



# Deep Generative Models

Aditya Grover

Stanford University

[aditya-grover.github.io](https://aditya-grover.github.io)

Deep Learning For Science School

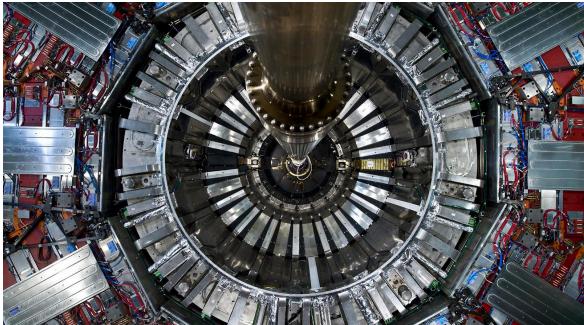
July 30, 2020



@adityagrover\_

# Age of Big **Unlabeled** Data

**Opportunity:** High-throughput data and computation



Particle Accelerators



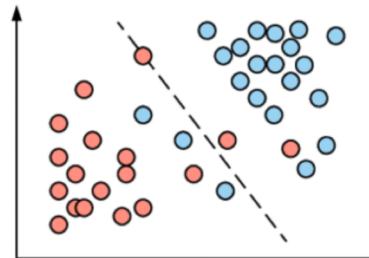
Materials Project



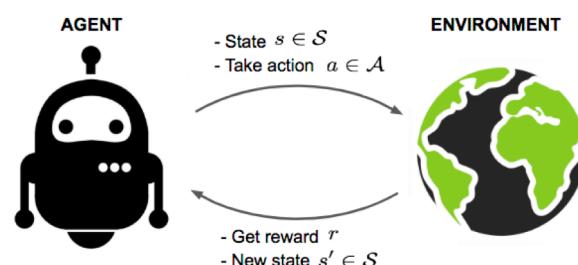
GenBank

**Challenge:** Supervision signals such as labels are *expensive*  
e.g., time, money, expertise, safety costs

# Learning With Limited Supervision

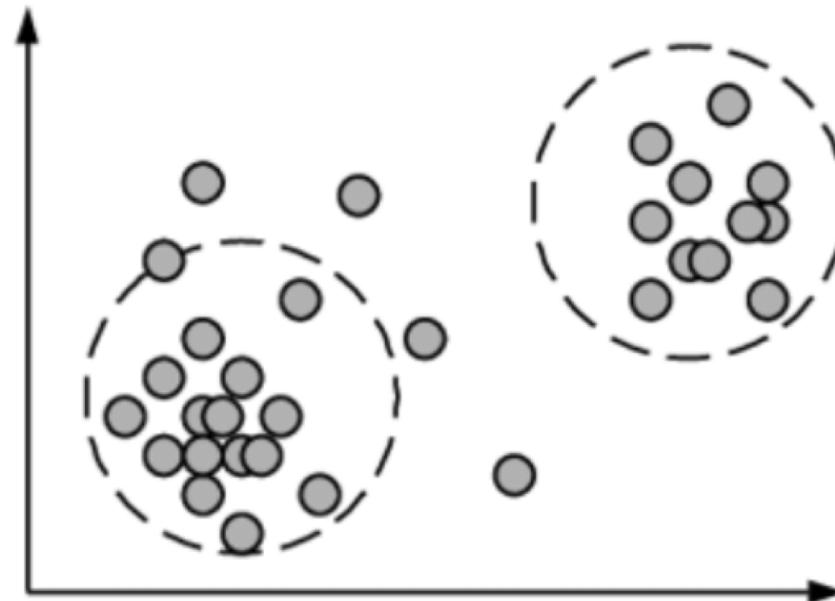


Classification  
w/ active querying



Reinforcement Learning  
w/ sparse rewards

Can we learn with *no* supervision?



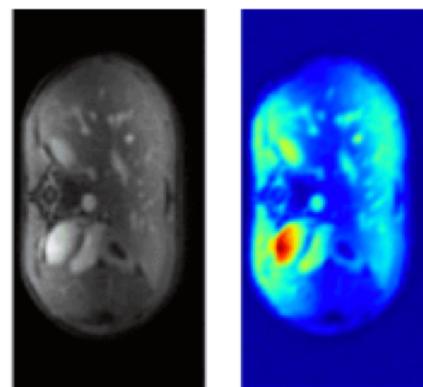
# Recent Advances in Unsupervised Learning



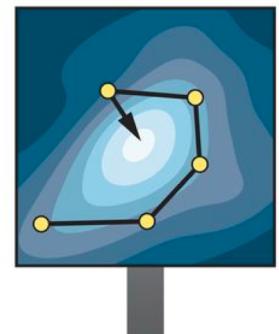
Karras et al.



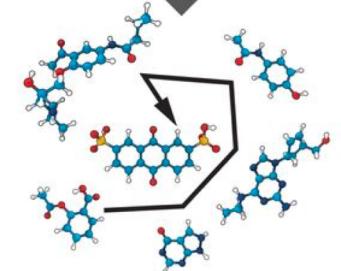
Huang et al.



Mardani et al.

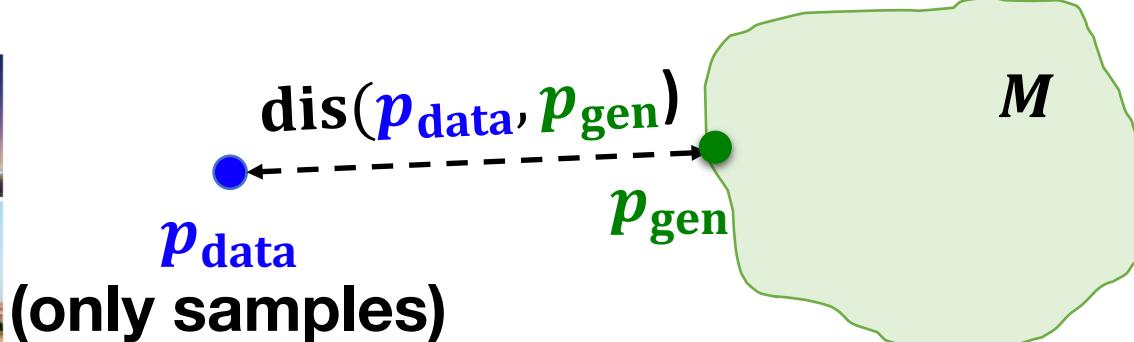


Optimization,  
evolutionary strategies,  
generative models (VAE,  
GAN, RL)



Sanchez-Lengelin et al.

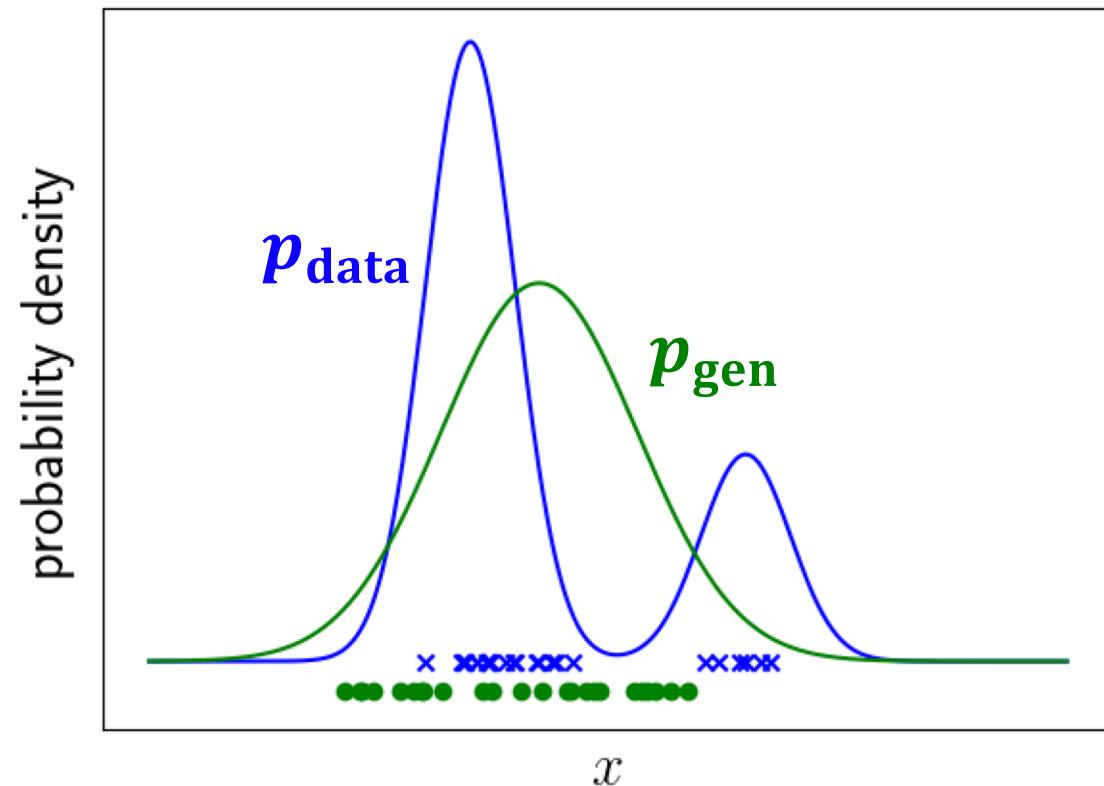
# Generative Models: Fit Data Distributions



$$\min_{p_{\text{gen}} \in M} \text{dis}(p_{\text{data}}, p_{\text{gen}})$$

# Example

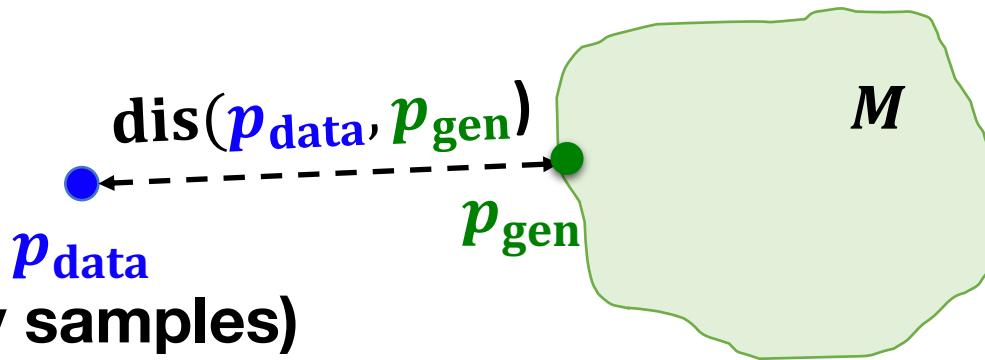
- $p_{\text{data}}$  data distribution  
(unknown)
- $\times$  data samples  
(known)
- $M$  mixture of 1d Gaussians
- $p_{\text{gen}}$  generative model
- generated samples



# Key Challenge: Curse of Dimensionality



(only samples)



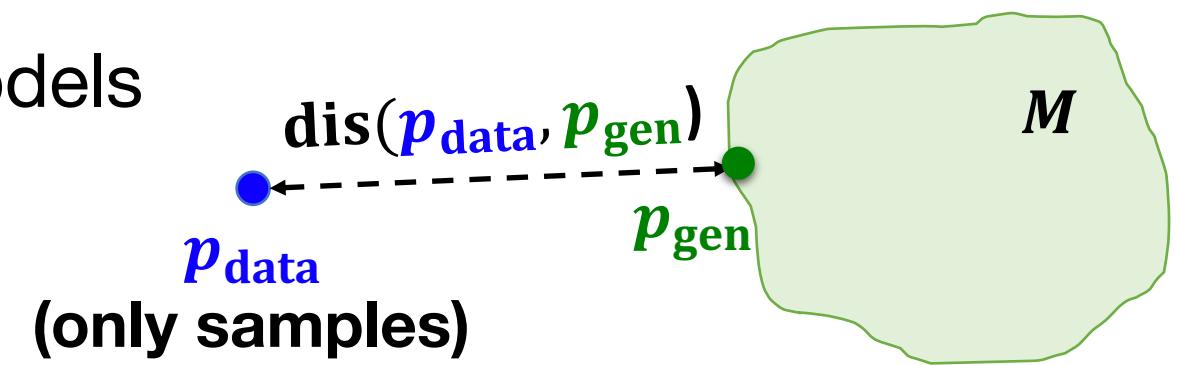
$$\min_{p_{\text{gen}} \in M} \text{dis}(p_{\text{data}}, p_{\text{gen}})$$

In high dimensions:

- (a)  $p_{\text{data}}$  can contain many modes
- (b) Training samples cover very small region of true support

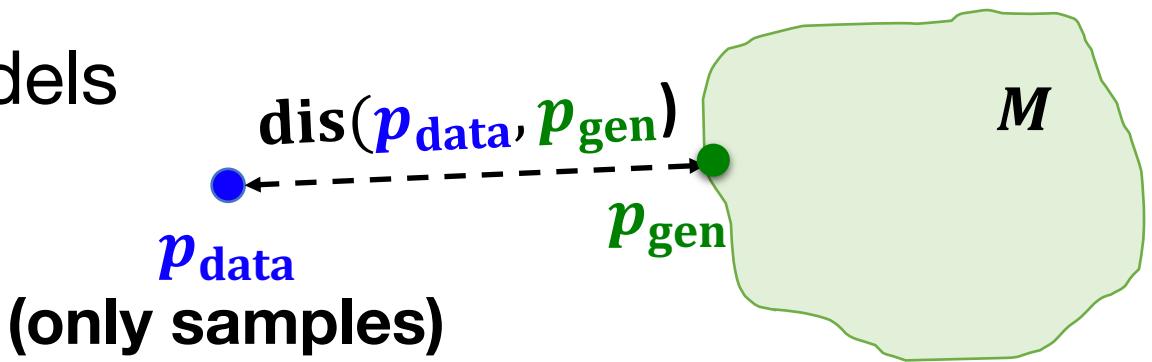
# Roadmap

- Introduction
- Likelihood-based Generative Models
  - Autoregressive Models
  - Variational Autoencoders
  - Normalizing Flow Models
- Likelihood-free Generative Models
  - Generative Adversarial Networks
- Applications In Scientific Discovery



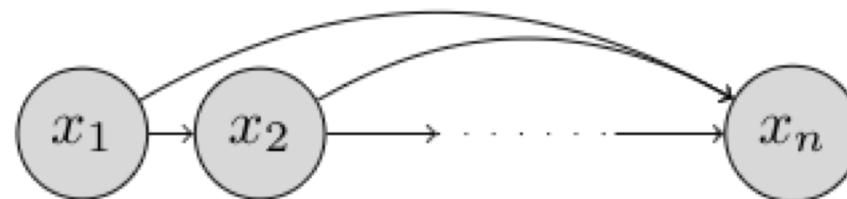
# Roadmap

- Introduction
- Likelihood-based Generative Models
  - Autoregressive Models
  - Variational Autoencoders
  - Normalizing Flow Models
- Likelihood-free Generative Models
  - Generative Adversarial Networks
- Applications In Scientific Discovery



# Autoregressive Generative Models

- Directed, fully-observed graphical models



- **Key idea:** Decompose the joint distribution as a **product of tractable conditionals**

$$p_{\theta}(\mathbf{x}) = \prod_{i=1}^n p_{\theta}(x_i | x_1, x_2, \dots, x_{i-1}) = \prod_{i=1}^n p_{\theta}(x_i | x_{<i})$$

# Learning and Inference

- **Learning** maximizes the model log-likelihood over the dataset  $\mathcal{D}$

$$\max_{\theta} \log p_{\theta}(\mathcal{D}) = \sum_{\mathbf{x} \in \mathcal{D}} \log p_{\theta}(\mathbf{x}) = \sum_{\mathbf{x} \in \mathcal{D}} \sum_{i=1}^n \log p_{\theta}(x_i | x_{<i})$$

- Tractable conditionals allow for **exact likelihood evaluation**
  - Conditional evaluated in parallel during training
- Directed model permits **ancestral sampling**, one variable at a time

$$x_1 \sim p_{\theta}(x_1), x_2 \sim p_{\theta}(x_2 | x_1), \dots, x_n \sim p_{\theta}(x_n | x_{<n})$$

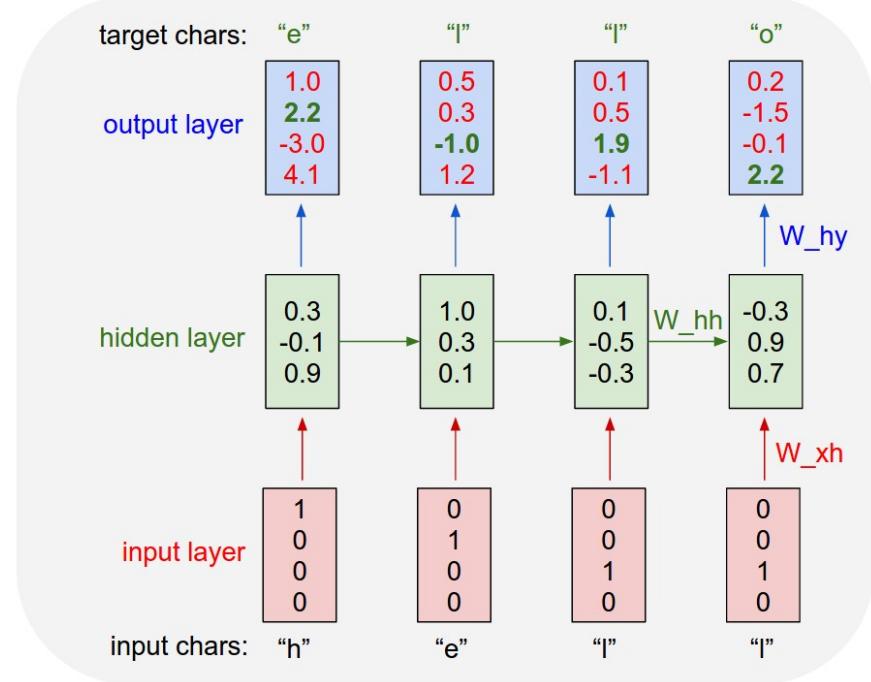
# Parameterizing AGMs

Linear/multi-layer NN based parameterizations

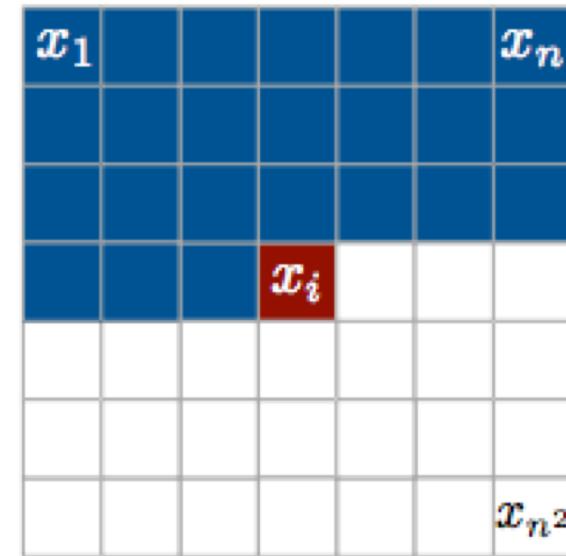
- Fully-visible sigmoid belief networks (FVSBN)
- Neural autoregressive density estimator (NADE): 1 hidden layer NN
- Masked autoencoder distribution estimation (MADE): multilayer NN

Neal, 1992, Larochelle & Murray, 2011, Germain et al., 2015

# RNN-based parameterizations



Char-RNN

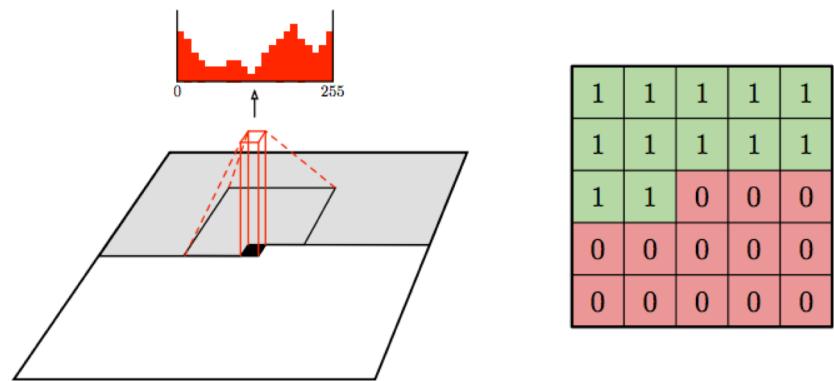


Pixel-RNN

Sutskevar et al., 2011, Karpathy, 2015, Theis & Bethge, 2015, van den Oord, 2016a

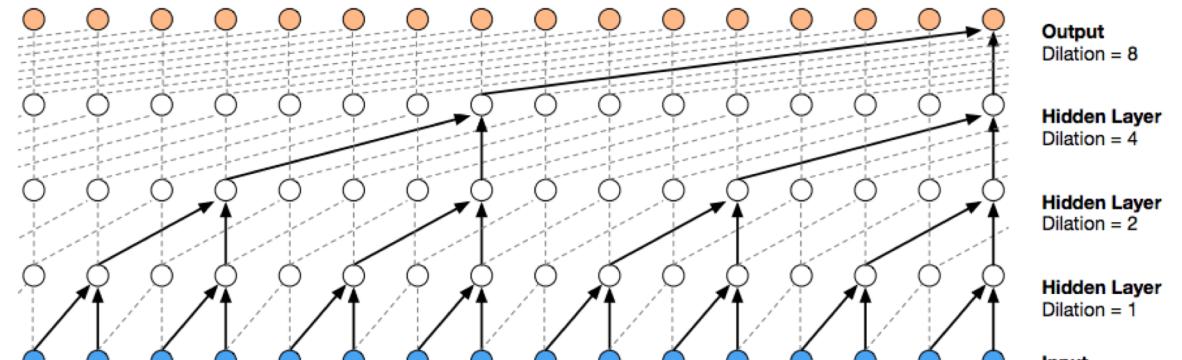
# CNN-based parameterizations

Masked convolutions preserve  
raster scan order



PixelCNN

Dilated convolutions increase  
the receptive field



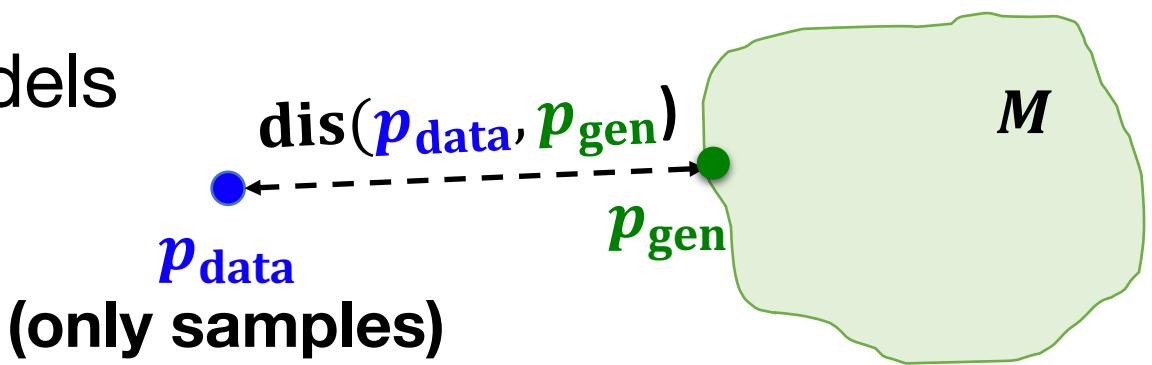
WaveNet

van den Oord et al., 2016b, c

Recent: Transformer-based parameterizations  
[Radford et al., 2018, 2019 Brown et al., 2020, Chen et al., 2020]

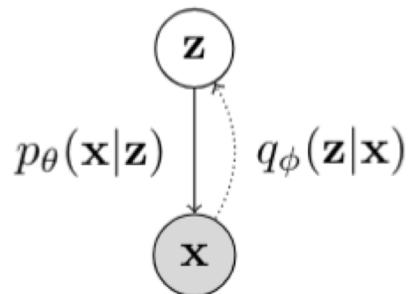
# Roadmap

- Introduction
- Likelihood-based Generative Models
  - Autoregressive Models
  - Variational Autoencoders
  - Normalizing Flow Models
- Likelihood-free Generative Models
  - Generative Adversarial Networks
- Applications In Scientific Discovery



# Variational Autoencoders

- Directed, latent-variable models, **with an inference network**



- **Goal:** Maximize the marginal log-likelihood of the dataset

$$\max_{\theta} \sum_{x \in \mathcal{D}} \log p_\theta(\mathbf{x})$$

# Variational Autoencoders

- **Challenge:** Marginal log-likelihood of any data point  $\mathbf{x}$  is intractable

$$\log p_\theta(\mathbf{x}) = \log \int p_\theta(\mathbf{x}, \mathbf{z}) d\mathbf{z}$$

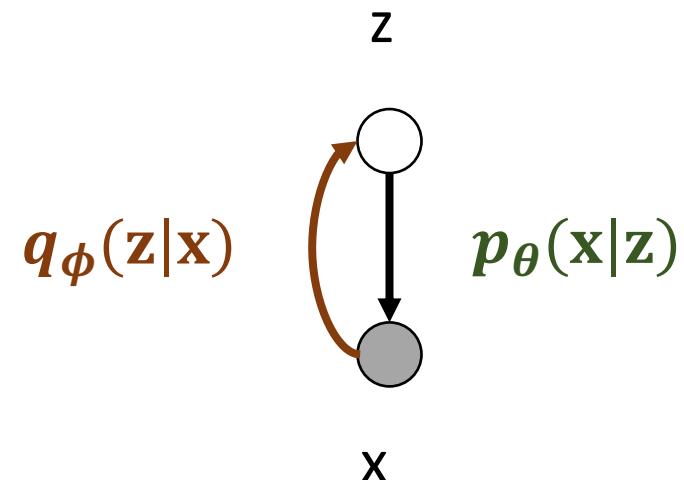
- **Key ideas**

- Approximate the posterior  $p_\theta(\mathbf{z}|\mathbf{x})$  with a simpler, tractable distribution  $q_\phi(\mathbf{z}|\mathbf{x})$
- Cast **inference as optimization** over  $\theta$  and  $\phi$

# Inference as Optimization

- Goal: Maximize the marginal log-likelihood of the data is **intractable!**

$$\log \int_{\mathbf{z}} p_{\theta}(\mathbf{x}, \mathbf{z}) = \log \int_{\mathbf{z}} q_{\phi}(\mathbf{z}|\mathbf{x}) \frac{p_{\theta}(\mathbf{x}, \mathbf{z})}{q_{\phi}(\mathbf{z}|\mathbf{x})}$$



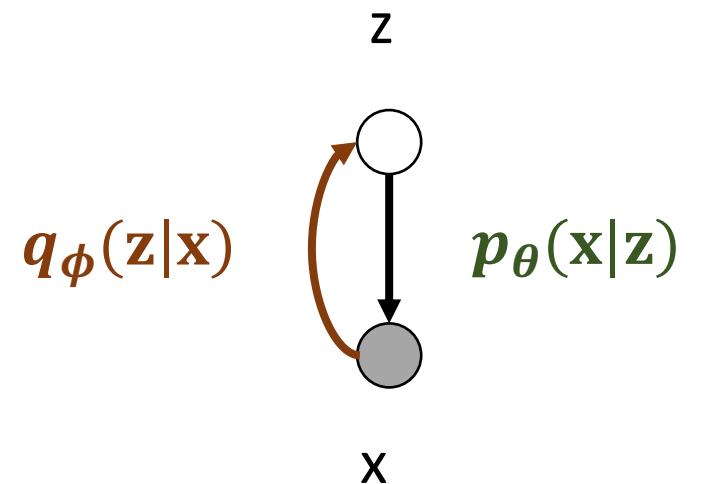
# Inference as Optimization

- **New** Goal: Maximize **lower bound** to the marginal log-likelihood of the data is **tractable!**

$$\begin{aligned}\log \int_{\mathbf{z}} p_{\theta}(\mathbf{x}, \mathbf{z}) &= \log \int_{\mathbf{z}} q_{\phi}(\mathbf{z}|\mathbf{x}) \frac{p_{\theta}(\mathbf{x}, \mathbf{z})}{q_{\phi}(\mathbf{z}|\mathbf{x})} \\ &\geq \underbrace{\int_{\mathbf{z}} q_{\phi}(\mathbf{z}|\mathbf{x}) \log \frac{p_{\theta}(\mathbf{x}, \mathbf{z})}{q_{\phi}(\mathbf{z}|\mathbf{x})}}_{\text{ELBO}(\mathbf{x}; \theta, \phi)}\end{aligned}$$

**Tightness Condition**

$$q_{\phi}(\mathbf{z}|\mathbf{x}) = p_{\theta}(\mathbf{z}|\mathbf{x})$$

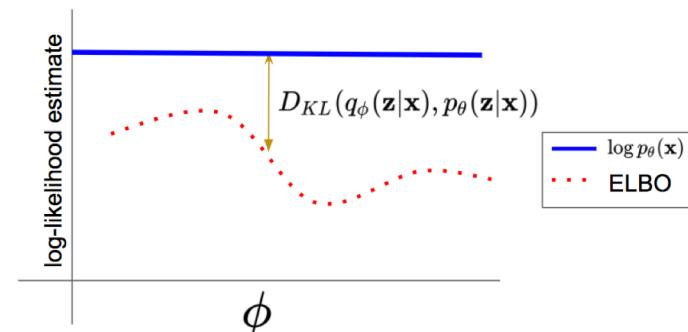


# Variational Bayes

- Evidence lower bound (ELBO) for the marginal log-likelihood of  $\mathbf{x}$

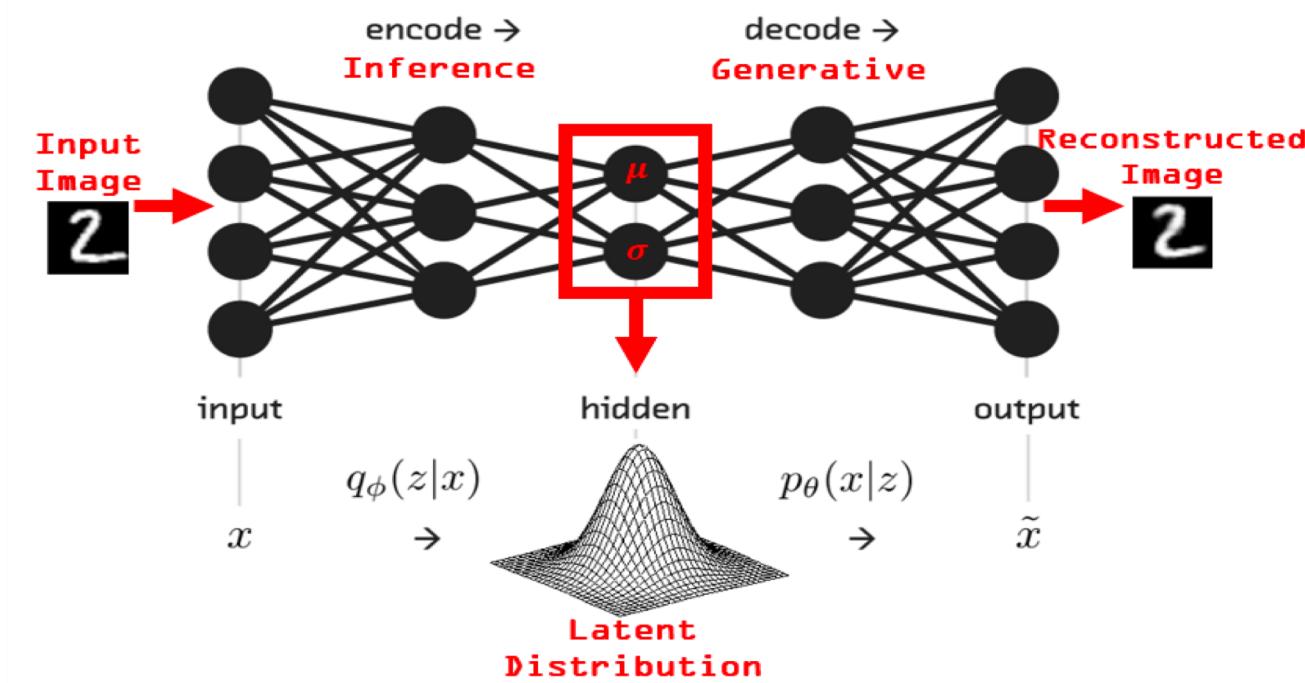
$$\log p_\theta(\mathbf{x}) = \underbrace{\mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})} \left[ \log \left( \frac{p_\theta(\mathbf{x}, \mathbf{z})}{q_\phi(\mathbf{z}|\mathbf{x})} \right) \right]}_{\text{ELBO}(\mathbf{x}; \theta, \phi)} + \text{KL}(q_\phi(\mathbf{z}|\mathbf{x}) \| p_\theta(\mathbf{z}|\mathbf{x}))$$

- KL gap depends on the quality of the variational approximation



# The Autoencoder Perspective

$$\text{ELBO}(\mathbf{x}; \theta, \phi) = \underbrace{\mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})} \left[ \log \left( \frac{p_\theta(\mathbf{x}, \mathbf{z})}{q_\phi(\mathbf{z}|\mathbf{x})} \right) \right]}_{\text{neg. reconstruction error}} - \underbrace{\text{KL}(q_\phi(\mathbf{z}|\mathbf{x}), p_\theta(\mathbf{z}))}_{\text{regularization penalty}}$$



# Learning and Inference

- Learning maximizes the ELBO for the observed data  $\mathcal{D}$
- ELBO gives **lower bounds** on true, intractable log-likelihood
- Directed model permits **ancestral sampling**

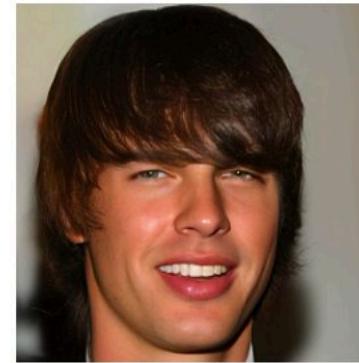
$$\mathbf{z} \sim p(\mathbf{z}), \mathbf{x} \sim p_{\theta}(\mathbf{x} | \mathbf{z})$$

- **Latent representations** for any  $\mathbf{x}$  can be inferred via  $q_{\phi}(\mathbf{z} | \mathbf{x})$

Neal, 1998

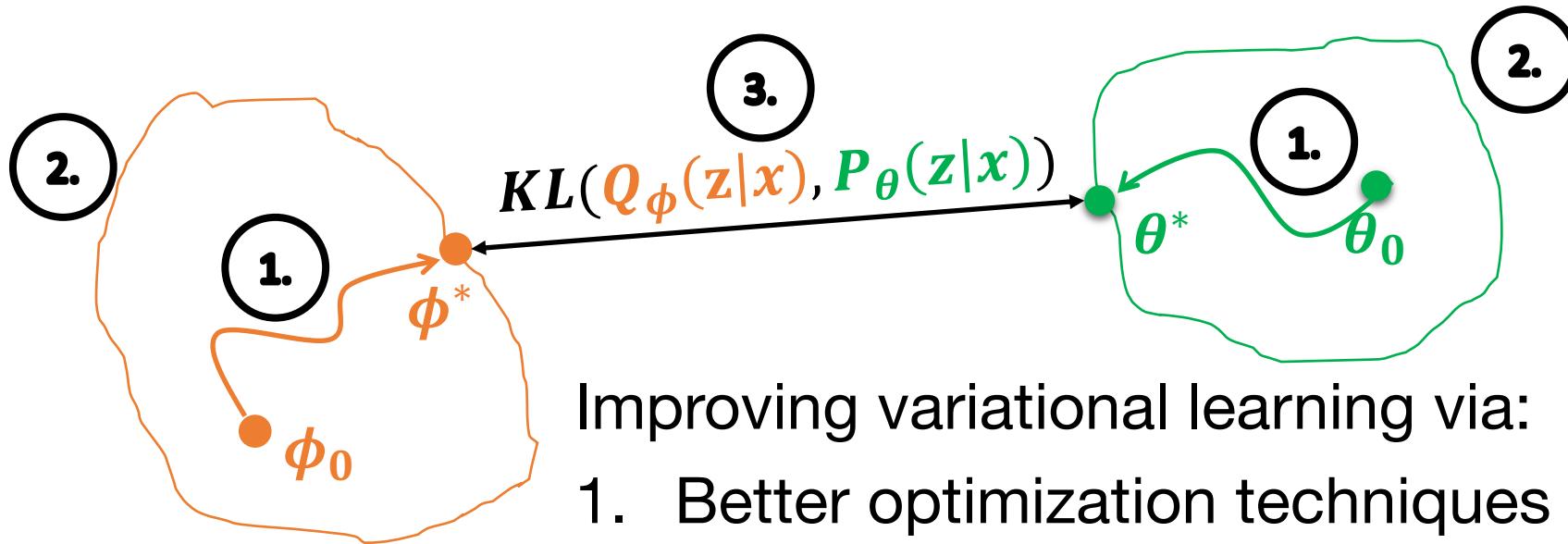
# NVAE

- Deep hierarchical VAEs with spectral regularization (and other tricks)



Vahdat & Kautz, 2020

# Research avenues in a nutshell

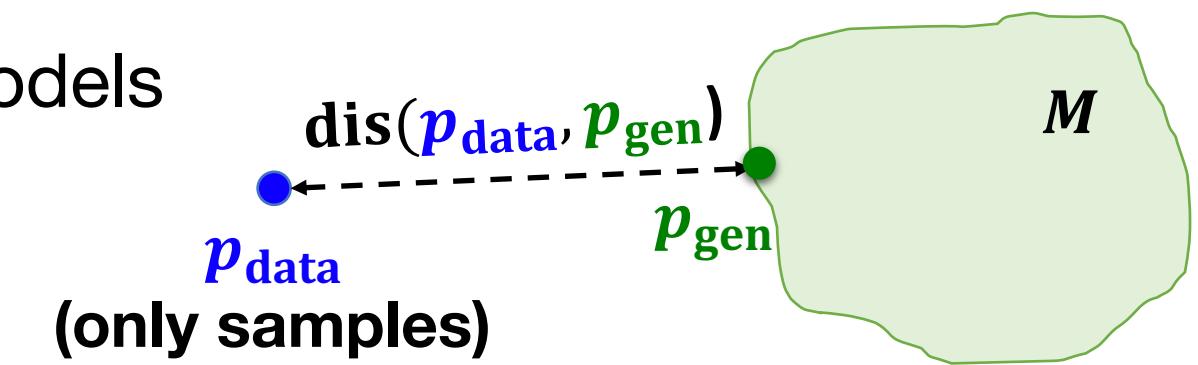


- Improving variational learning via:
1. Better optimization techniques
  2. More expressive approximating families
  3. Alternate loss functions

Figure inspired from David Blei's keynote at AISTATS 2018

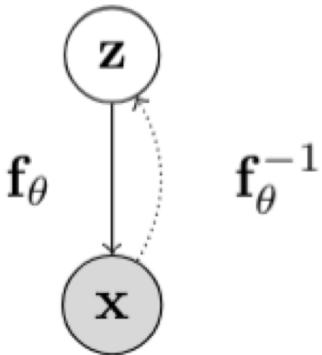
# Roadmap

- Introduction
- Likelihood-based Generative Models
  - Autoregressive Models
  - Variational Autoencoders
  - Normalizing Flow Models
- Likelihood-free Generative Models
  - Generative Adversarial Networks
- Applications In Scientific Discovery



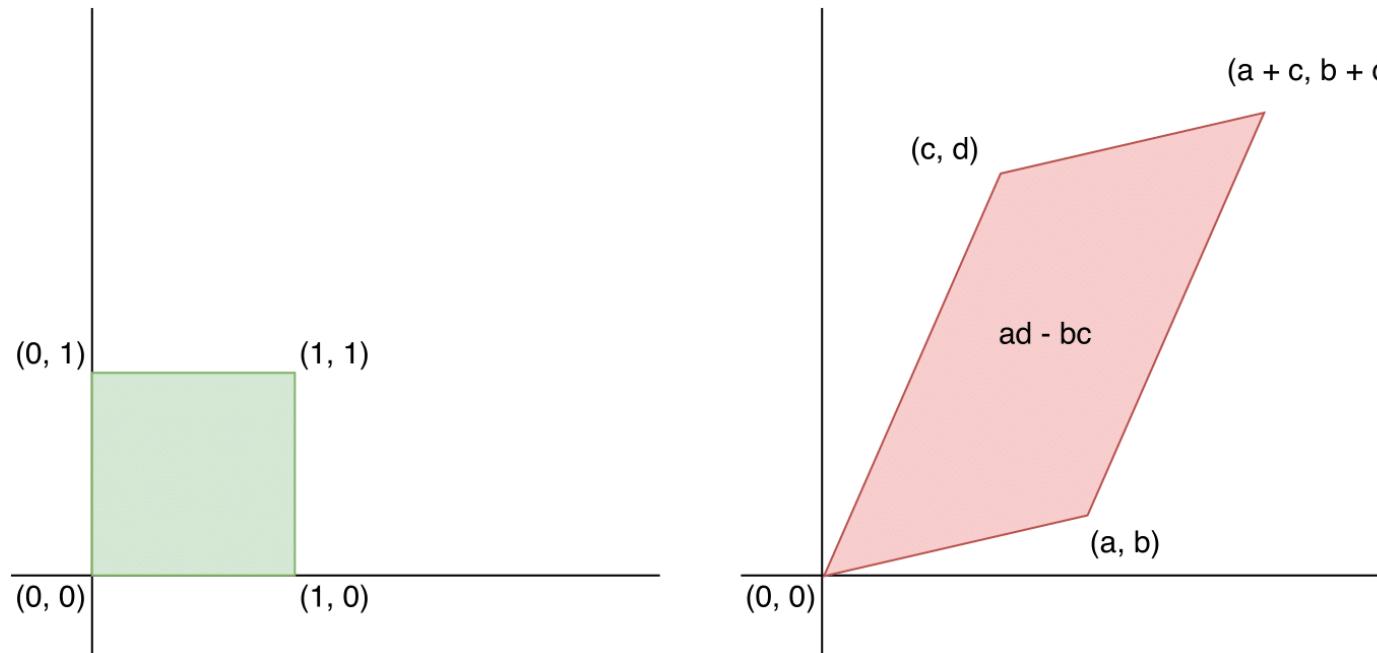
# Normalizing flow models

Directed, latent-variable **invertible** models



**Key idea:** The mapping between  $z$  and  $x$ , given by  $f_\theta: \mathbb{R}^n \rightarrow \mathbb{R}^n$ , is **deterministic and invertible** such that  $x = f_\theta(z)$  and  $z = f_\theta^{-1}(x)$ .

# Example: Linear Transformation



# Normalizing flow models

- Use **change-of-variables** to relate densities in the spaces defined by  $\mathbf{z}$  and  $\mathbf{x}$ :

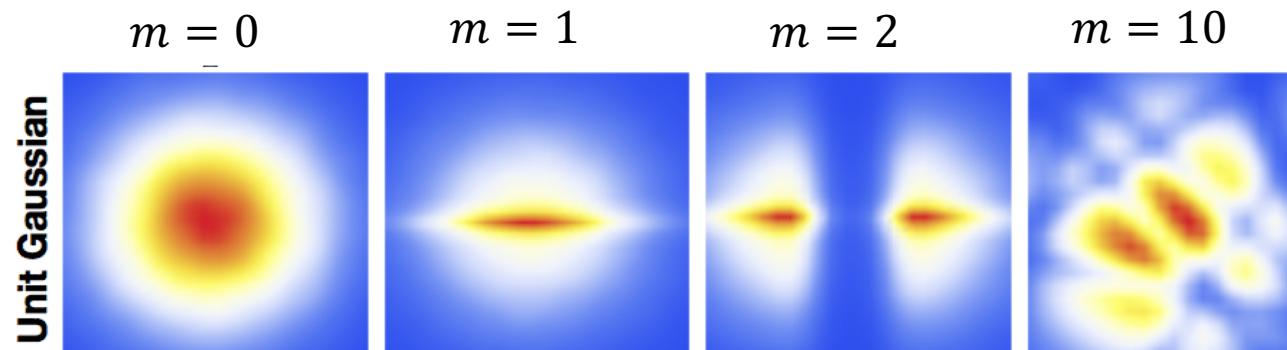
$$p_X(\mathbf{x}; \theta) = p_Z(\mathbf{z}) \left| \det \frac{\partial(\mathbf{f}_\theta)^{-1}}{\partial X} \right|_{X=\mathbf{x}}$$

- Transformations shape densities via **volume expansions and contractions**
- Jacobian determinant quantifies the per-unit change in volume

# A flow of transformations

Invertible transformations can be composed with each other

$$\mathbf{z}^M \triangleq \mathbf{x} = \mathbf{f}_\theta^M \circ \cdots \circ \mathbf{f}_\theta^1(\mathbf{z}^0); \quad p_X(\mathbf{x}; \theta) = p_{Z^0}(\mathbf{z}^0) \prod_{m=1}^M \left| \det \frac{\partial (\mathbf{f}_\theta^m)^{-1}}{\partial Z^m} \right|_{Z^m=\mathbf{z}^m}$$



Rezende & Mohamed, 2016

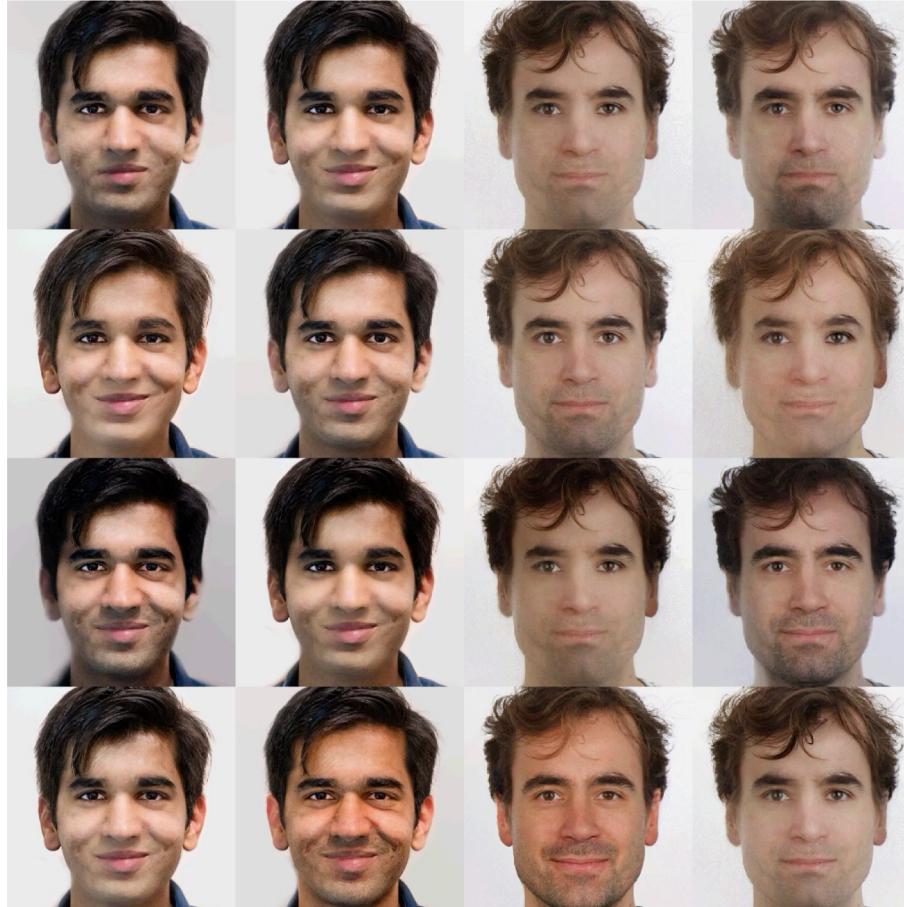
# Learning and inference

- **Learning** maximizes the model likelihood over the dataset  $\mathcal{D}$

$$\max_{\theta} \log p_X(\mathcal{D}; \theta) = \sum_{\mathbf{x} \in \mathcal{D}} \left( \log p_Z(\mathbf{z}) - \log \left| \det \frac{\partial (\mathbf{f}_{\theta})^{-1}}{\partial \mathbf{x}} \right|_{\mathbf{x}=\mathbf{x}} \right)$$

- **Exact likelihood evaluation** via inverse transformations
- **Ancestral sampling** via forward transformations
$$\mathbf{z} \sim p_Z(\mathbf{z}), \quad \mathbf{x} = \mathbf{f}_{\theta}(\mathbf{z})$$
- **Latent representations** inferred via inverse transformations
$$\mathbf{z} = (\mathbf{f}_{\theta})^{-1}(\mathbf{x})$$

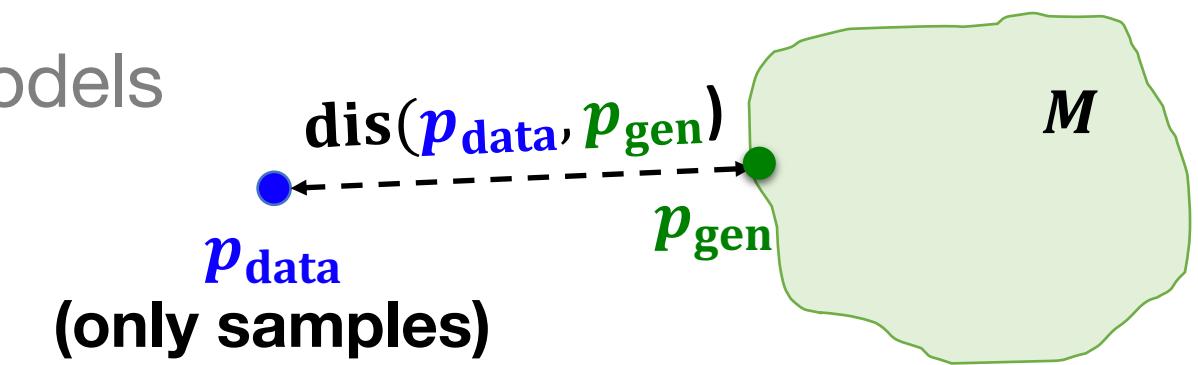
# Glow



Kingma & Dhariwal, 2018

# Roadmap

- Introduction
- Likelihood-based Generative Models
  - Autoregressive Models
  - Variational Autoencoders
  - Normalizing Flow Models
- Likelihood-free Generative Models
  - Generative Adversarial Networks
- Applications In Scientific Discovery



# Two Sample Testing

- Compare samples from  $p_{\text{data}}$  and  $p_{\text{gen}}$  over set of real-valued functions  $F$

$$\min_{p_{\text{gen}} \in M} \max_{f \in F} \mathbb{E}_{x \sim p_{\text{data}}} [f(x)] - \mathbb{E}_{x \sim p_{\text{gen}}} [f(x)]$$

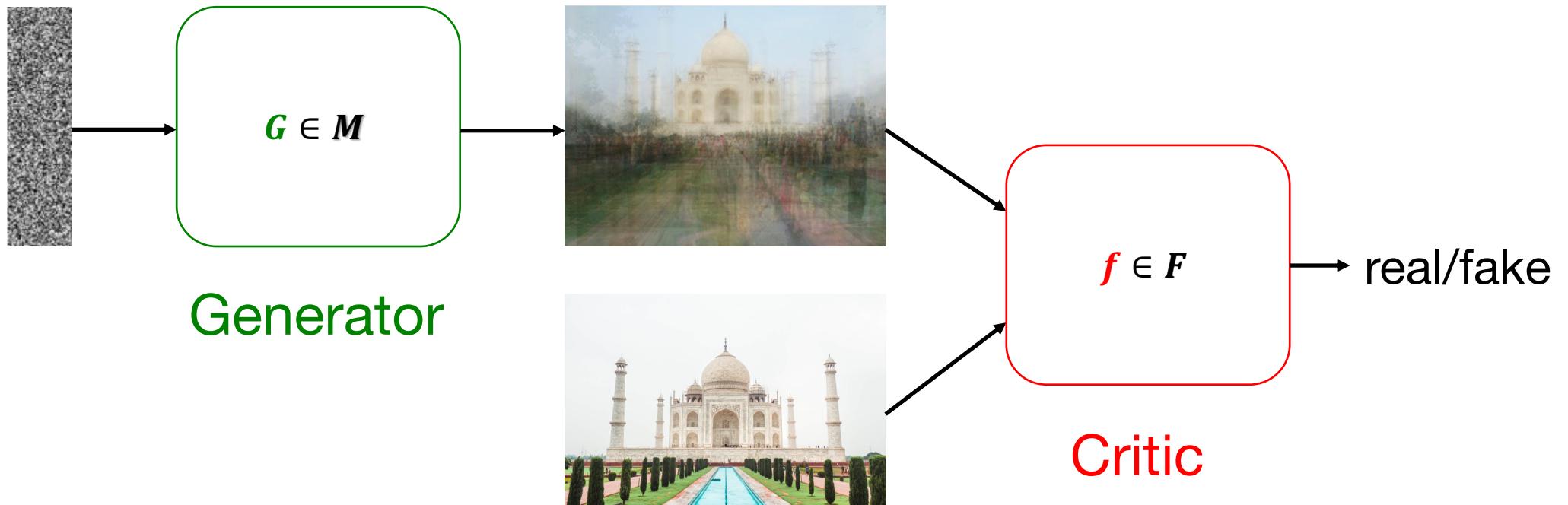
$\text{dis}(p_{\text{data}}, p_{\text{gen}})$

- Different choices of  $F$  lead to different discrepancy metrics, e.g., total-variation, Wasserstein, maximum mean discrepancy

# Generative Adversarial Nets (GAN)

$z \sim p(z)$  (e.g., Gaussian)

$x = \mathbf{G}(z)$

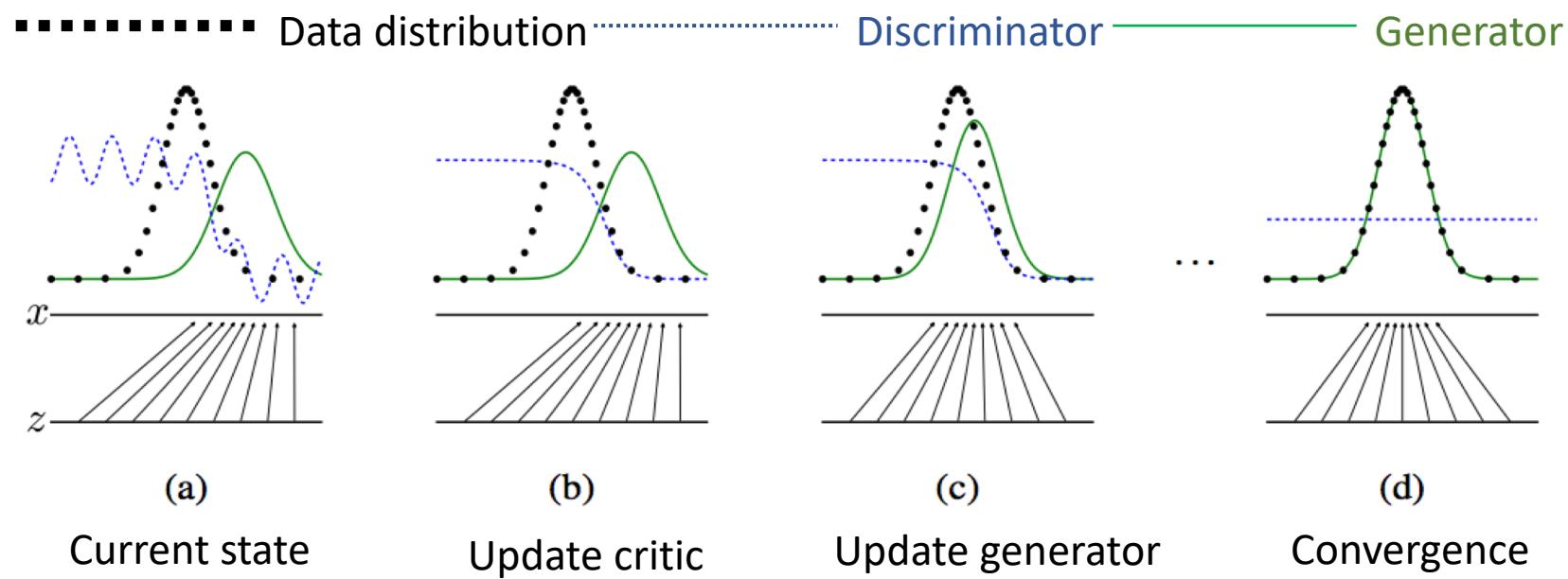


$$\min_{\mathbf{G} \in \mathcal{M}} \max_{\mathbf{f} \in \mathcal{F}} \mathbb{E}_{x \sim \mathbf{p}_{\text{data}}} [\mathbf{f}(x)] - \mathbb{E}_{z \sim p(z)} [\mathbf{f}(\mathbf{G}(z))]$$

# A minimax learning objective

During learning, generator and critic are updated alternatively

$$\min_{G \in M} \max_{f \in F} \mathbb{E}_{x \sim p_{\text{data}}} [f(x)] - \mathbb{E}_{z \sim p(z)} [f(G(z))]$$



Goodfellow et al., 2014

# Inference

- **Likelihoods** may not be defined or tractable
  - Except when the generative model is invertible [Grover et al., 2018]
- Directed model permits **ancestral sampling**
$$z \sim p(z) \text{ (e.g., Gaussian)}$$
$$x = \mathbf{G}(z)$$

# GANs Over the Years

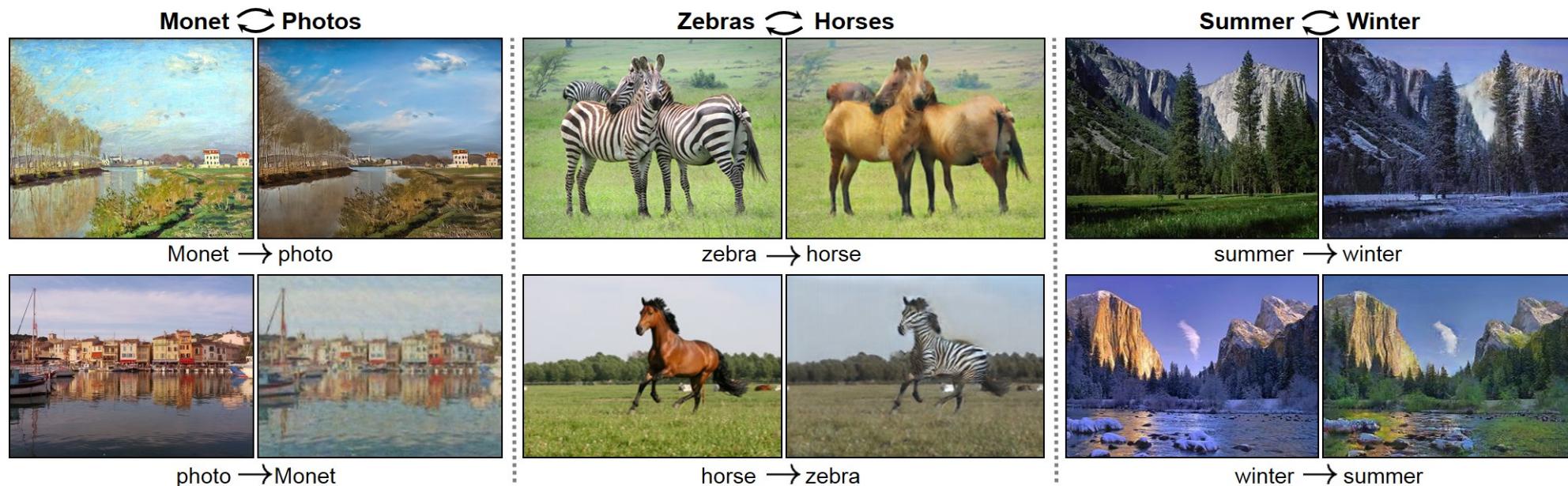


Pic credit: Ian Goodfellow

None of these individuals are real!  
Have these models passed the ‘generation Turing test’?

# Domain Translation & Adptation

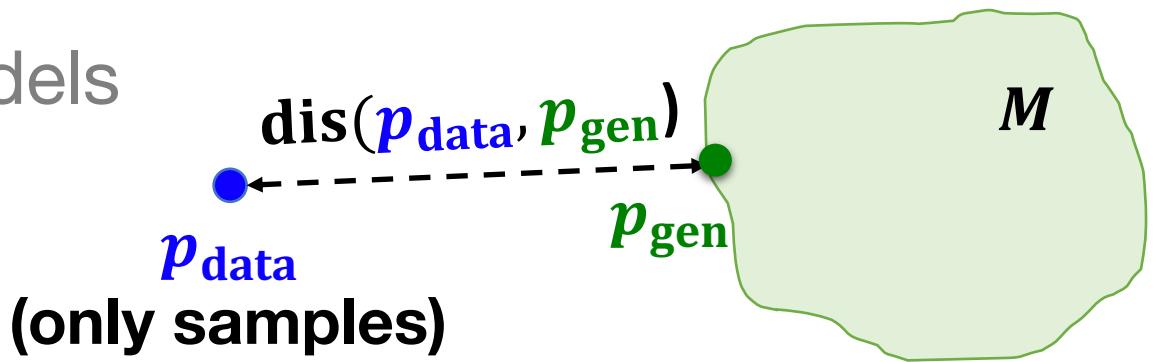
- **Conditional GANs** for image translation



Pix2Pix (Isola et al., 2017), CycleGAN (Zhu et al., 2017)

# Roadmap

- Introduction
- Likelihood-based Generative Models
  - Autoregressive Models
  - Variational Autoencoders
  - Normalizing Flow Models
- Likelihood-free Generative Models
  - Generative Adversarial Networks
- Applications In Scientific Discovery



# Many, many applications

---

**Physics-aware Deep Generative Models for  
Creating Synthetic Microstructures**

---

**Automatic Chemical Design Using a Data-Driven Continuous  
Representation of Molecules**

**Deep Generative Adversarial Neural Networks  
for Compressive Sensing MRI**

**Generative deep fields: arbitrarily sized, random synthetic  
astronomical images through deep learning**

Michael J. Smith<sup>★</sup> & James E. Geach<sup>†</sup>

*Centre for Astrophysics Research, School of Physics, Astronomy & Mathematics, University of Hertfordshire, Hatfield, AL10 9AB*  
*Centre of Data Innovation Research, School of Physics, Astronomy & Mathematics, University of Hertfordshire, Hatfield, AL10 9AB*

# What's Next for Deep Generative Models?

- Emerging classes of models
  - Energy-based models [Du et al., 2018], generalized generator-discriminator energy models [Grover et al. 2019, Arbel et al., 2020], score matching [Song et al., 2019], diffusion models [Ho et al., 2020]
- Evaluation metrics for sampling and representation learning
- Bias & fairness in generative modeling [Choi et al., 2020]
- Applications
  - Causal discovery and inference
  - Model-based reinforcement learning
  - Scientific discovery

# Check out [deepgenerativemodels.github.io](https://deepgenerativemodels.github.io) for materials!

Deep Generative Models  
CS236 - Fall 2019

---

**Course Description**  
Generative models are widely used in many subfields of AI and Machine Learning. Recent advances in parameterizing these models using deep neural networks, combined with progress in stochastic optimization methods, have enabled scalable modeling of complex, high-dimensional data including images, text, and speech. In this course, we will study the probabilistic foundations and learning algorithms for deep generative models, including variational autoencoders, generative adversarial networks, autoregressive models, and normalizing flow models. The course will also discuss application areas that have benefitted from deep generative models, including computer vision, speech and natural language processing, graph mining, and reinforcement learning.

---

[Course Notes](#)   [Syllabus](#)   [Piazza](#)   [Office Hours](#)

---

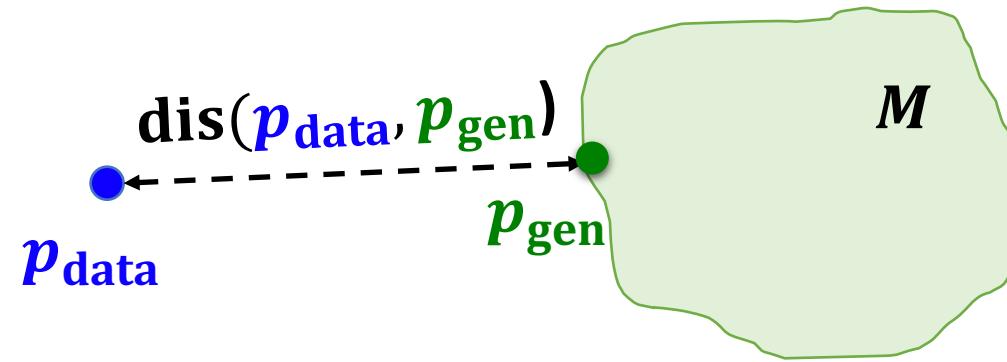
**Course Instructors**

						
<a href="#">Stefano Ermon</a>	<a href="#">Aditya Grover</a>	<a href="#">Kristy Choi</a>	<a href="#">Yang Song</a>	<a href="#">Rui Shu</a>	<a href="#">Amaury Sabran</a>	<a href="#">Kaidi Cao</a>

**Course Assistants**

				
<a href="#">Preerna Dhareshwar</a>	<a href="#">Sriram Somasundaram</a>	<a href="#">Arnaud Autef</a>	<a href="#">Xingyu Liu</a>	<a href="#">Kevin Zakka</a>

# Thank You



*“Imagination is the living power and prime agent of all human perception”*  
- Samuel Taylor Coleridge