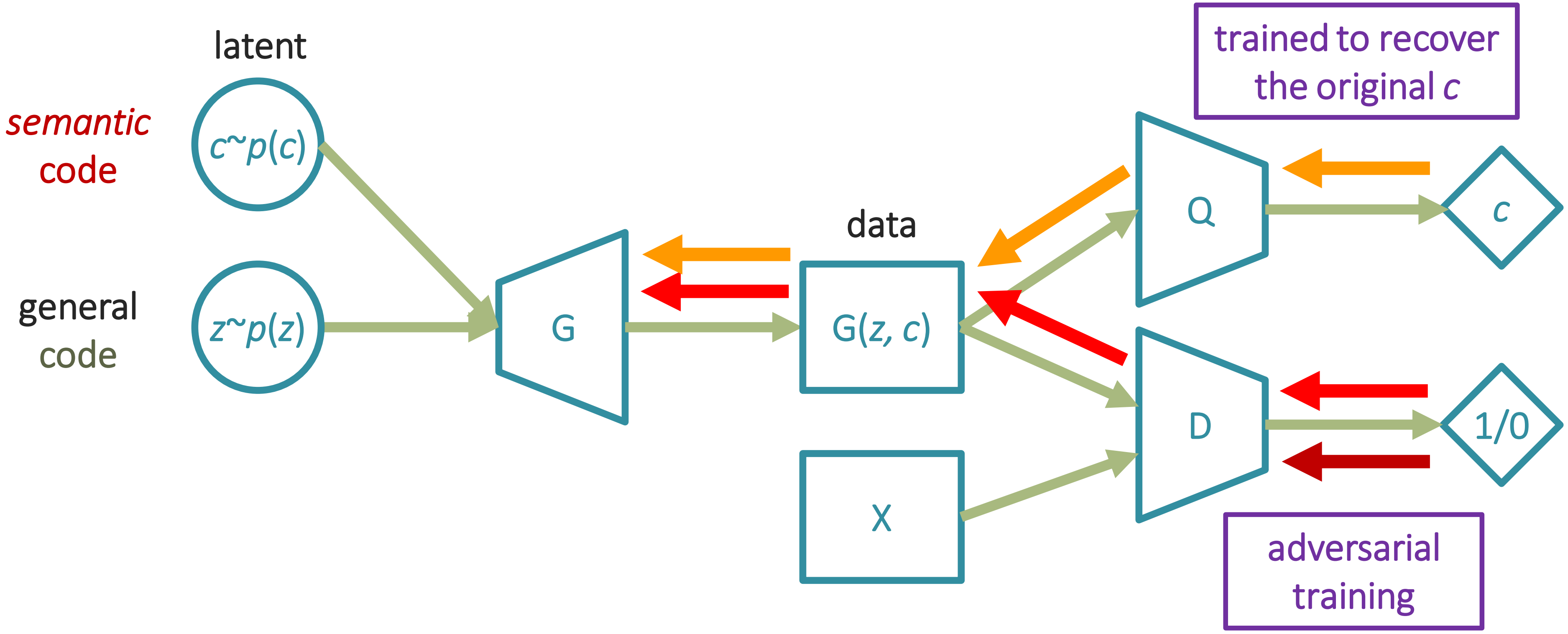
InfoGAN (Information Maximizing GAN)

- minimize **divergence** between the distribution of real data and generated samples
- minimize **reconstruction loss** of the semantic code



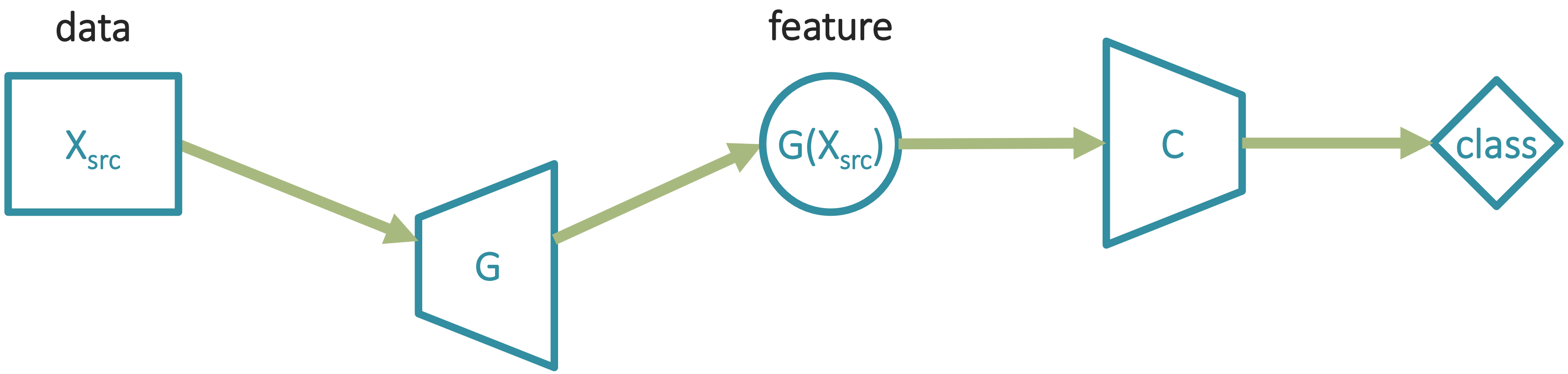
InfoGAN (Information Maximizing GAN)

- minimize **divergence** between the distribution of real data and generated samples
- minimize **reconstruction loss** of the semantic code



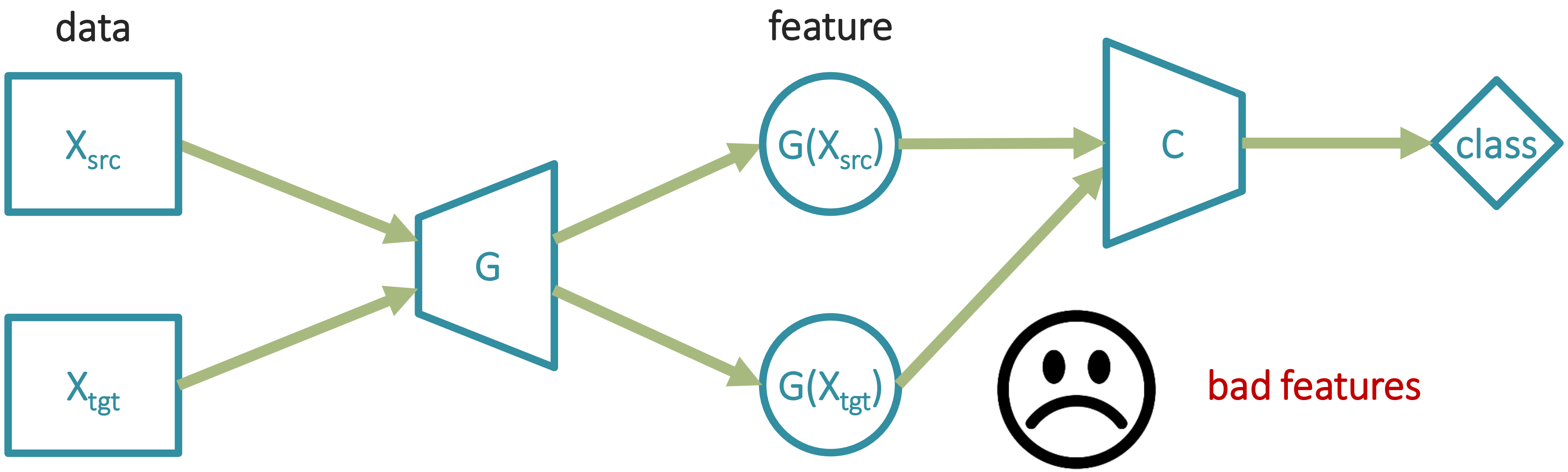
ADA (Adversarial Domain Adaption)

- **Goal**: given labeled data in source domain, classify unlabeled data in target domain.



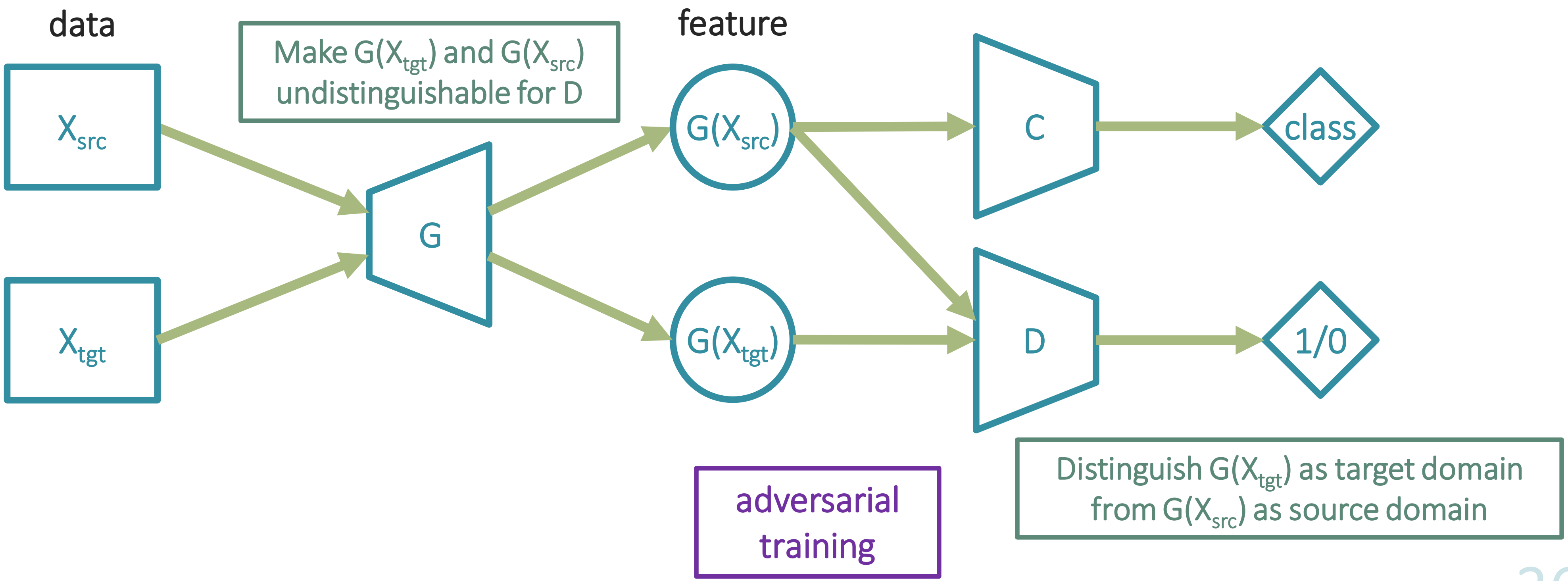
ADA (Adversarial Domain Adaption)

- **Goal:** given labeled data in source domain, classify unlabeled data in target domain.



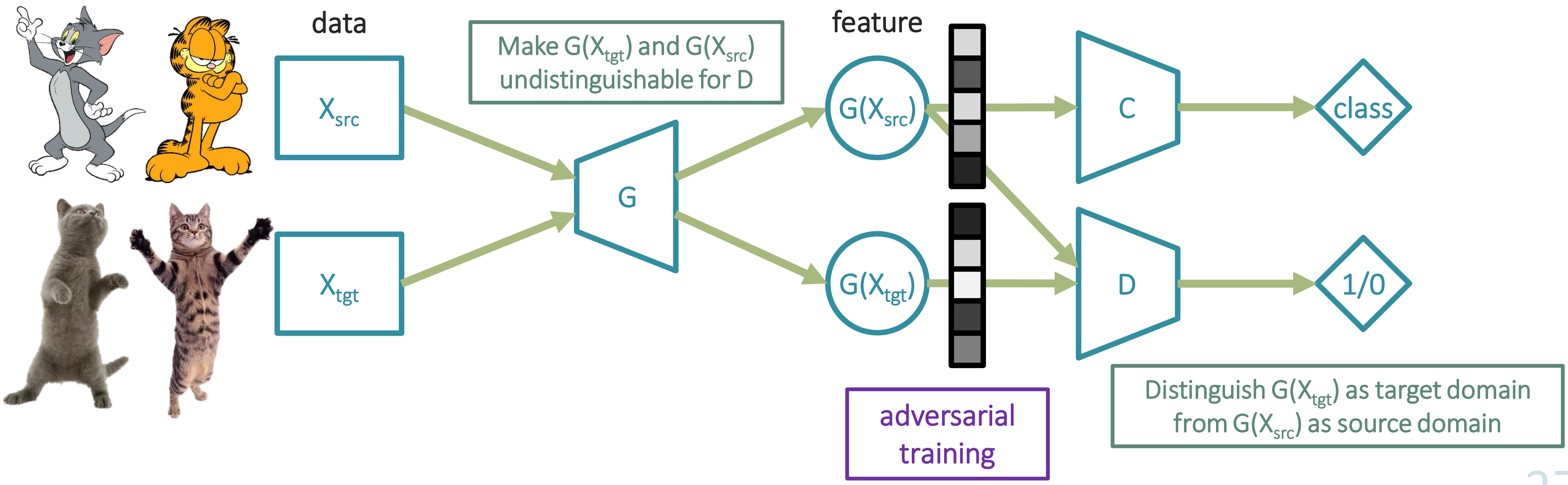
ADA (Adversarial Domain Adaption)

- **Goal**: given labeled data in source domain, classify unlabeled data in target domain.



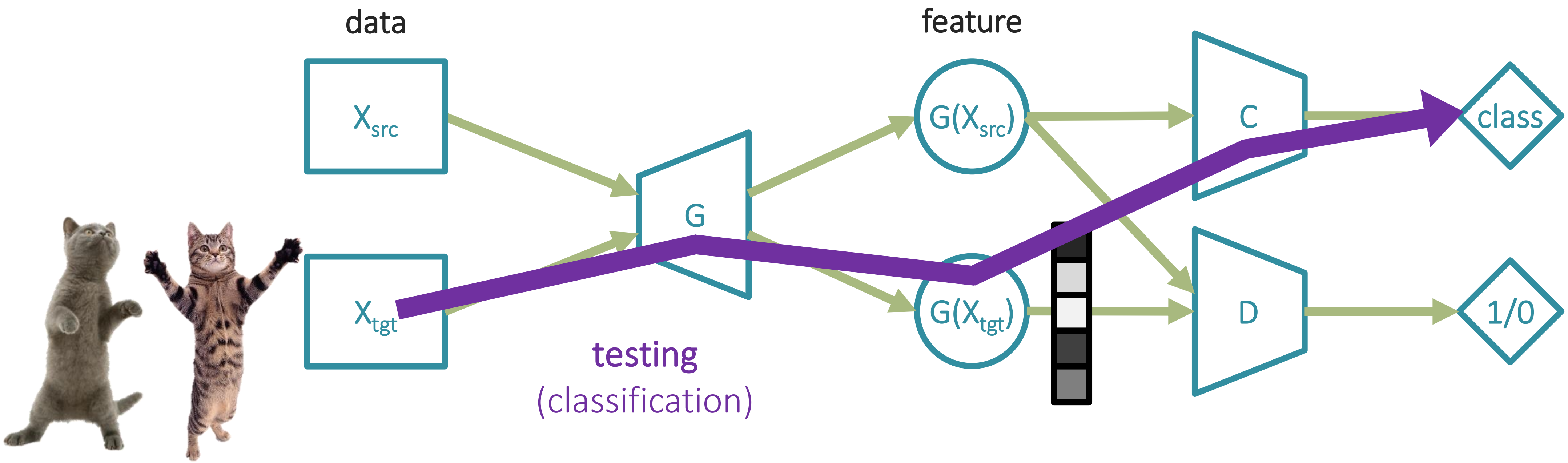
ADA (Adversarial Domain Adaption)

- **Goal**: given labeled data in source domain, classify unlabeled data in target domain.



ADA (Adversarial Domain Adaption)

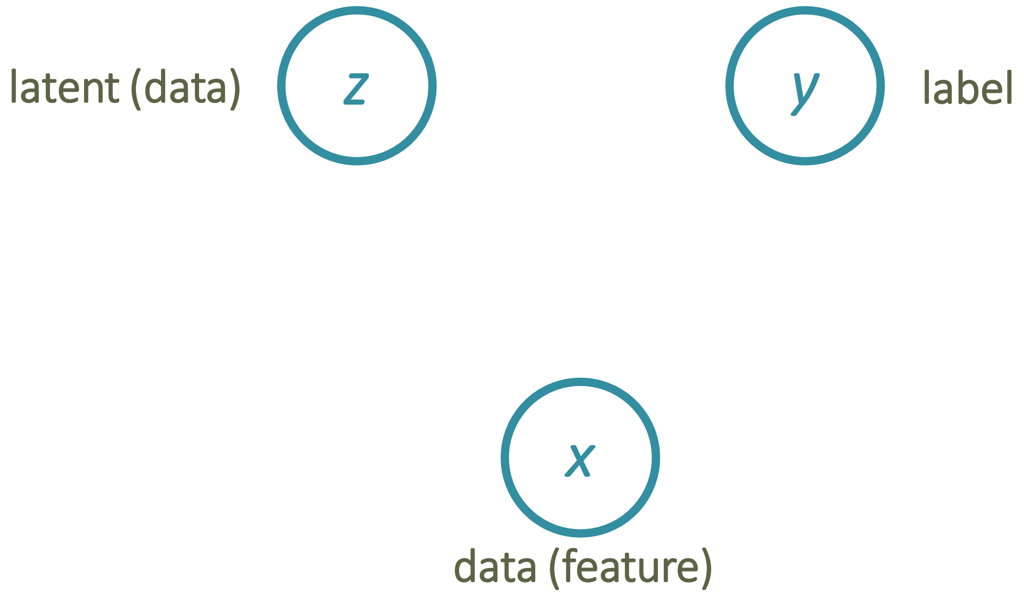
- **Goal**: given labeled data in source domain, classify unlabeled data in target domain.



Schematic graphical model

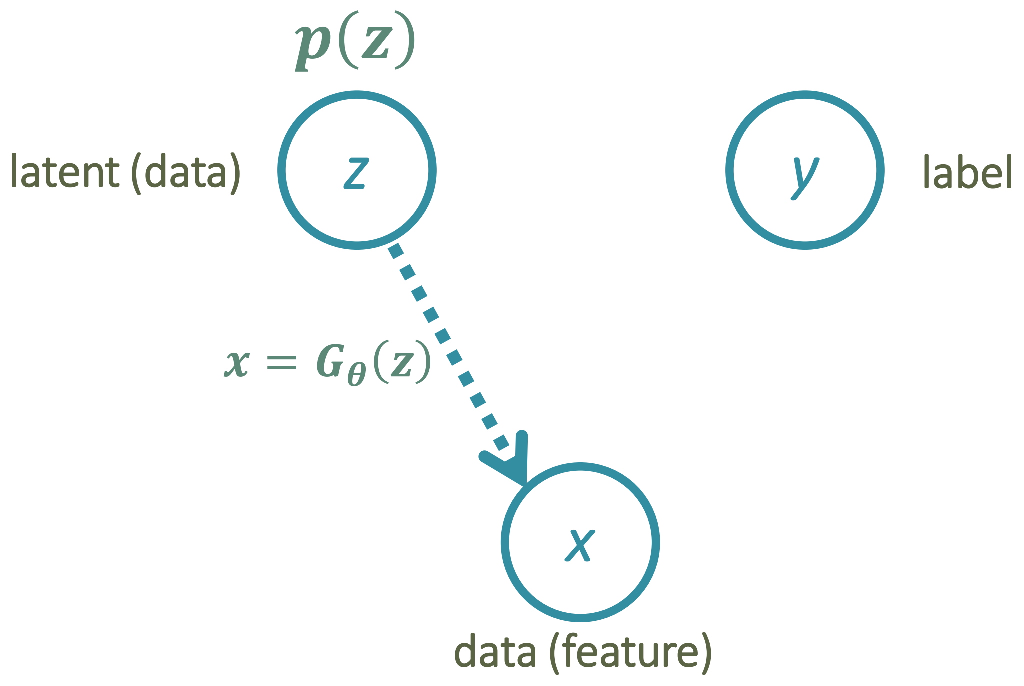
A representation for a unified view on DGMs

Schematic graphical model



Schematic graphical model

- $\boldsymbol{G_\theta} - \theta$ are parameters in generator
- $\boldsymbol{D_\phi} - \phi$ are parameters in generator

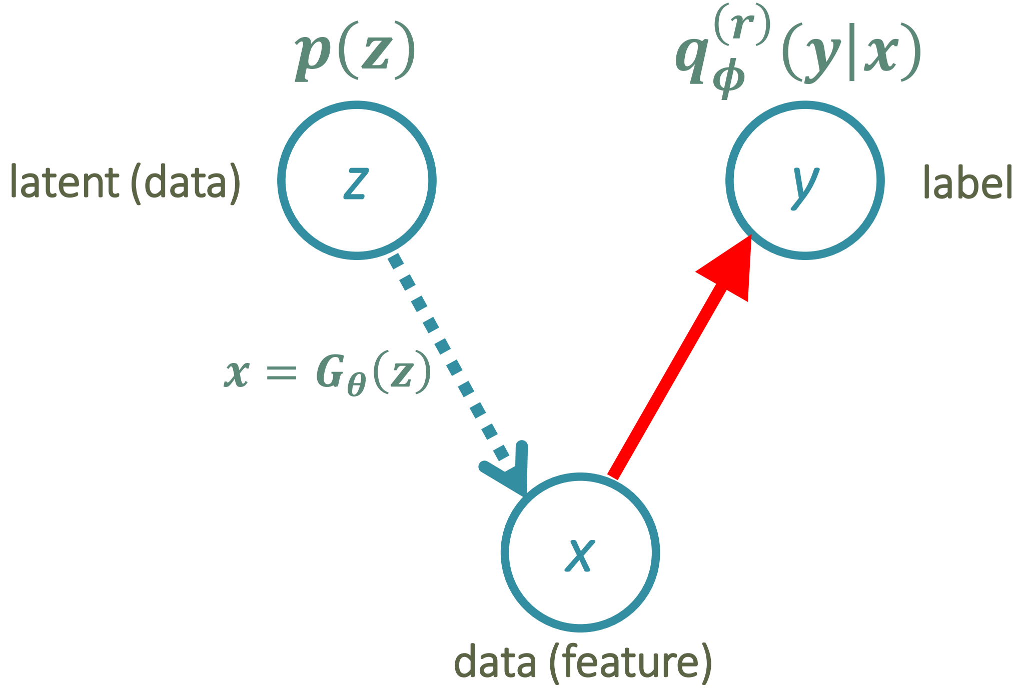


Schematic graphical model

- G_θ – θ are parameters in generator
- D_ϕ – ϕ are parameters in generator

- **Solid line** – generative process
- **Dashed line** – inference process
- **Hollow arrow** – deterministic transformation
- **Red arrow** – adversarial mechanism

- $q_\phi^{(r)}(y|x)$ denotes both $q_\phi(y|x)$ and $q_\phi(1-y|x)$



GAN $\quad y = \begin{cases} 1, & if\ x\ is\ real \\ 0, & if\ x\ is\ fake \end{cases}$

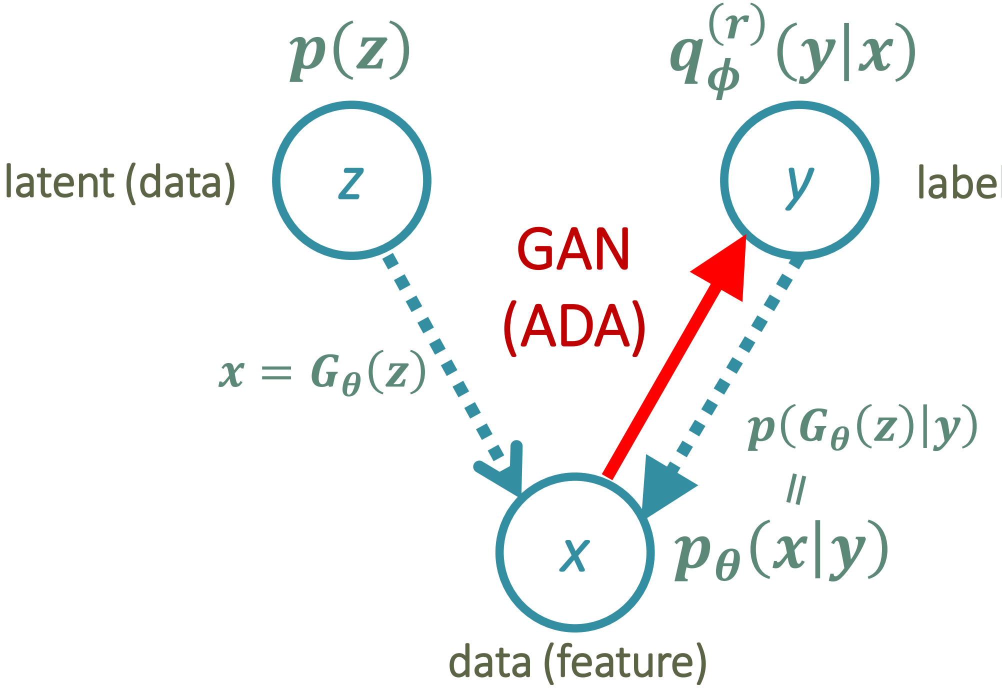
ADA $\quad y = \begin{cases} 1, & if\ x\ is\ in\ source\ domain \\ 0, & if\ x\ is\ in\ target\ domain \end{cases}$

Schematic graphical model

- $\boldsymbol{G_\theta} - \theta$ are parameters in generator
- $\boldsymbol{D_\phi} - \phi$ are parameters in generator

- **Solid line** – generative process
- **Dashed line** – inference process
- **Hollow arrow** – deterministic transformation
- **Red arrow** – adversarial mechanism

- $\boldsymbol{q_\phi^{(r)}(y|x)}$ denotes both $\boldsymbol{q_\phi(y|x)}$ and $\boldsymbol{q_\phi(1-y|x)}$



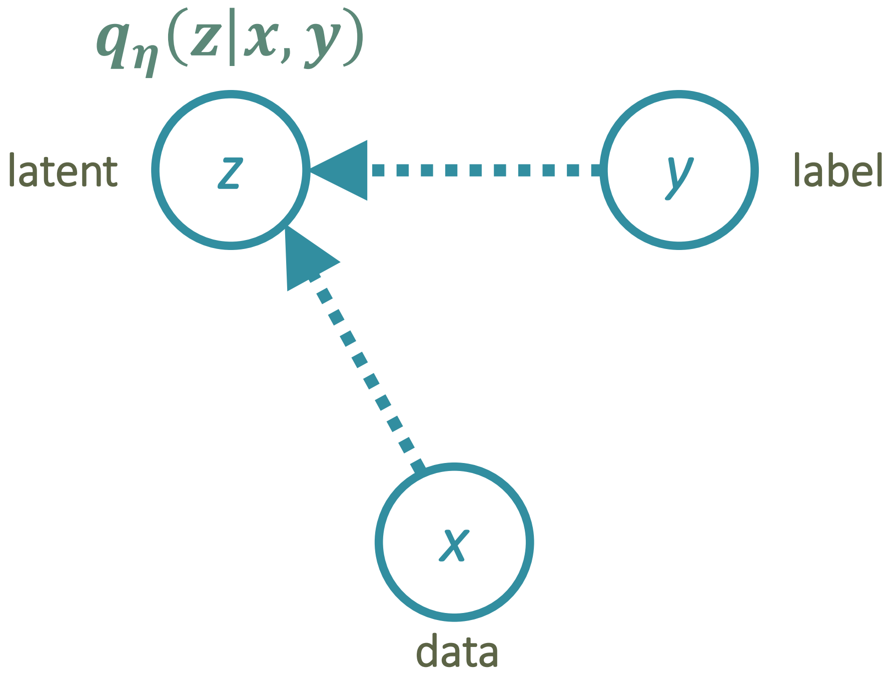
GAN $\quad y = \begin{cases} 1, & \text{if } x \text{ is real} \\ 0, & \text{if } x \text{ is fake} \end{cases}$

ADA $\quad y = \begin{cases} 1, & \text{if } x \text{ is in source domain} \\ 0, & \text{if } x \text{ is in target domain} \end{cases}$

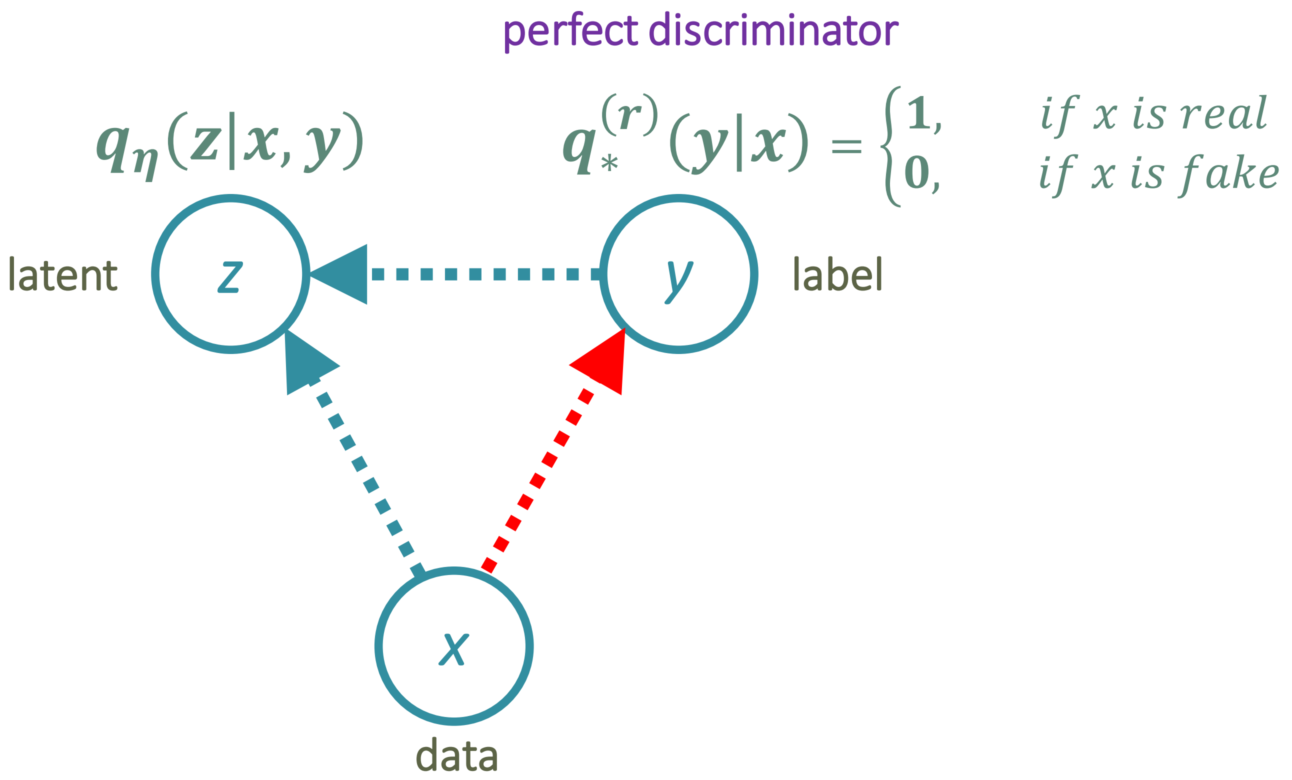
Reformulating DGMs

Using the schematic graphical model representation

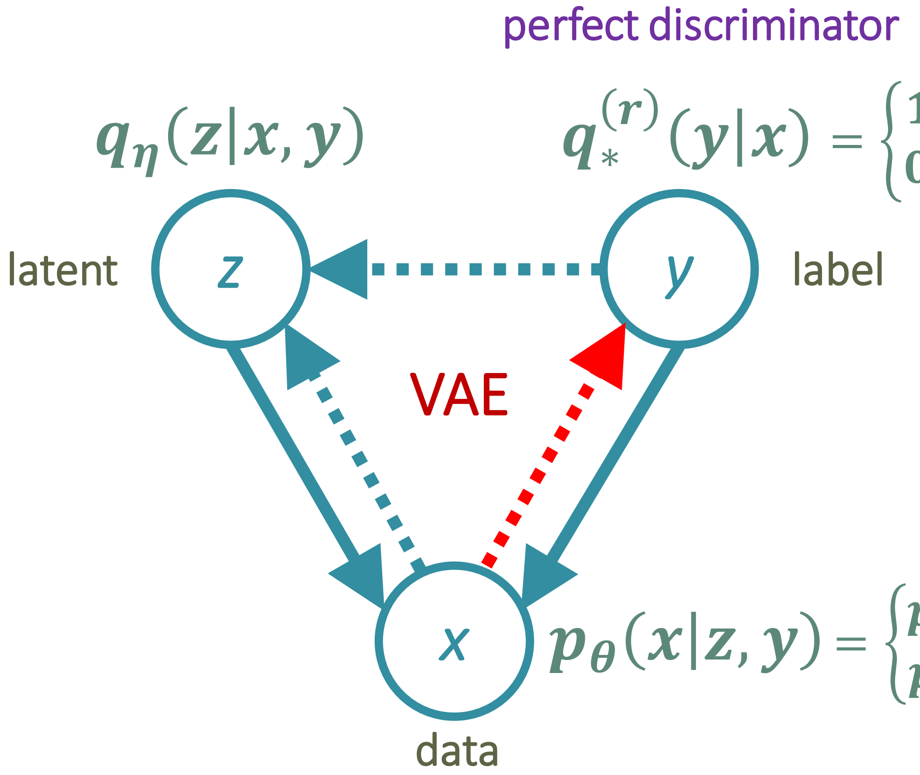
Reformulating VAEs



Reformulating VAEs



Reformulating VAEs



$$q_\eta(\boldsymbol{z}|\boldsymbol{x},\boldsymbol{y})$$

$$\boldsymbol{q}_*^{(\boldsymbol{r})}(\boldsymbol{y}|\boldsymbol{x}) = \begin{cases} \mathbf{1}, & if\ x\ is\ real \\ \mathbf{0}, & if\ x\ is\ fake \end{cases}$$

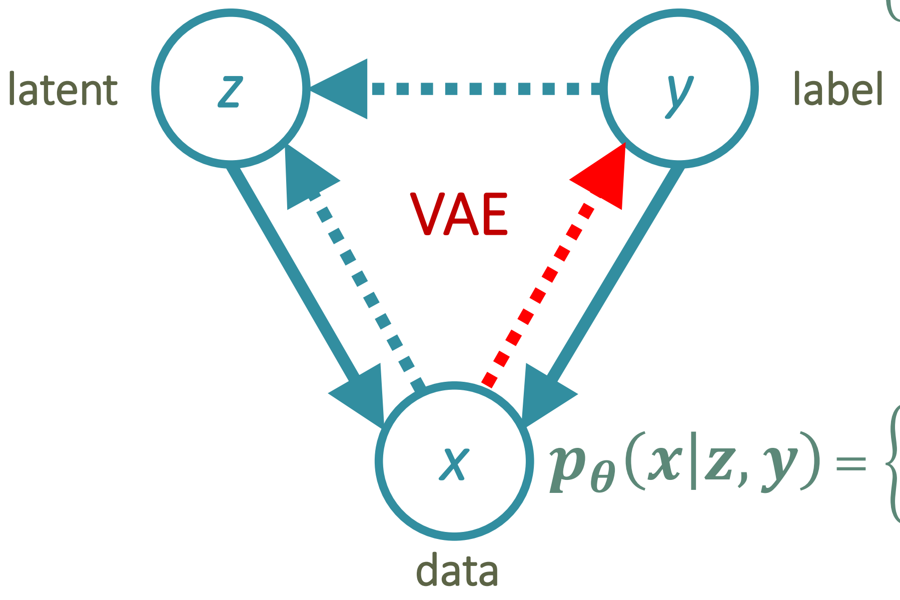
$$\boldsymbol{p}_\theta(\boldsymbol{x}|\boldsymbol{z},\boldsymbol{y}) = \begin{cases} \boldsymbol{p}_\theta(\boldsymbol{x}|\boldsymbol{z}), & if\ y = 0 \\ \boldsymbol{p}_{data}(\boldsymbol{x}), & if\ y = 1 \end{cases}$$

Reformulating VAEs

"VAEs in our interpretation contain a **degenerated adversarial mechanism** that blocks out generated samples and only allows real examples for model training."

perfect discriminator

$$q_\eta(z|x,y) \qquad q_*^{(r)}(y|x) = \begin{cases} 1, & if\ x\ is\ real \\ \cancel{0,} & \cancel{if\ x\ is\ fake} \end{cases}$$



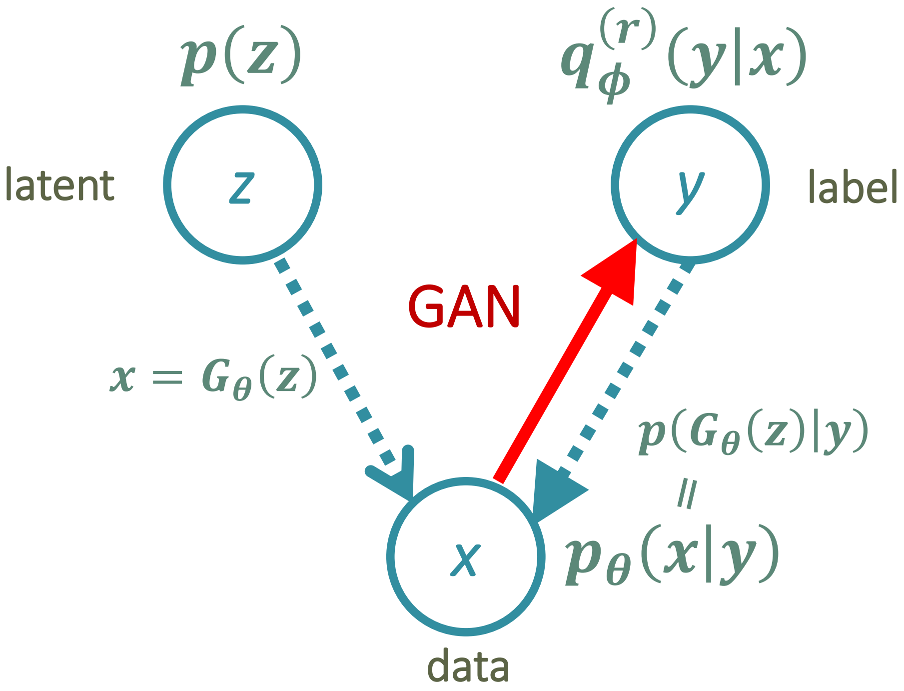
degenerated adversarial discriminator

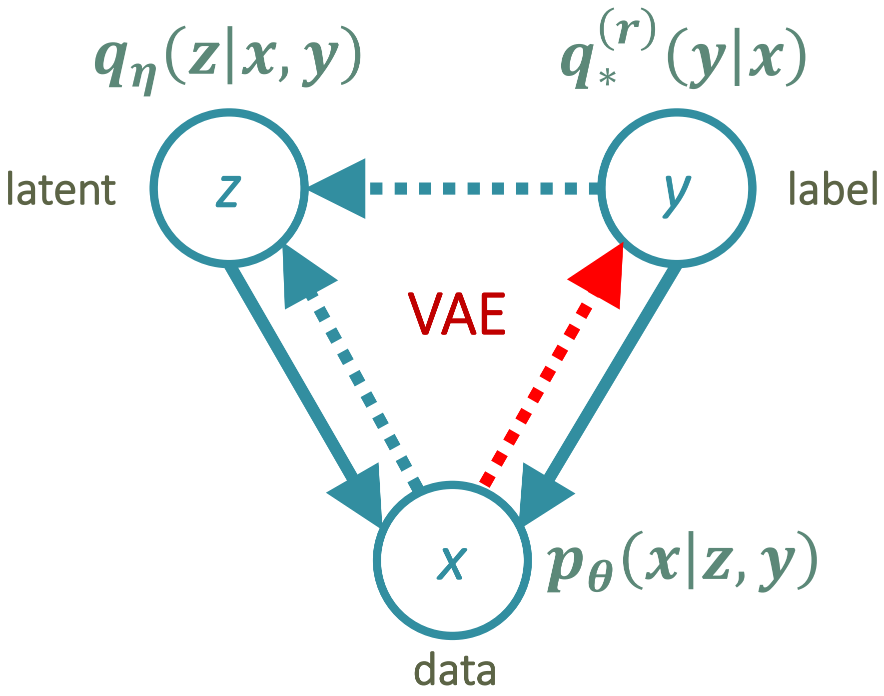
$$p_\theta(x|z,y) = \begin{cases} \cancel{p_\theta(x|z),} & \cancel{if\ y = 0} \\ p_{data}(x), & if\ y = 1 \end{cases}$$

GANs vs VAEs

"We develop a reformulation of GANs that interprets generation of samples as performing posterior inference, leading to an objective that resembles variational inference as in VAEs."

"VAEs in our interpretation contain a degenerated adversarial mechanism that blocks out generated samples and only allows real examples for model training."





Connecting GANs and VAEs

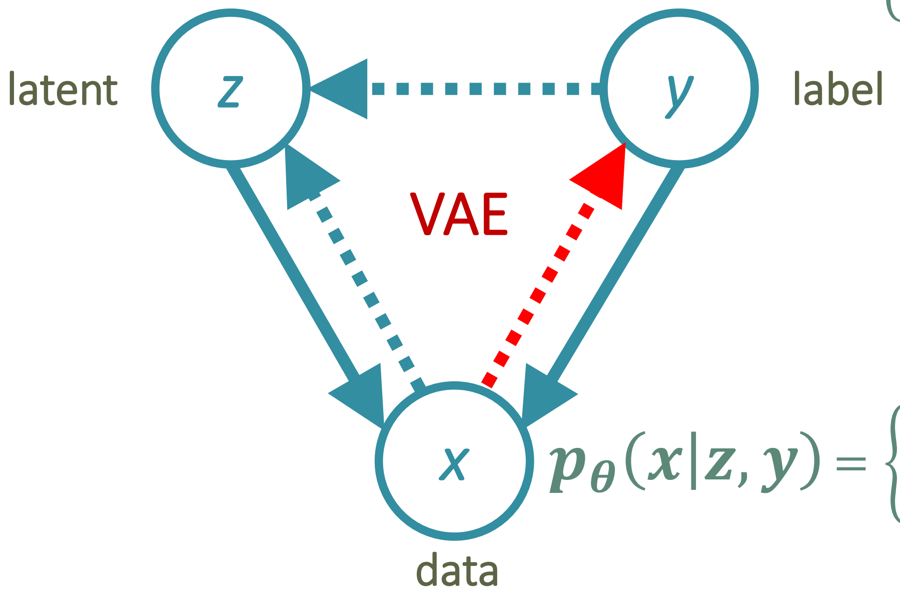
- GANs now also relate to the variational inference algorithm as with VAEs.
- VAEs with also include an adversarial mechanism as in GANs. The discriminator is perfect and degenerated, disabling generated samples to help with learning.
- The generator parameters θ are **placed in the opposite directions in the two KLDs**. The asymmetry of KLD leads to distinct model behaviors.
  - For instance, GANs are able to generate sharp images but tend to collapse to one or few modes of the data (i.e., **mode missing**).
  - In contrast, the KLD of VAEs tends to drive generator to cover all modes of the data distribution but also small-density regions (i.e., **mode covering**), which tend to result in blurred samples.
- GANs and VAEs have inverted latent-visible treatments of (z, y) and x, since we **interpret sample generation in GANs as posterior inference**. Such inverted treatments strongly relates to the symmetry of the sleep and wake phases in the wake-sleep algorithm.

Adversary Activated VAEs (AAVAEs)

perfect discriminator

$$q_*^{(r)}(y|x) = \begin{cases} 1, & if\ x\ is\ real \\ \cancel{0,} & \cancel{if\ x\ is\ fake} \end{cases}$$

$$q_\eta(z|x,y)$$



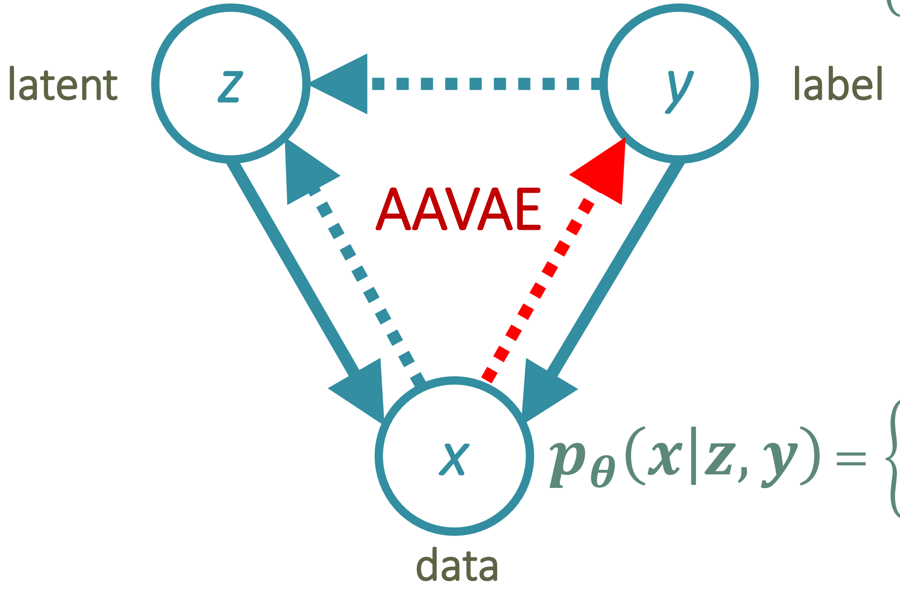
degenerated adversarial discriminator

$$p_\theta(x|z,y) = \begin{cases} \cancel{p_\theta(x|z),} & \cancel{if\ y = 0} \\ p_{data}(x), & if\ y = 1 \end{cases}$$

Adversary Activated VAEs (AAVAEs)

perfect discriminator

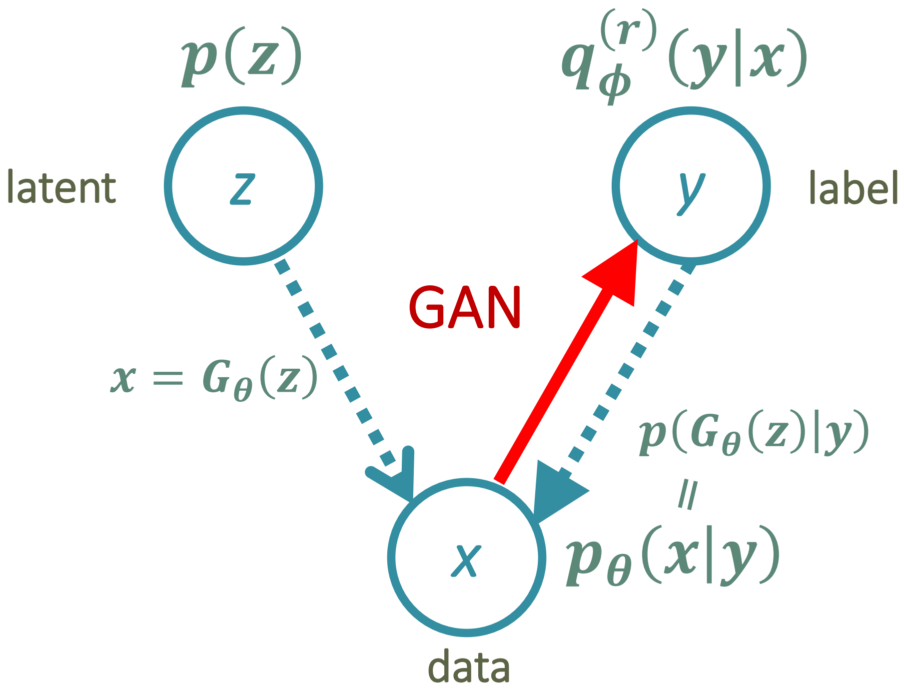
$$q_\eta(\boldsymbol{z}|\boldsymbol{x},\boldsymbol{y}) \qquad q_*^{(r)}(\boldsymbol{y}|\boldsymbol{x}) = \begin{cases} \mathbf{1}, & if\ x\ is\ real \\ \mathbf{0}, & if\ x\ is\ fake \end{cases}$$

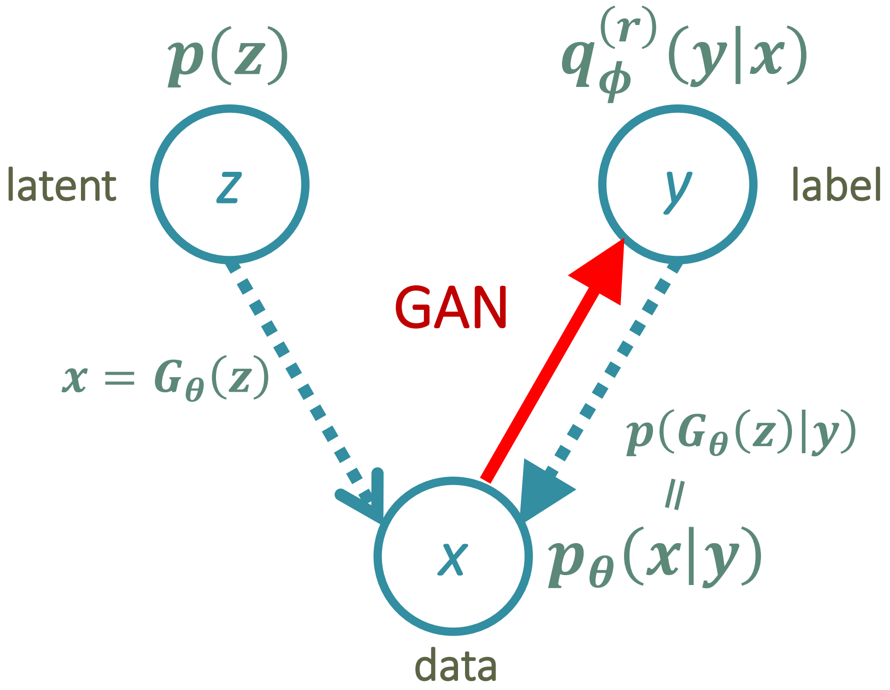


$$p_\theta(\boldsymbol{x}|\boldsymbol{z},\boldsymbol{y}) = \begin{cases} p_\theta(\boldsymbol{x}|\boldsymbol{z}), & if\ y = 0 \\ p_{data}(\boldsymbol{x}), & if\ y = 1 \end{cases}$$

degenerated adversarial discriminator → adversary activated VAEs (AAVAEs)

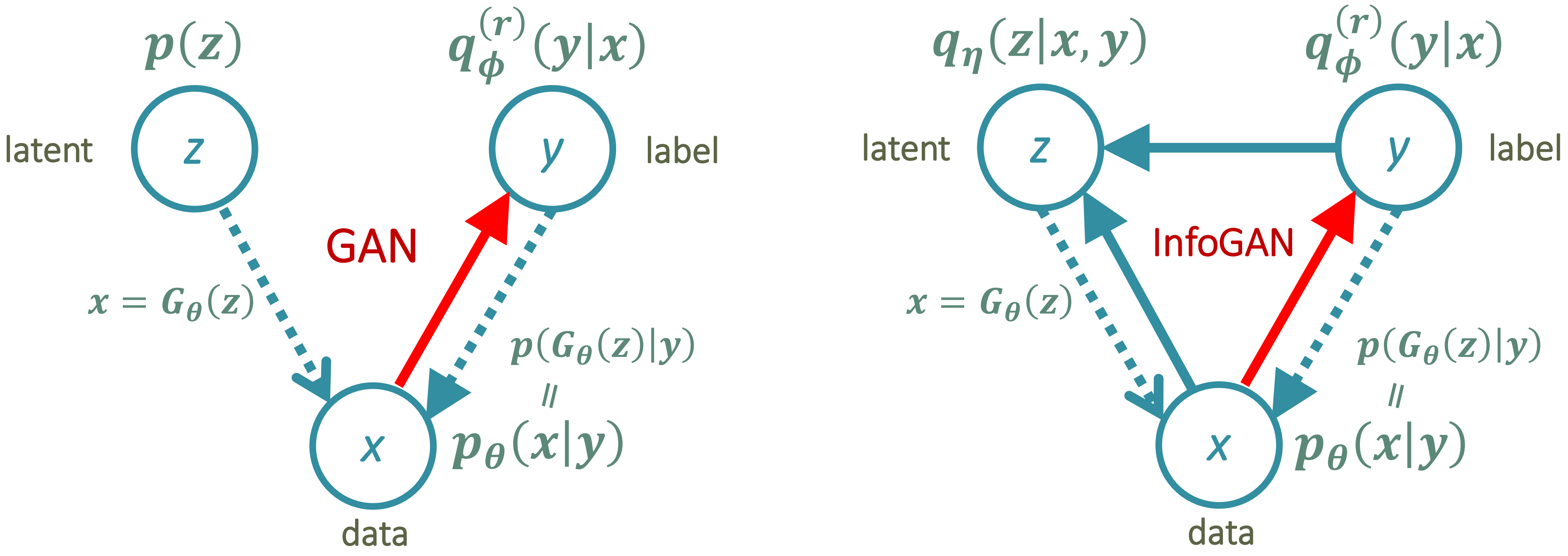
GANs vs InfoGANs





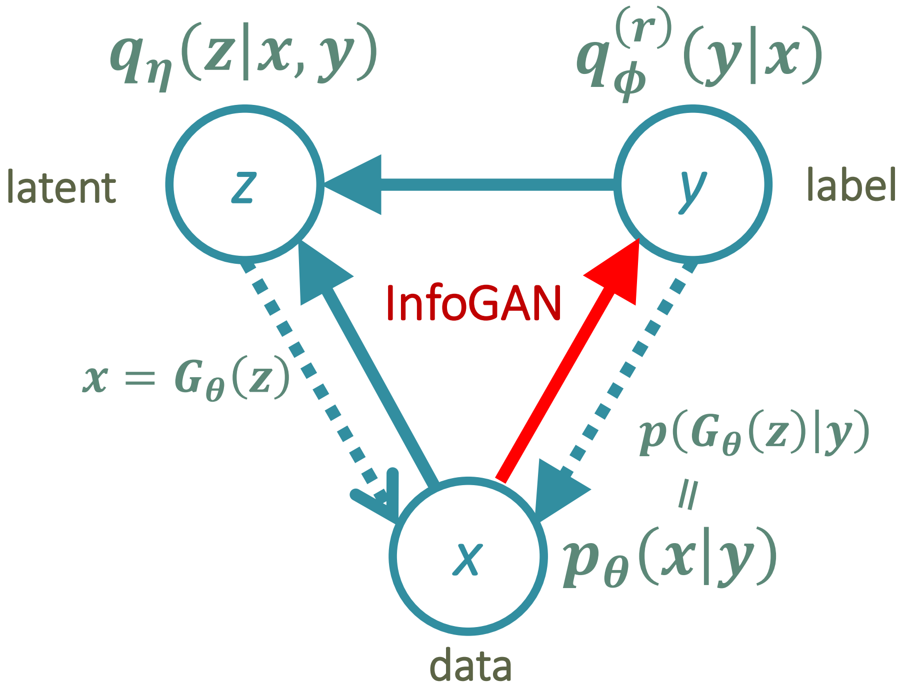
GANs vs InfoGANs

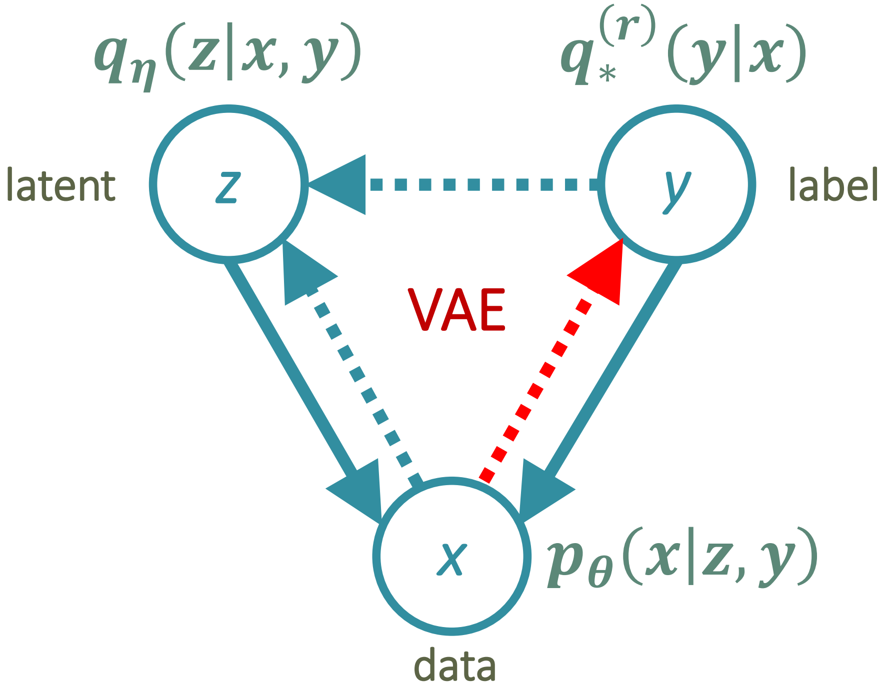
"Schematic graphical model of InfoGAN, which, compared to GANs, adds conditional generative process of code z with distribution $q_\eta(z|x,y)$."



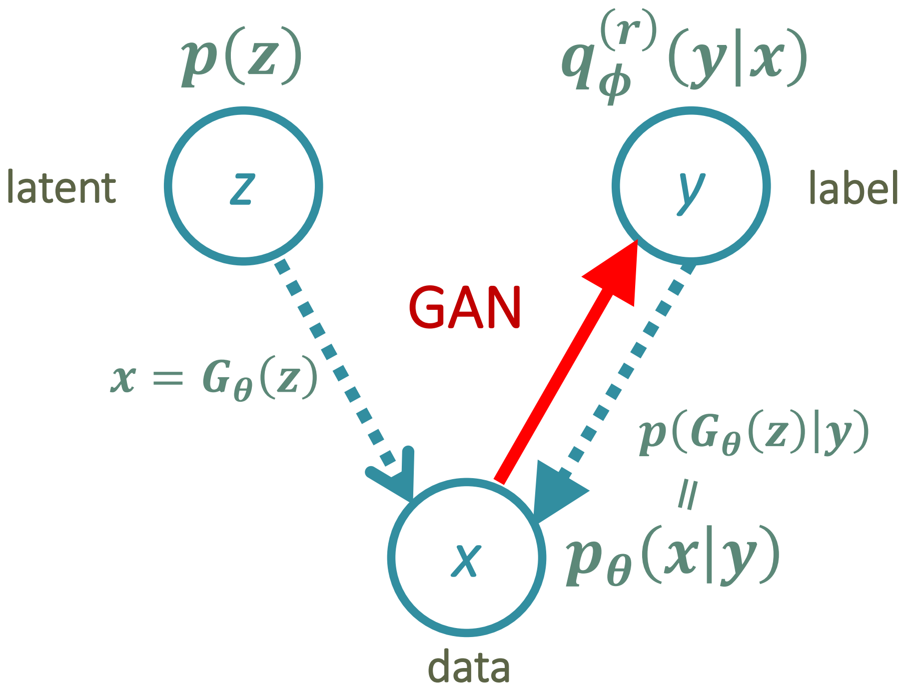
InfoGANs vs VAEs

“Schematic graphical model of VAEs, which is obtained by swapping the generative and inference processes of InfoGAN.”

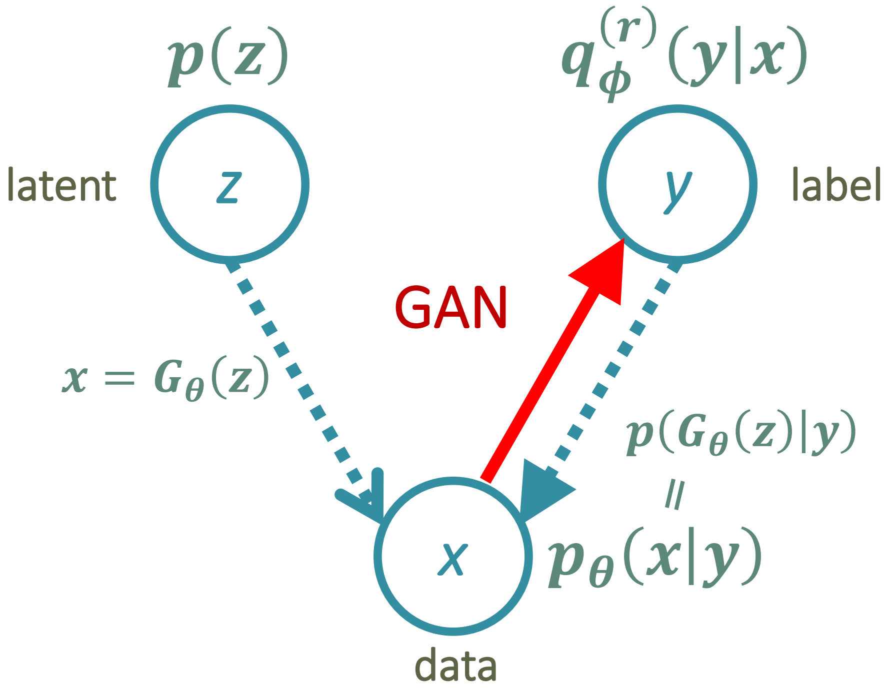




GANs vs AAEs

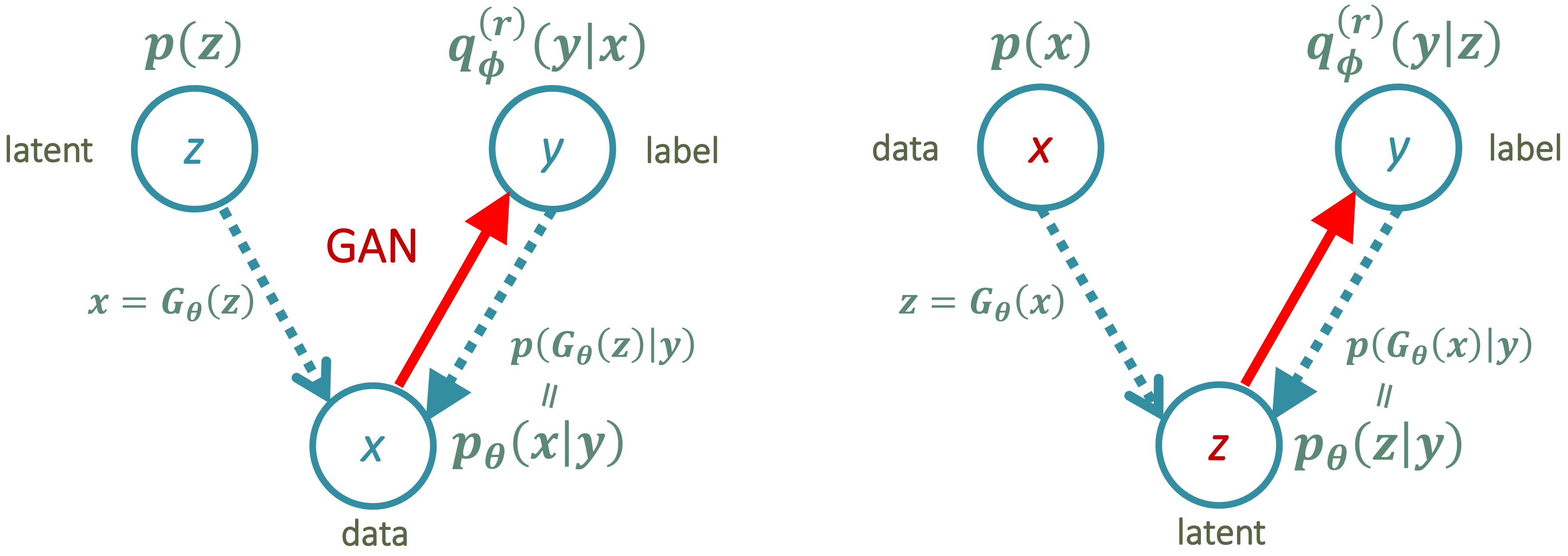


$p(z)$
$q_\phi^{(r)}(y|x)$
latent
z
y
label
GAN
$x = G_\theta(z)$
$p(G_\theta(z)|y)$
||
x
$p_\theta(x|y)$
data

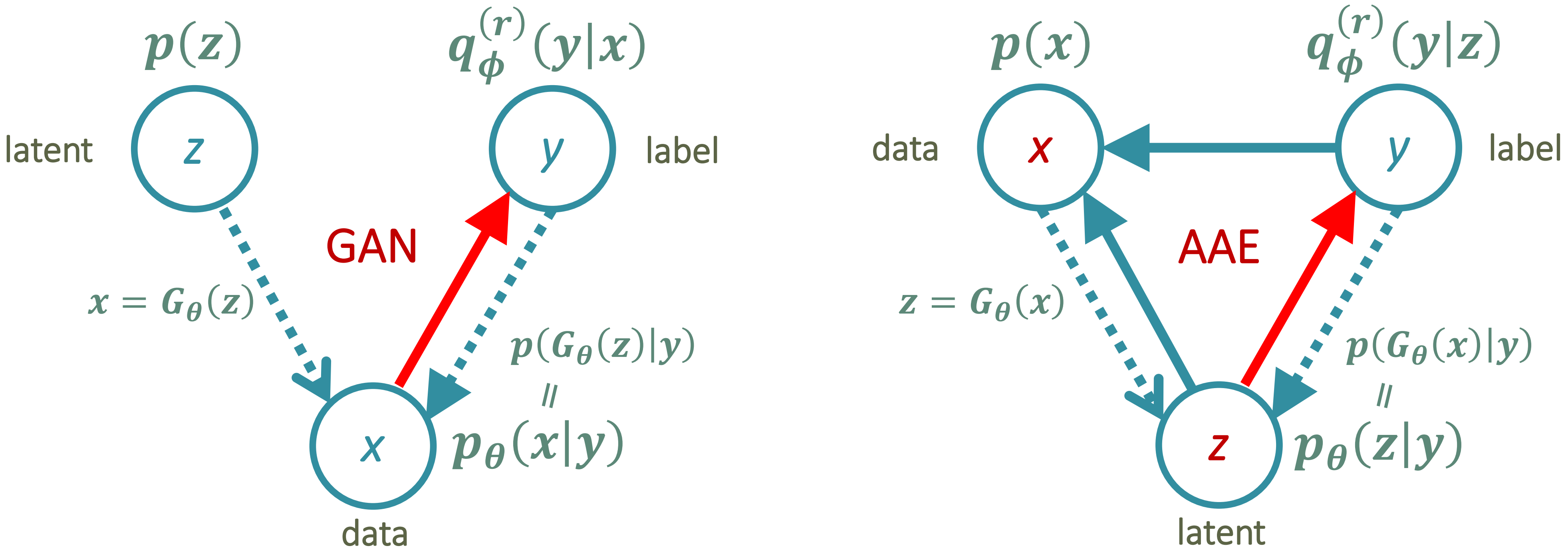


$p(z)$
$q_\phi^{(r)}(y|x)$
latent
z
y
label
GAN
$x = G_\theta(z)$
$p(G_\theta(z)|y)$
||
x
$p_\theta(x|y)$
data

GANs vs AAEs


$p(z)$
$q_\phi^{(r)}(y|x)$
latent
z
y
label
GAN
$x = G_\theta(z)$
$p(G_\theta(z)|y)$
=
$p_\theta(x|y)$
x
data
$p(x)$
$q_\phi^{(r)}(y|z)$
data
x
y
label
$z = G_\theta(x)$
$p(G_\theta(x)|y)$
=
$p_\theta(z|y)$
z
latent

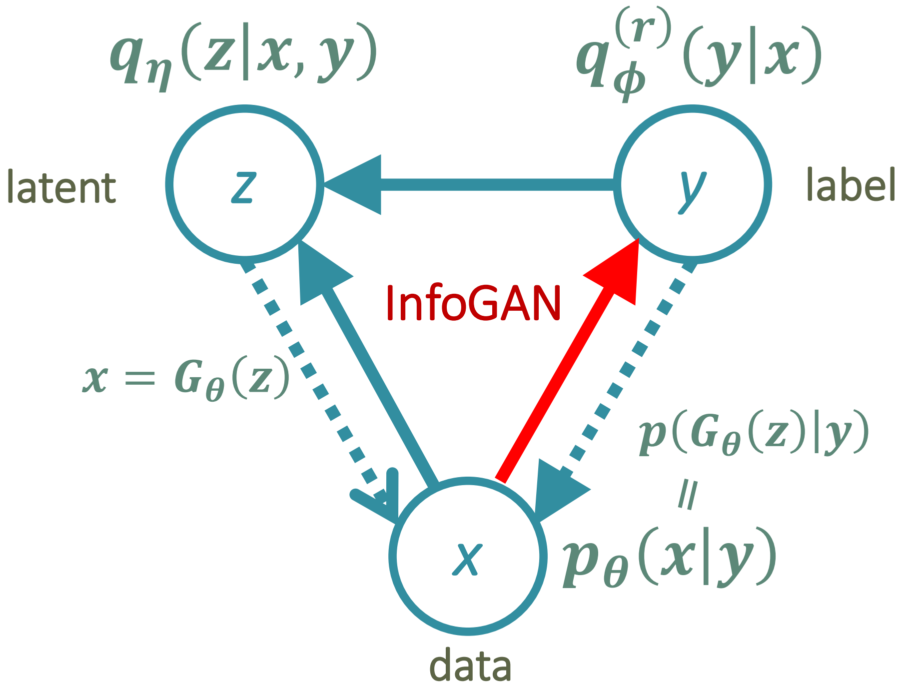
GANs vs AAEs

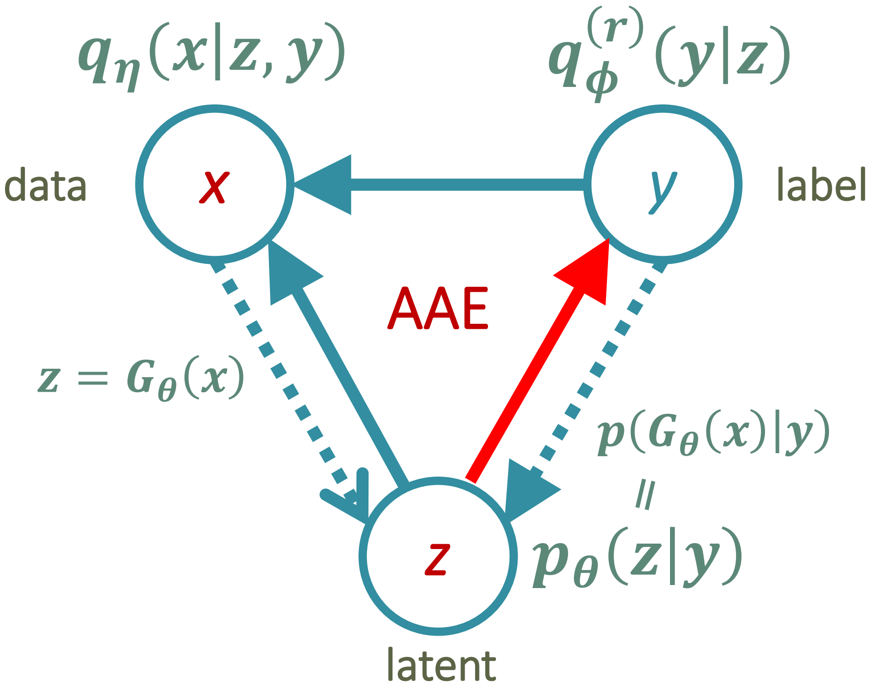


$p(z)$
$q_\phi^{(r)}(y|x)$
latent
z
y
label
GAN
$x = G_\theta(z)$
$p(G_\theta(z)|y)$
$\|$
x
$p_\theta(x|y)$
data
$p(x)$
$q_\phi^{(r)}(y|z)$
data
x
y
label
AAE
$z = G_\theta(x)$
$p(G_\theta(x)|y)$
$\|$
z
$p_\theta(z|y)$
latent

InfoGANs vs AAEs

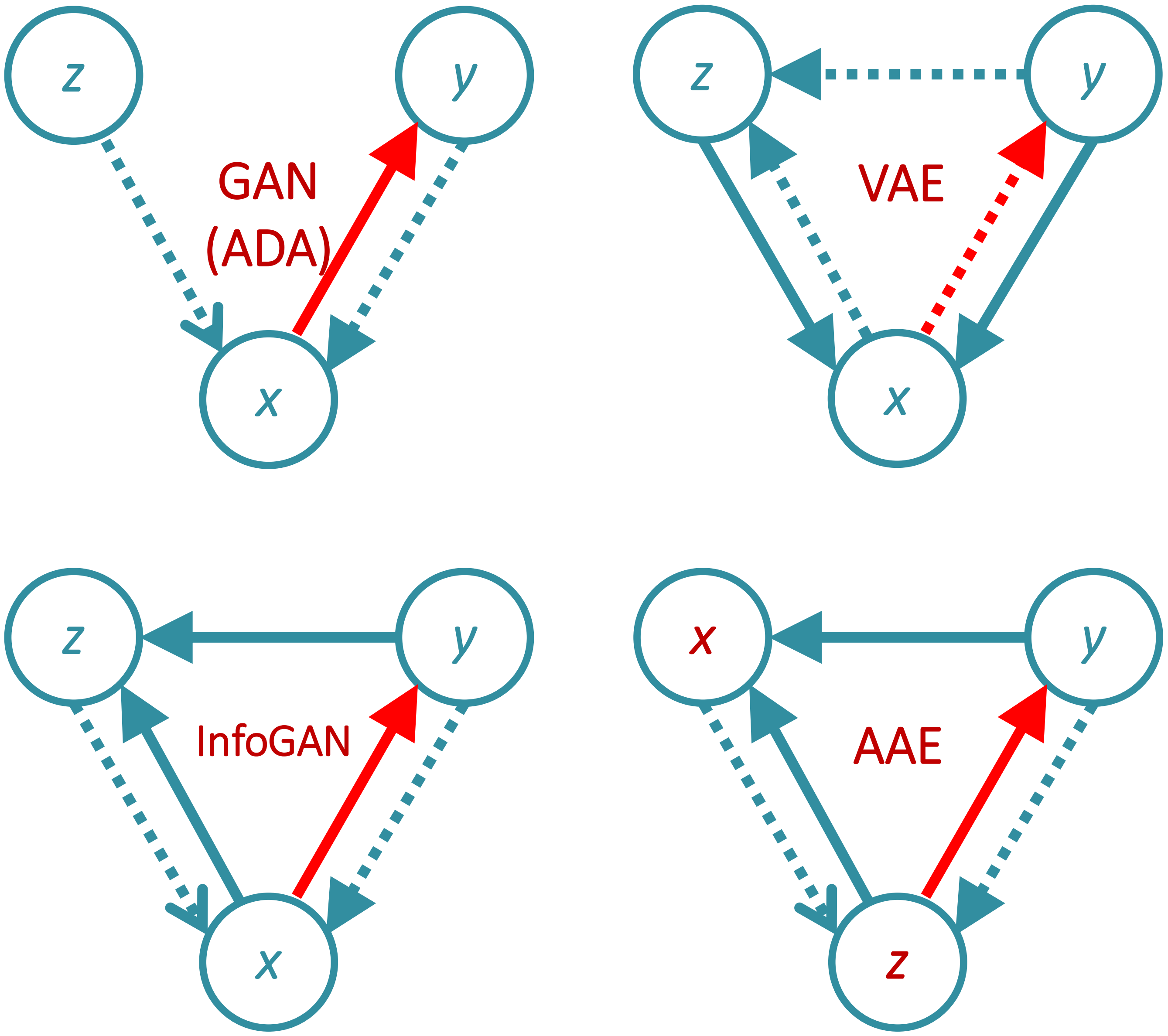
"Schematic graphical model of Adversarial Autoencoder (AAE), which is obtained by **swapping data x and code z** in InfoGAN."





Summary

- The schematic graphical model representation reveals some interesting connections among different DGMs.

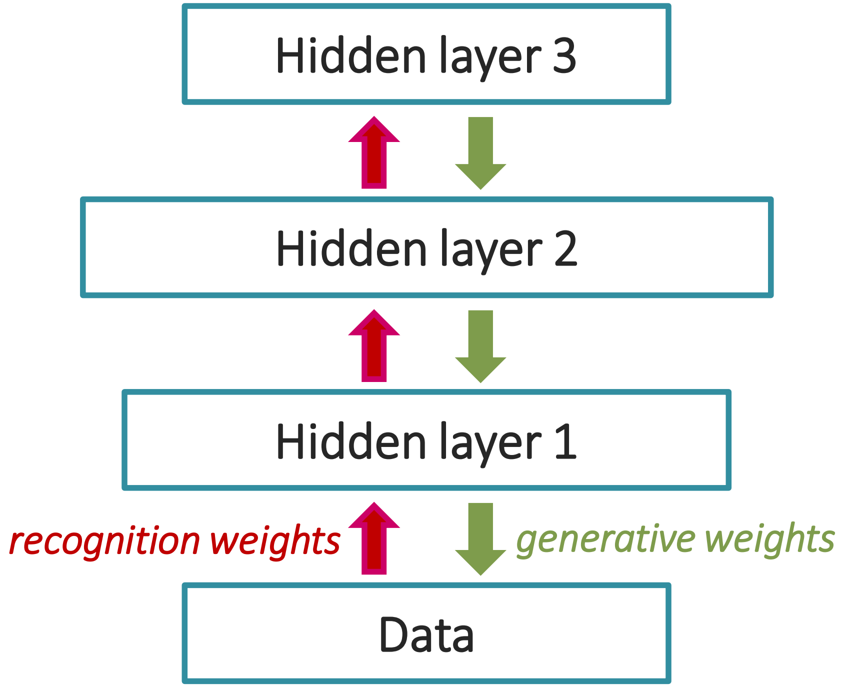


Discussions

Connection to the wake-sleep algorithm
Similarities and differences between visible and latent variables

Wake-sleep (WS) algorithm

- Wake phase
  - Use ***recognition weights*** for bottom-up pass
  - Train the ***generative weights*** to reconstruct activities in each layer from the layer above
  - $\max_{\theta} \mathbb{E}_{q_\lambda(h|x)p_d(x)}[\log p_\theta(x|h)]$
- Sleep phase
  - Use ***generative weights*** to generate samples
  - Train the ***recognition weights*** to reconstruct activities in each layer from the layer below
  - $\max_{\lambda} \mathbb{E}_{p_\theta(x|h)p(h)}[\log q_\lambda(h|x)]$



Connections between VAEs and WS

(Wake phase) $\max_{\theta} \mathbb{E}_{q_\lambda(h|x)p_d(x)}[\log p_\theta(x|h)]$

(VAEs) $\max_{\theta,\eta} \mathbb{E}_{q_\eta(z|x)p_d(x)}[\log p_\theta(x|z)] \; \underbrace{- \mathbb{E}_{p_d(x)}\big[\mathrm{KL}\big(q_\eta(\mathrm{z}|x) \| p(z)\big)\big]}_{\text{an additional prior regularization on the latent variables}}$

also optimize the inference model

Connections between GANs and WS

(Sleep phase) $\max_{\lambda} \mathbb{E}_{p_\theta(x|h)p(h)}[\log q_\lambda(h|x)]$

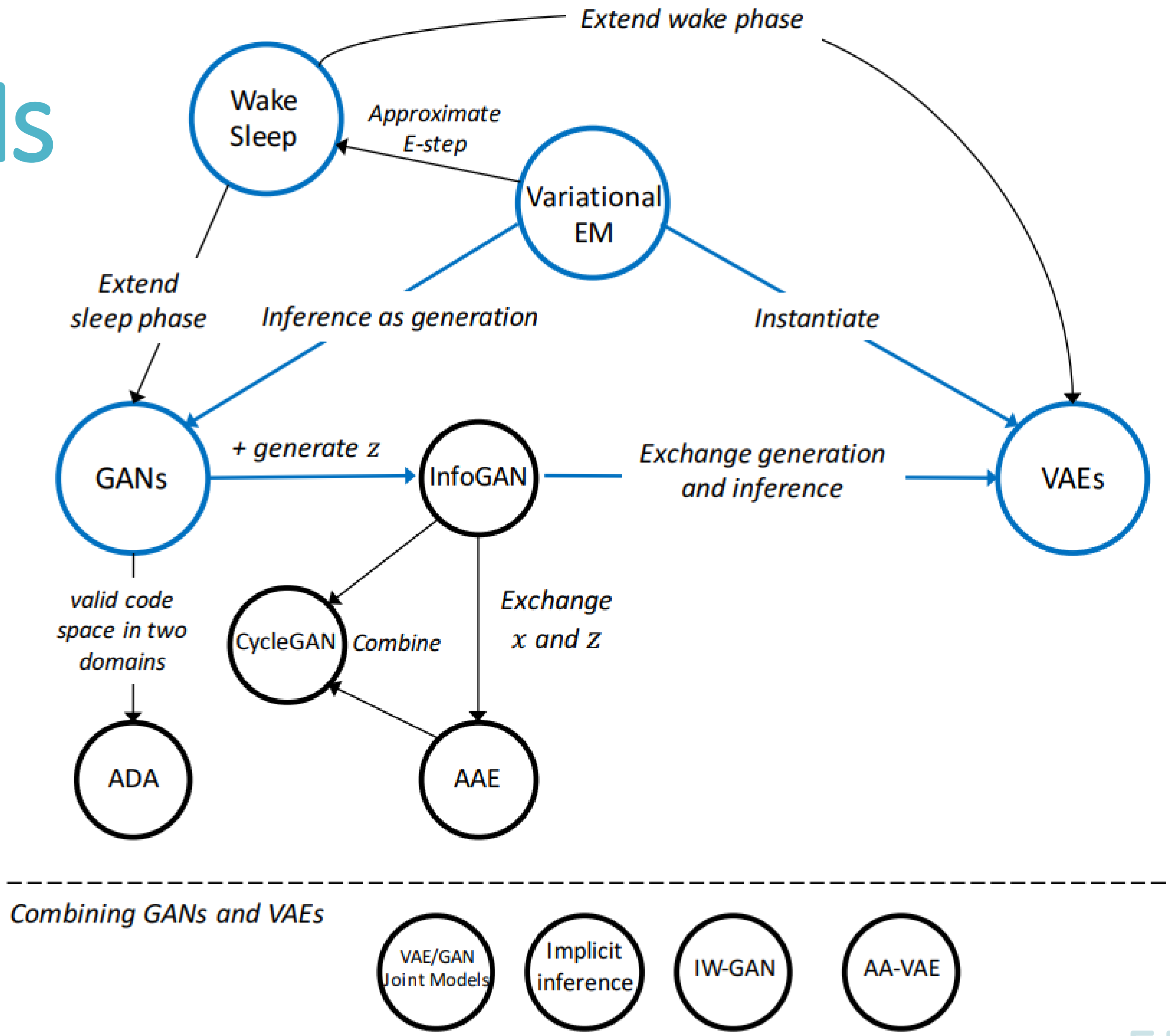
(GANs) $\max_{\phi} \mathbb{E}_{p_\phi(x|y)p(y)}[\log q_\phi(y|x)]$

the discriminator training resembles the sleep phase

$\max_{\theta} \mathbb{E}_{p_\theta(x|y)p(y)}[\log q_\phi(1-y|x)]$

also optimize the generative model to reconstruct $1-y$

Relations of DGMs



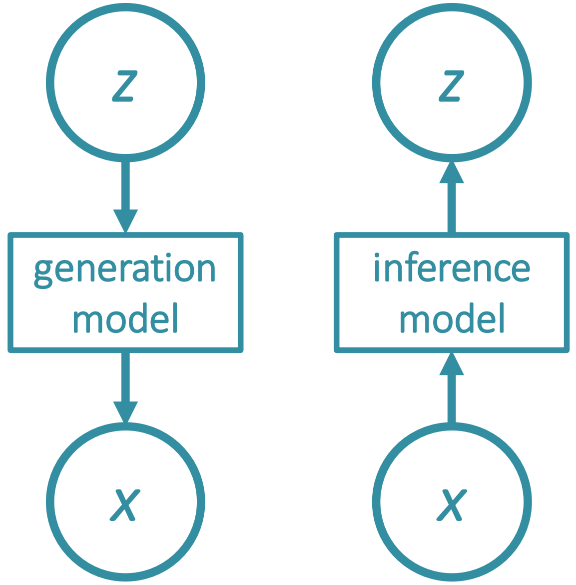
Symmetric view on visibles and latents

- Traditional modeling approaches
  - usually distinguish between latent and visible variables clearly
  - *treat them in very different ways*
- Classic wake-sleep algorithm
  - Visible and latent variables are ***treated in a completely symmetric manner***
    - **Wake phase**: reconstruct visible variables conditioned on latent variables
    - **Sleep phase**: reconstruct latent variables conditioned on visible variables

Symmetric view on visibles and latents

- Sschematic graphical model representation
  - **Visible variables**—*sampled from some (empirical) data distribution*
  - **Latent variables**—*sampled from some prior distribution*
  - **Inference**—*mapping from visible to latent variables*
  - **Generation**—*mapping from latent to visible variables*
- Treating visible and latent variables as a symmetric pair
  - reveals interesting connections among different DGMs
  - helps with modeling and understanding

$x \sim p_{data}(x)$
$z \sim f'_{black-box}(x)$



$z \sim p_{prior}(x)$
$x \sim f_{black-box}(z)$

Differences between visibles and latents

Visible space	Latent space
high-dimensional	low-dimensional (manifold assumption)
complex	simple (sometimes designed to be)
implicit (easy to draw samples from but intractable for evaluating likelihood)	explicit (amenable to likelihood evaluation)

can also be implicit with recent tools
for implicit generative modeling
(e.g., adversarial losses)

Differences between visibles and latents

- Differences between visible and latent variables might be *intentionally introduced*.
- **For feasible likelihood evaluation**
  - Recent tools can implicitly model distributions
- **For enforcing prior beliefs on latent manifolds**
  - Priors should be reasonable
  - But sometimes we are just guessing
- ***Choose the model that best suits your needs!***

References

- Z. Hu, Z. Yang, R. Salakhutdinov and E. P. Xing, "**On unifying deep generative models**," *ICLR,* 2018.
- D. P. Kingma and M. Welling, "**Auto-Encoding Variational Bayes**," *ICLR*, 2014.
- I. J. GoodFellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville and Y. Bengio, "**Generative Adversarial Nets**," *NeurIPS*, 2014.
- A. Makhzani, J. Shlens, N. Jaitly, I. Goodfellow and B. Frey, "**Adversarial Autoencoders**," *ICLR*, 2016.
- X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever and P. Abbeel, "**InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets**," *NeurIPS*, 2016.
- G. E. Hinton, P. Dayan, B. J. Frey and R. M. Neal, "**The wake-sleep algorithm for unsupervised neural networks**," *Science*, 1995.