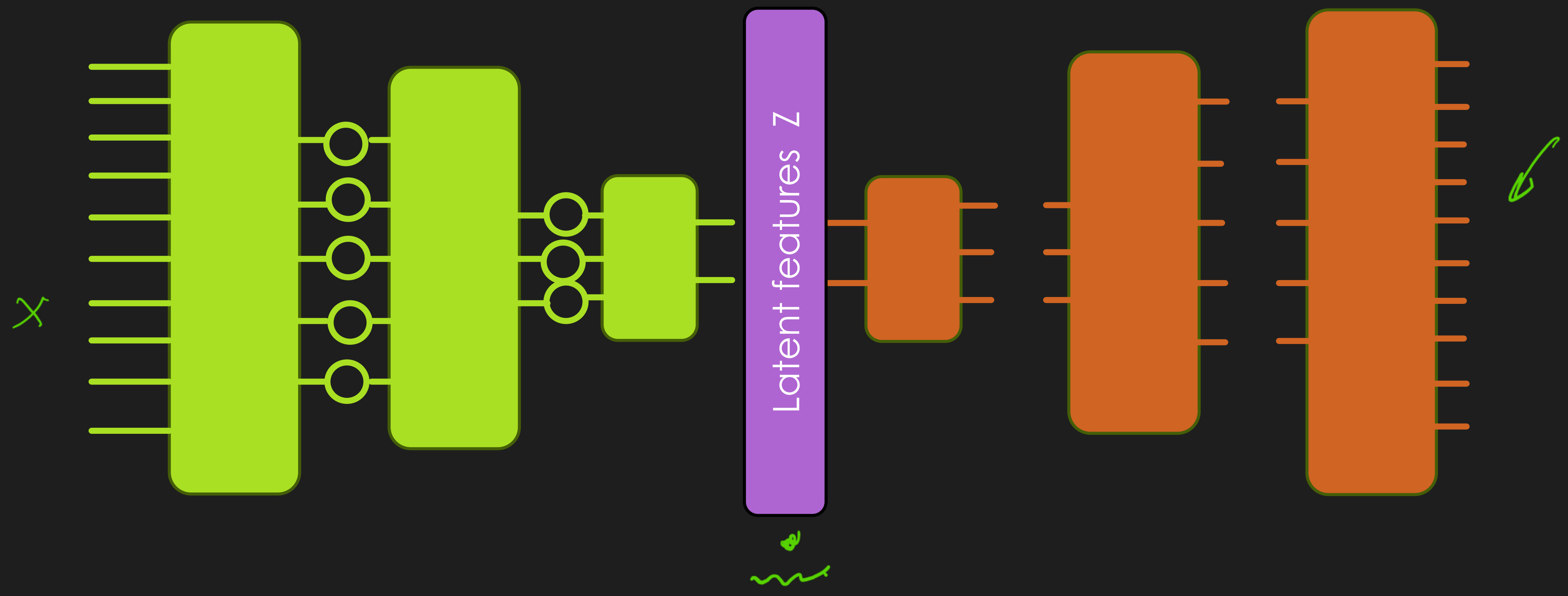
INTRODUCTION TO MACHINE LEARNING COMPSCI 4ML3

Lecture 27

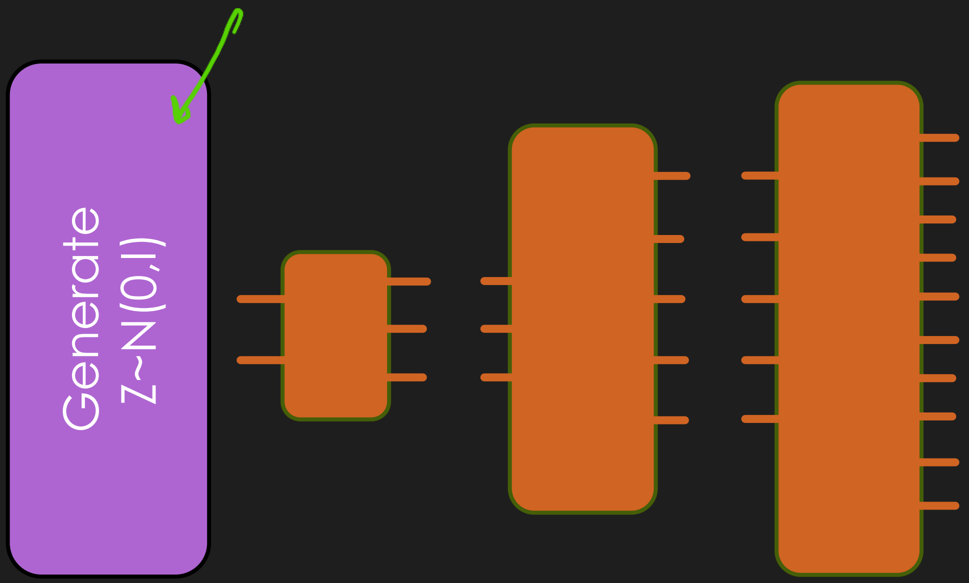
Hassan Ashtiani

REVIEW AUTOENCODER

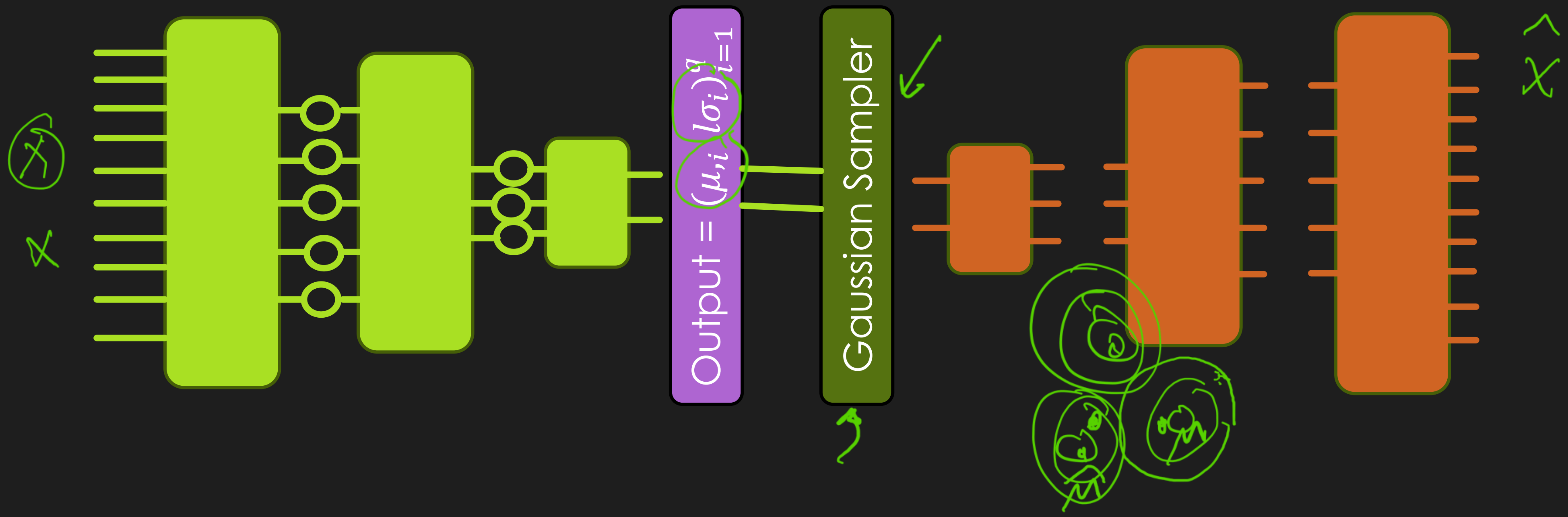


REVIEW: TAMING AUTOENCODERS

- Idea: train autoencoder in a way that the latent space distribution looks like isotopic Gaussian(?!)
  - Decoder learns to turn Gaussian noise into new images
- For generating new images, simply feed Gaussian noise to the decoder



SIMPLIFIED VARIATIONAL AUTOENCODER



$$\text{MIN}\left(\sum_{x}\|Dec(Sampler(Enc(x))) - x\|_2^2 + \lambda \sum_{i=1}^{q}\left(-\text{LOG}\left(\sigma_i^2\right) + \sigma_i^2 + \mu_i^2\right)\right)$$

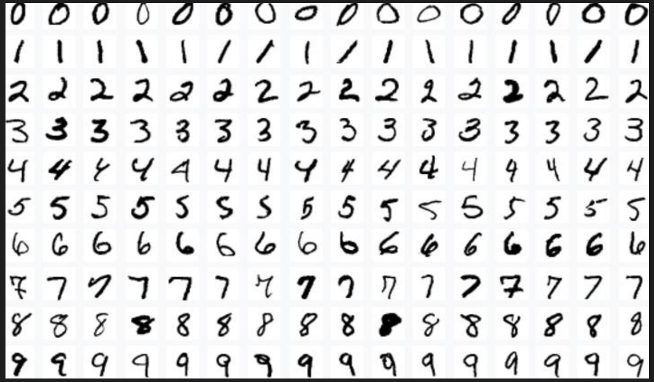
VARIATIONAL AUTOENCODERS

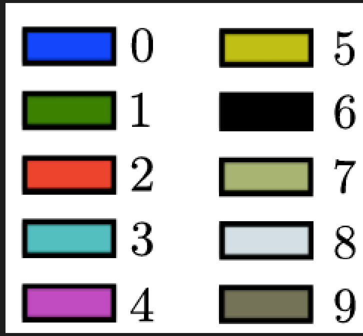
- VAE's original formulation is more complicated
  - Based on evidence lower bound [ELBO], which is not covered in this course
- The generated images are sometimes blurry…
  - Although $\boxed{P(z|x)}$ is Gaussian by design…
    - $P(z)$ may not be… $p(z) = \sum p(z,x) = \sum_x p(z|x)p(x)$
  - Other issues?

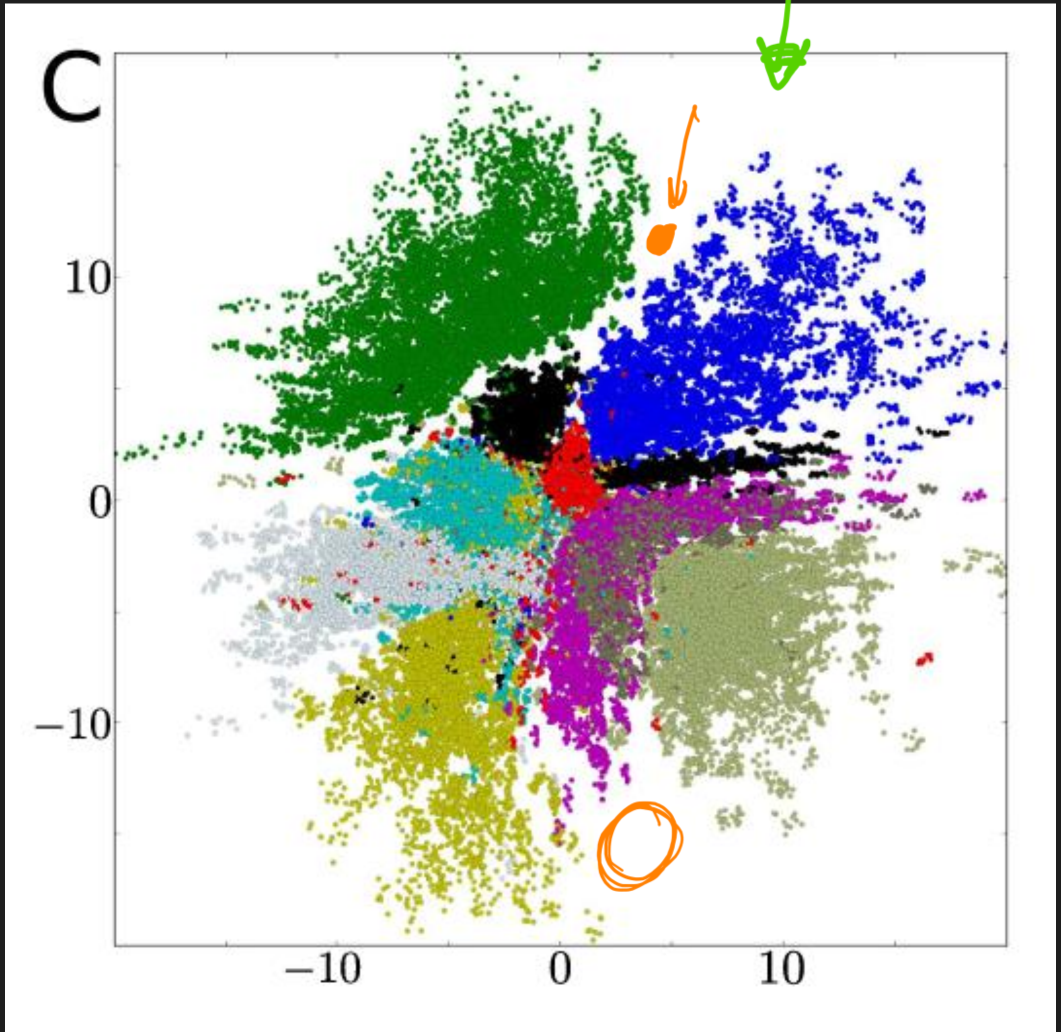
* Minimizing reconstruction error, …

LATENT SPACE VISUALIZATION

- VAE FOR MNIST







MAKHZANI ET AL

-

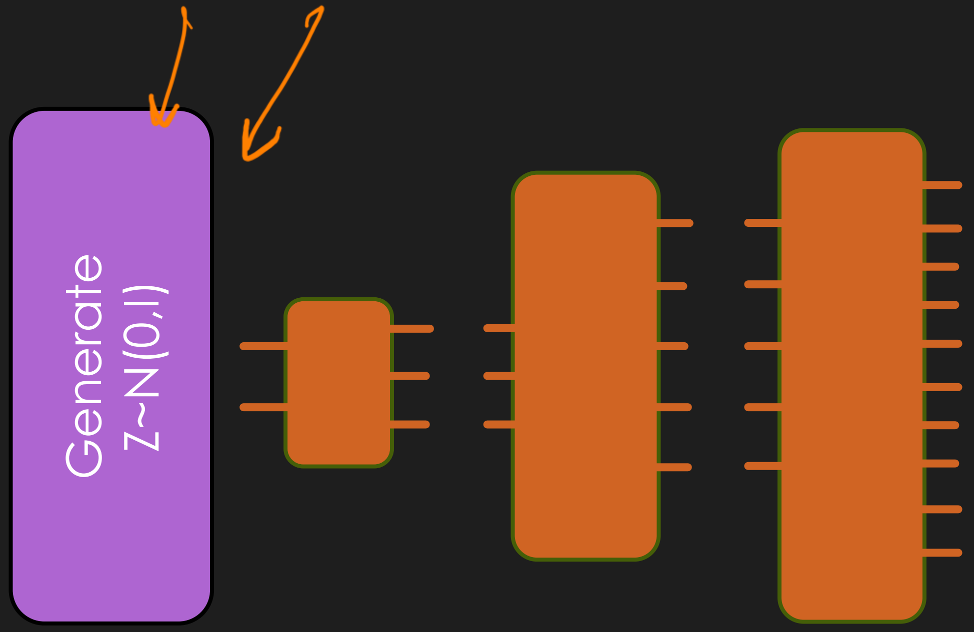
GENERATIVE ADVERSARIAL NETWORKS

- Can generate realistic images
  - But a bit hard to train

- HTTPS://THISPERSONDOESNOTEXIST.COM/

GENERATIVE ADVERSARIAL NETWORKS

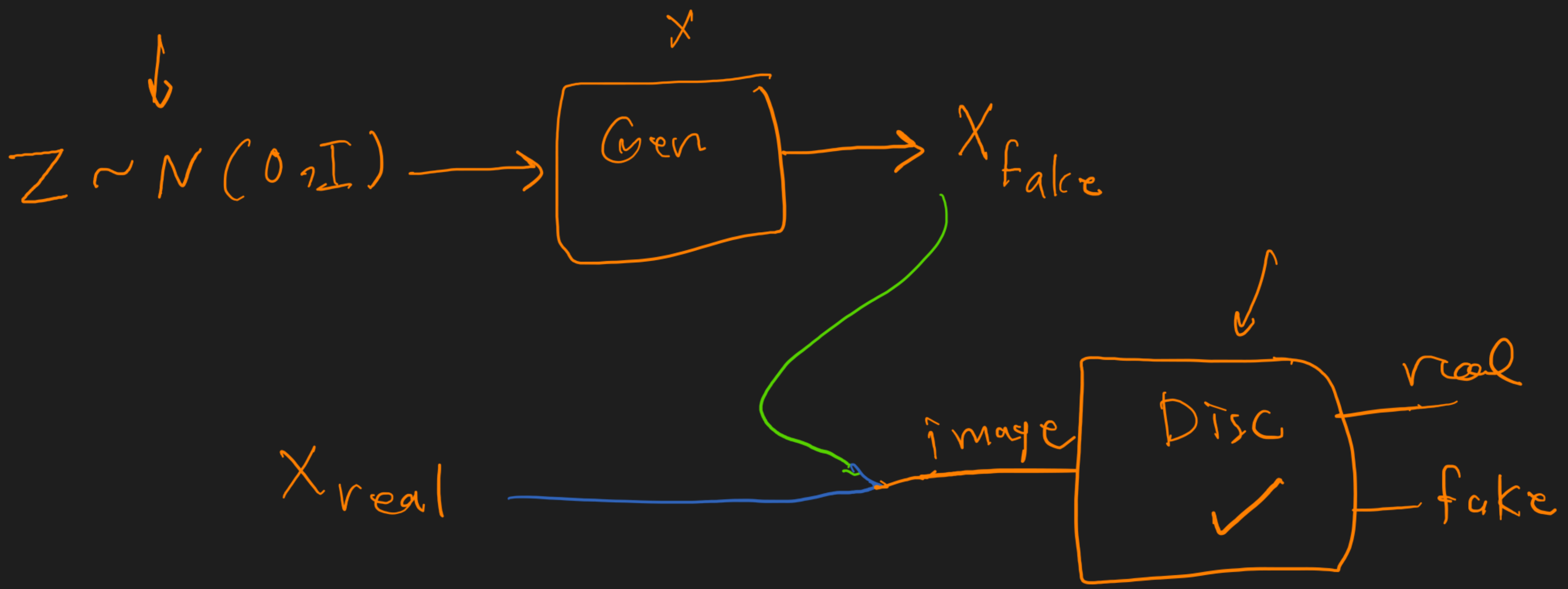
- Turn Gaussian samples into realistic images?
  - Only train a decoder (so no latent rep)
  - But how to train it?
- IDEA: a set of generated images are realistic if it is hard to distinguish them from real images
  - Train a classifier that tries to distinguish real from fake images



ADVERSARIAL TRAINING

- GENERATE GENERATES AN IMAGE
  - GOAL: FOOL THE ADVERSARY

- THE ADVERSARY RECEIVES AN IMAGE AND DECIDES WHETHER THE IMAGE WAS REAL OR FAKE
  - GOAL: CLASSIFY CORRECTLY



$z \sim N(0, I)$
Gen
X_{fake}
X_{real}
image
Disc
real
fake

- Discriminator's objective
  - $\max_{Disc} \sum_{x \in Data} \log Disc(x) + \sum_{z \in N(0,I)} \log(1 - Disc(G(z)))$
- Generator's objective
  - $\min_{G} \sum_{z \in N(0,I)} \log(1 - Disc(G(z)))$
- Train them at the same time
  - Delicate optimization, can be unstable
  - Ideally, the generator would converge to a solution that manages to fool the discriminator

label of fake : 0
label of real : 1

MODE COLLAPSE

Examples:

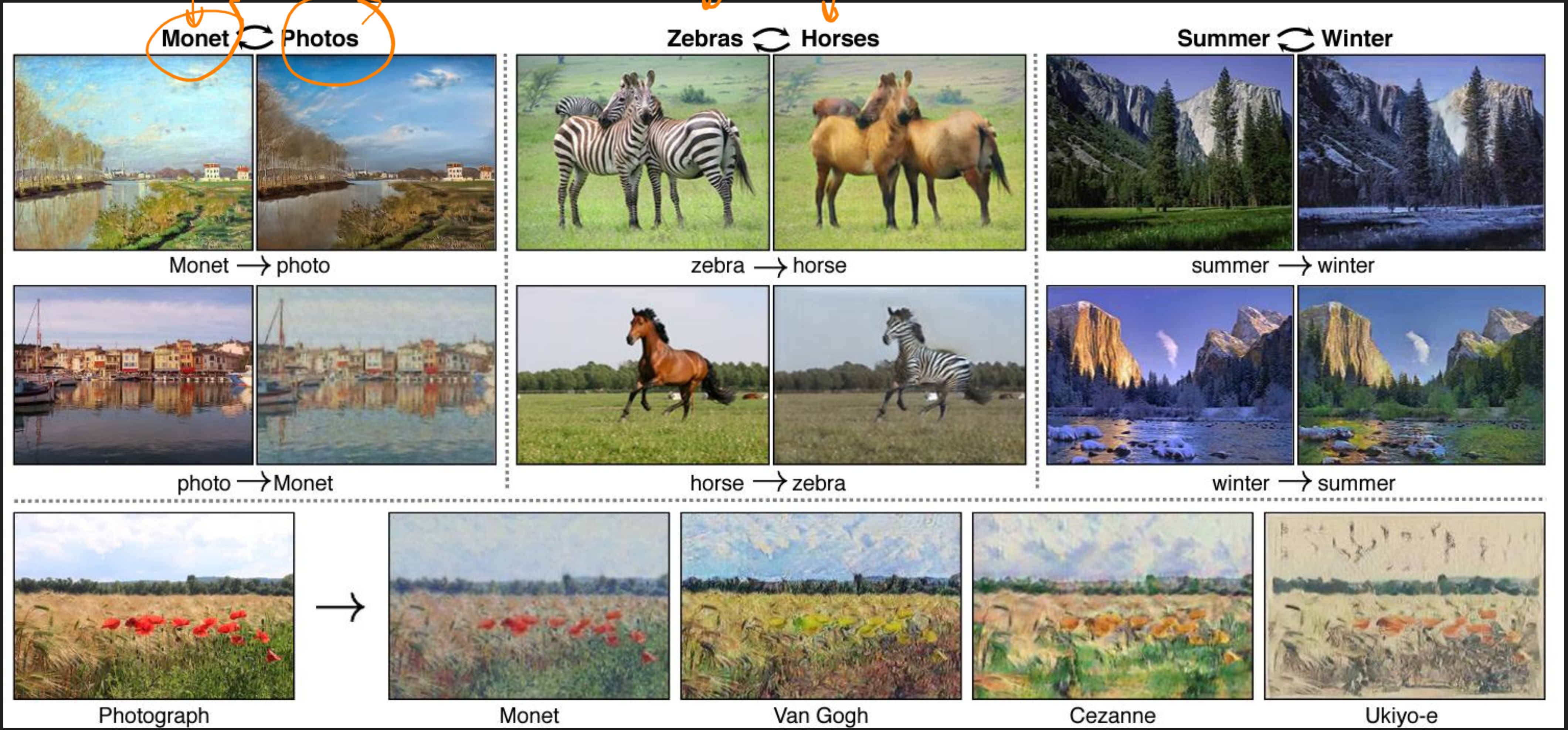
* Generator generates good images but background is blue.
* Disc: any image with blue background is fake.
* gen: background red
* Disc: any red background is fake

- Some tricks for training GANs
  - Balance the power of disc/gen
  - In the beginning the discriminator can easily win
    - Original GAN paper used an alternative objective function … to help the generator
  - Use of ADAM instead of SGD
    - Especially for the generator
  - https://github.com/soumith/ganhacks

SGD ALTERNATIVES

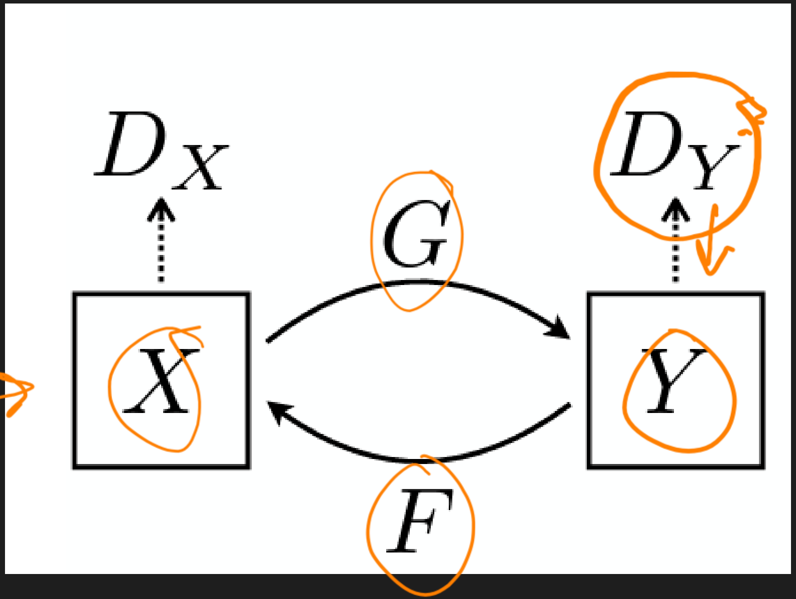
- $w^{t+1} = w^t - \alpha \nabla_W(\mathrm{E}(w))$
  - Performance is sensitive to the choice of α
  - Variable rate decay
- Alternatives to gradient descent
  - Using momentum
  - RMSProp
  - ADAM
  - …

STYLE TRANSFER VIA (CYCLE) GANS



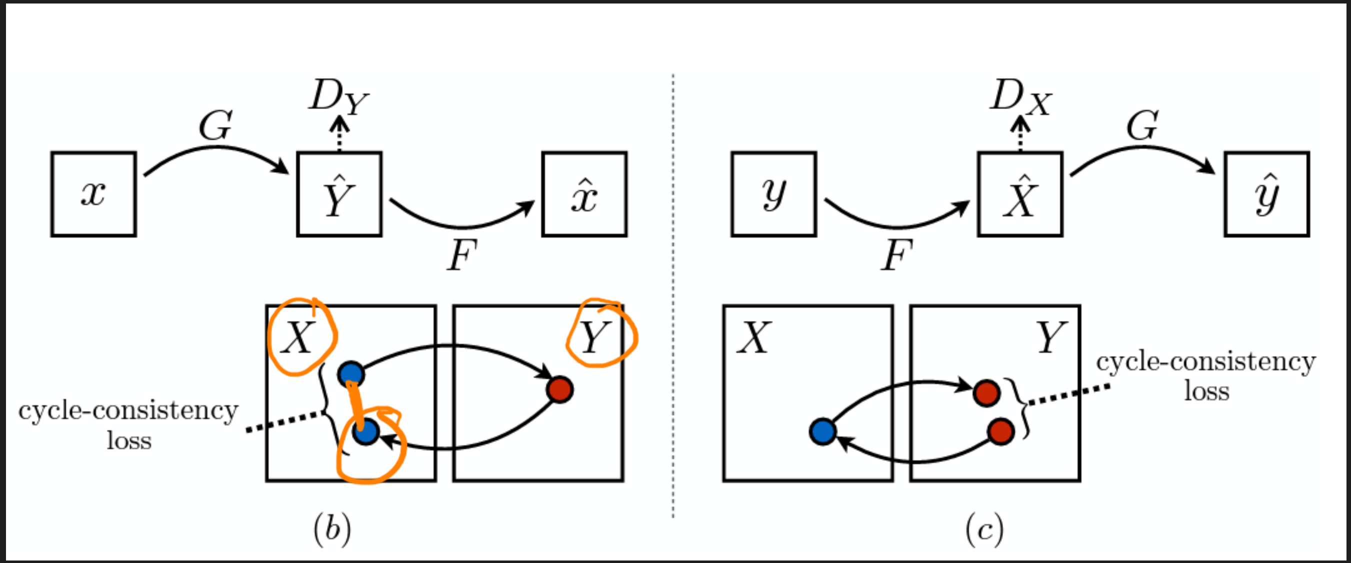
Cycle GAN

- Two data sets (domain X and Y)
  - Find a transformation that maps X to Y and vice versa
  - Unpaired data
  - No labels either
- GANs: map Gaussians to images…
  - Here: Map one domain to the other…
  - …and vice versa
  - Train a discriminator/gen for each domain



CYCLE-CONSISTENCY CONSTRAINT

- ADDITIONAL CONSTRAINT/LOSS:

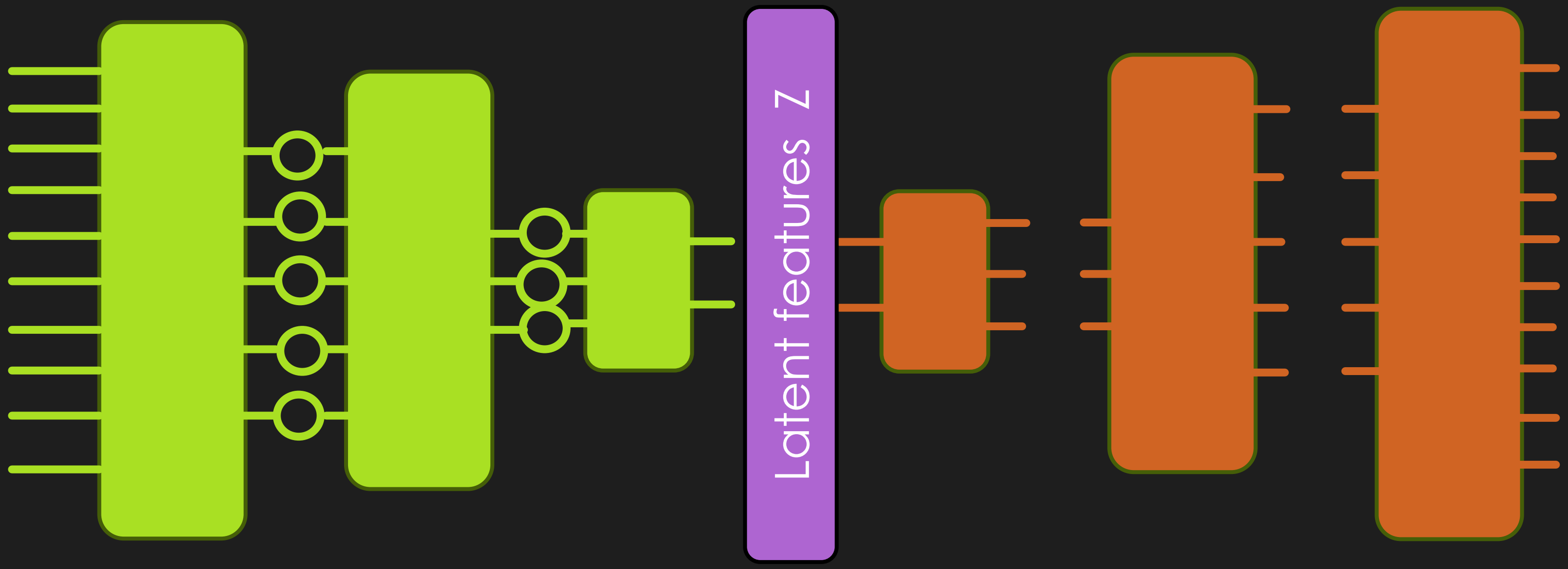


$||X - F(G(x))||_2^2$

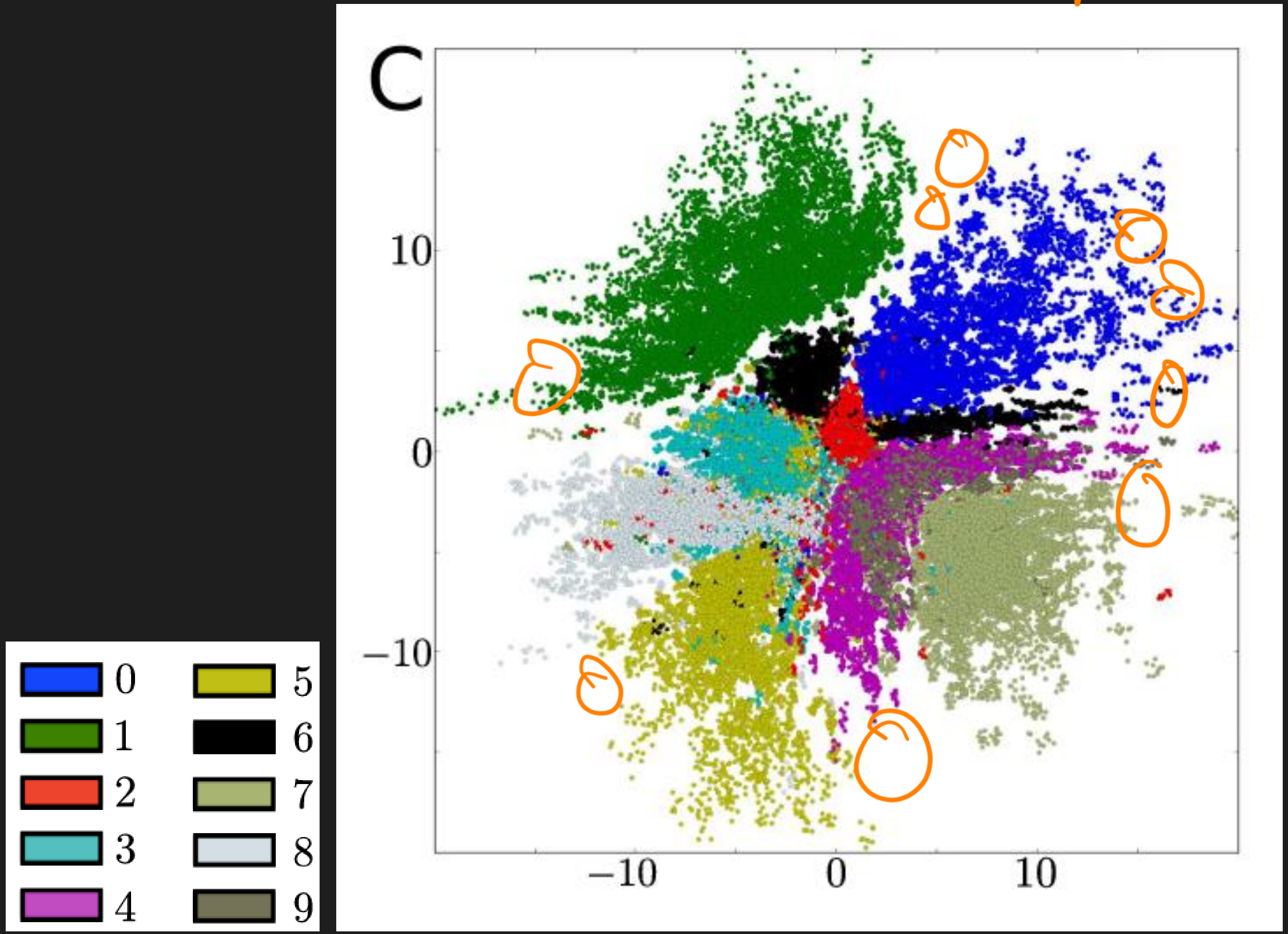
Translation example

ADVERSARIAL AUTOENCODER

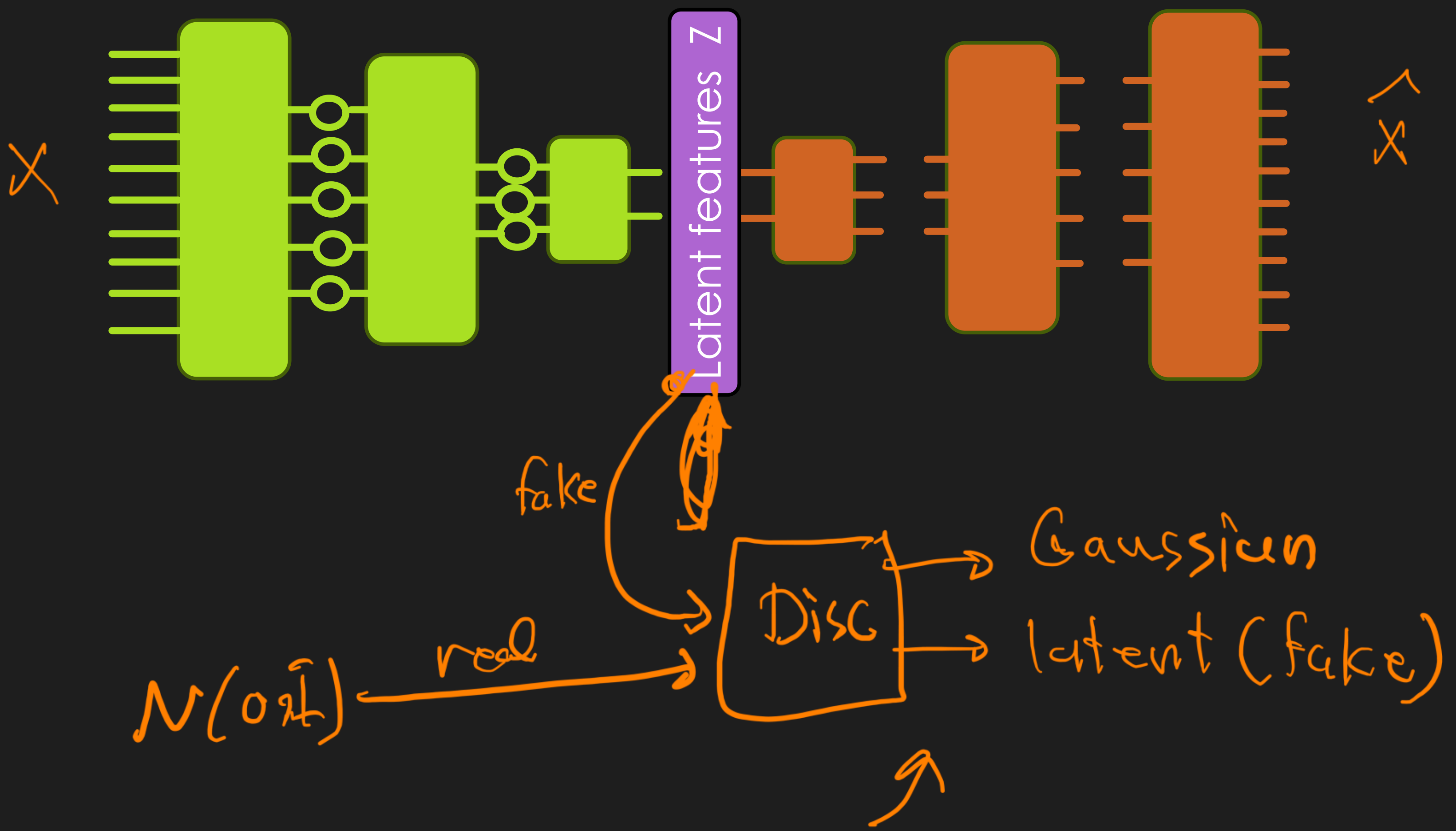
- How to enforce the latent space to be Gaussian?

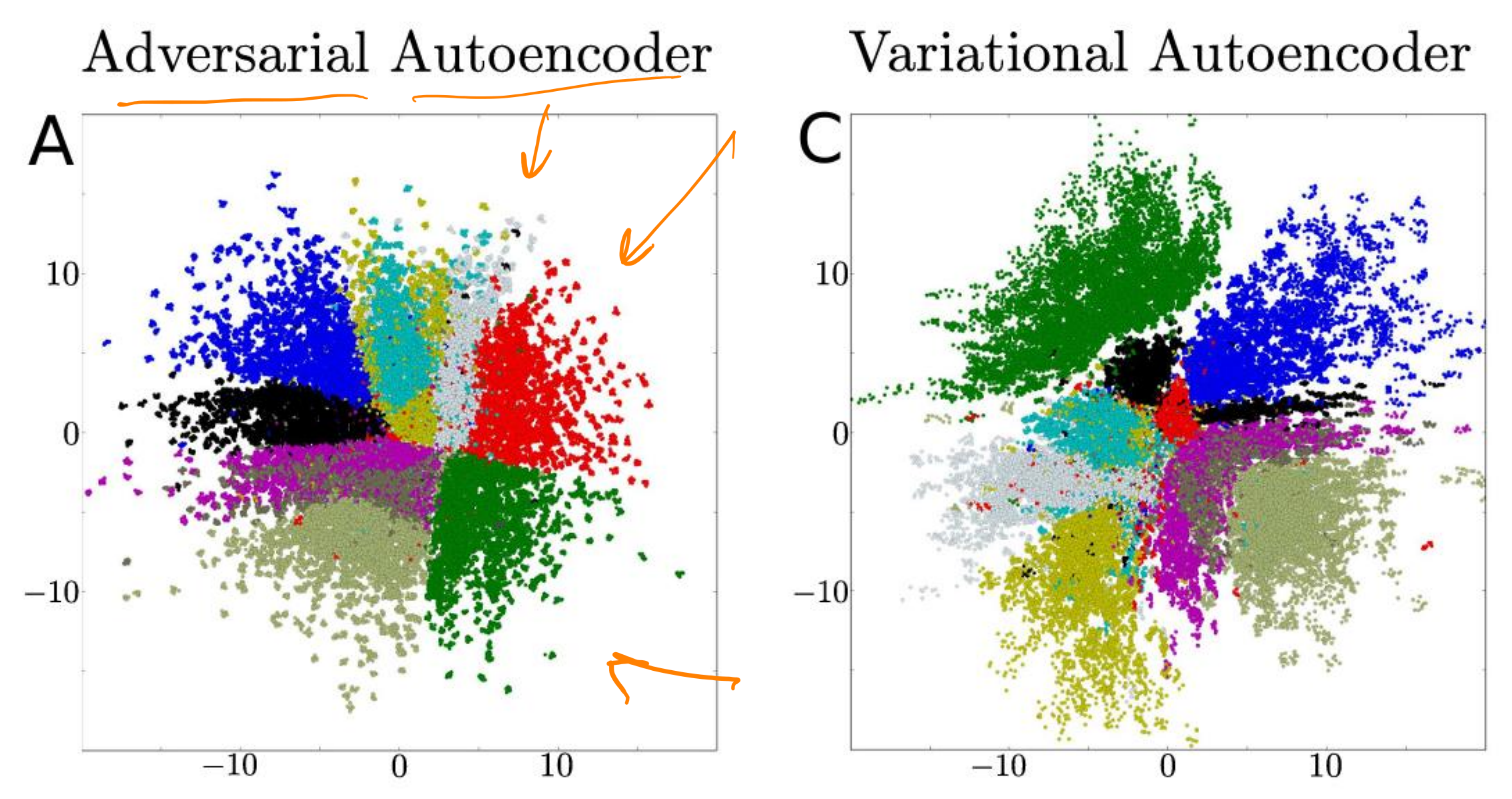


PITFALLS OF VAE



ADVERSARIAL AUTOENCODERS





Makhzani et al