Adaptive Density Estimation for Generative Models





Thomas Lucas

Konstantin Shmelkov*

Karteek Alahari

Cordelia Schmid

Jakob Verbeek

* Now at Huawei

Generative modelling

Goal

Given samples from target distribution p^*, train a model p_θ to match p^*

Generative modelling

Goal

Given samples from target distribution p^*, train a model p_θ to match p^*

- Maximum likelihood: Eval. training points under the model

Generative modelling

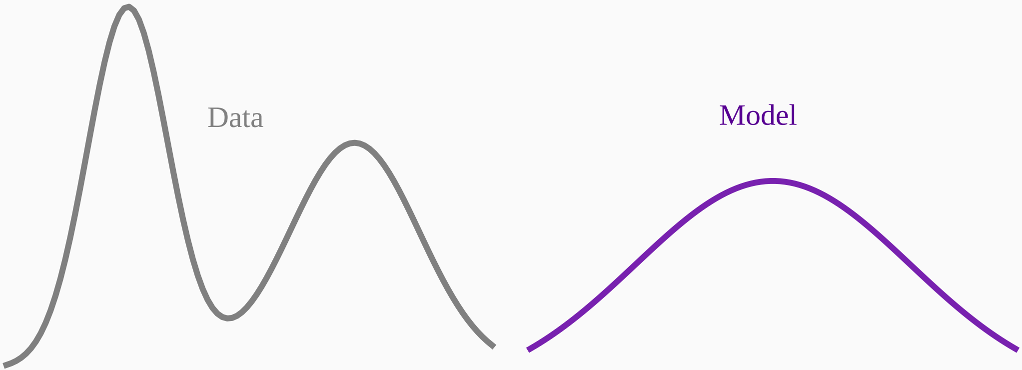
Goal

Given samples from target distribution p^*, train a model p_θ to match p^*

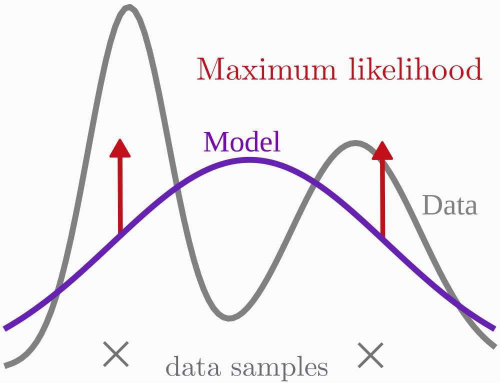
- Maximum likelihood: Eval. training points under the model
- Adversarial training[1]: Eval. samples under (approximation of) p^*

[1] Ian Goodfellow et al. (2014). "Generative adversarial nets". In: **NIPS**.

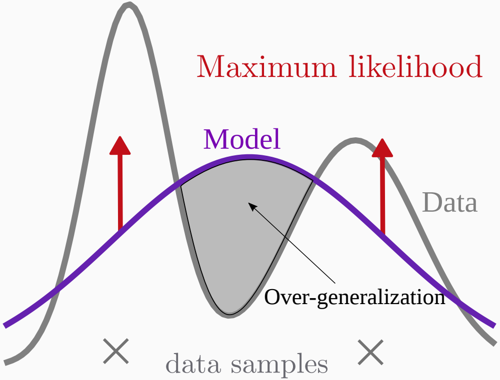
Schematic illustration



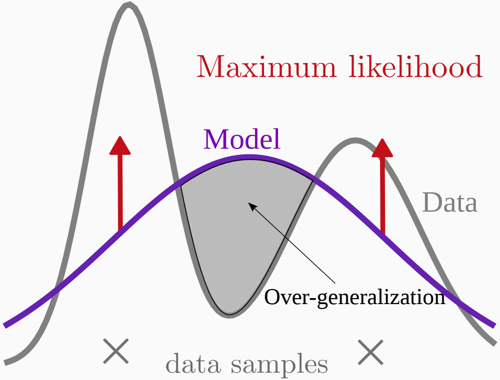
Maximum likelihood



Maximum likelihood



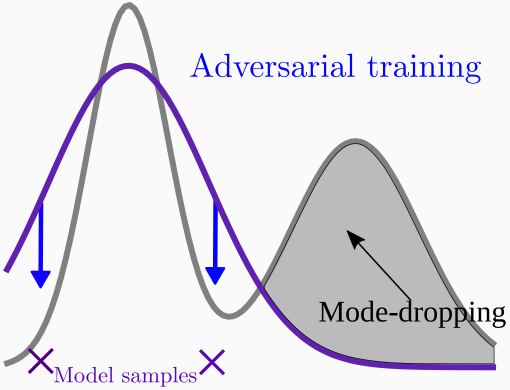
Maximum likelihood



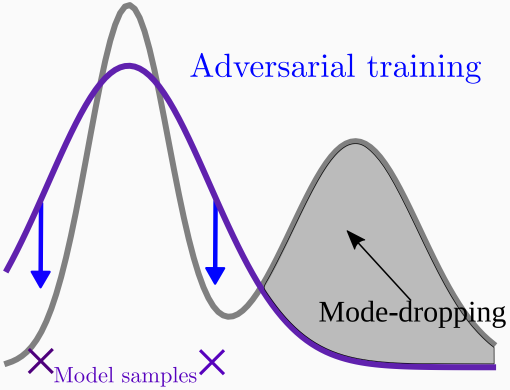
Consequences

- MLE covers **full support** of distribution
- Produces **unrealistic samples**

Adversarial training



Adversarial training



Consequences

- Production of **high quality** samples
- Parts of the support are **dropped**

Hybrid training approach

Goal

- Explicitly optimize both dataset **coverage** and sample **quality**

Hybrid training approach

Goal

- Explicitly optimize both dataset **coverage** and sample **quality**
- Discriminator can be seen as a **learnable** inductive bias

Hybrid training approach

Goal

- Explicitly optimize both dataset **coverage** and sample **quality**
- Discriminator can be seen as a **learnable** inductive bias
- Retain valid **likelihood** to evaluate support coverage

Hybrid training approach

Goal

- Explicitly optimize both dataset **coverage** and sample **quality**
- Discriminator can be seen as a **learnable** inductive bias
- Retain valid **likelihood** to evaluate support coverage

Challenges

- **Tradeoff** between the two objectives: need more flexibility

Hybrid training approach

Goal

- Explicitly optimize both dataset **coverage** and sample **quality**
- Discriminator can be seen as a **learnable** inductive bias
- Retain valid **likelihood** to evaluate support coverage

Challenges

- **Tradeoff** between the two objectives: need more flexibility
- Limiting **parametric assumptions** required for tractable MLE, e.g. Gaussianity, conditional independence

Hybrid training approach

Goal

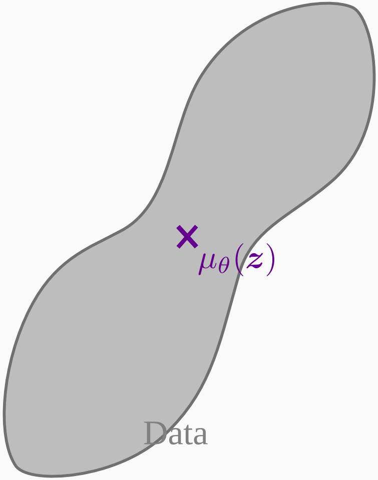
- Explicitly optimize both dataset **coverage** and sample **quality**
- Discriminator can be seen as a **learnable** inductive bias
- Retain valid **likelihood** to evaluate support coverage

Challenges

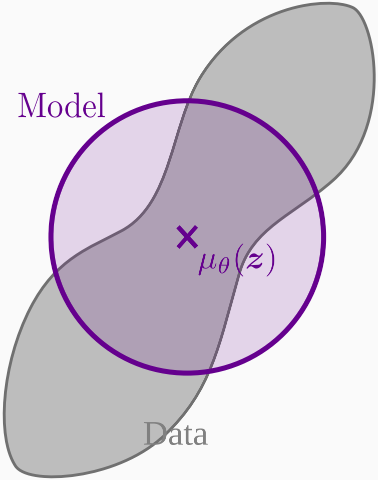
- **Tradeoff** between the two objectives: need more flexibility
- Limiting **parametric assumptions** required for tractable MLE, e.g. Gaussianity, conditional independence
- Often no likelihood in pixel space[2]

[2] A. Larsen et al. (2016). "Autoencoding beyond pixels using a learned similarity metric". In: **ICML**.

Conditional independence

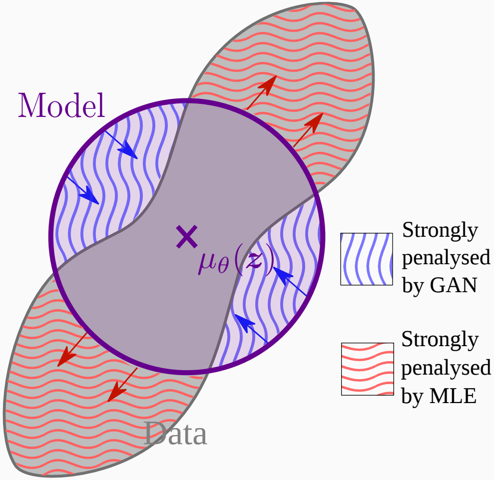


Conditional independence



$$p(\boldsymbol{x}|\boldsymbol{z}) = \prod_{i}^{N} \mathcal{N}(x_i|\mu_\theta(\boldsymbol{z}), \sigma \mathsf{I}_n)$$

Conditional independence



$$p(\boldsymbol{x}|\boldsymbol{z}) = \prod_i^N \mathcal{N}(x_i|\mu_\theta(\boldsymbol{z}), \sigma \mathsf{I}_n)$$

Going beyond conditional independence

Avoiding strong parametric assumptions

- Lift reconstruction losses into a **feature** space

Going beyond conditional independence

Avoiding strong parametric assumptions

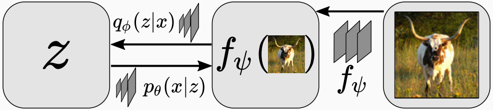
- Lift reconstruction losses into a **feature** space
- Deep **invertible** models: valid density in image space

Going beyond conditional independence

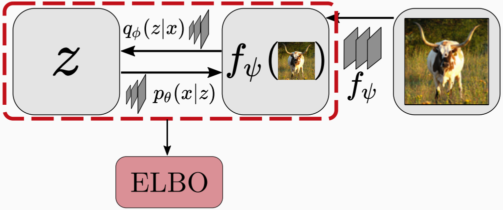
Avoiding strong parametric assumptions

- Lift reconstruction losses into a **feature** space
- Deep **invertible** models: valid density in image space
- Retain **fast** sampling for adversarial training

Maximum likelihood estimation with feature targets



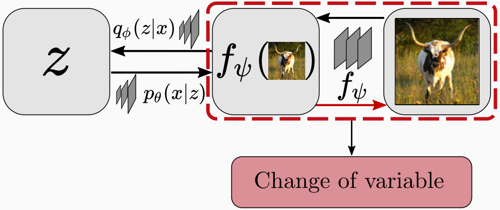
Maximum likelihood estimation with feature targets



Amortized Variational inference in feature space:

$$\mathcal{L}_{\theta,\phi,\psi}(\boldsymbol{x}) = \underbrace{-\mathbb{E}_{q_\phi(\boldsymbol{z}|\boldsymbol{x})}\left[\ln(p_\theta(f_\psi(\boldsymbol{x})|\boldsymbol{z}))\right] + D_{KL}(q_\phi(\boldsymbol{z}|\boldsymbol{x})||p_\theta(\boldsymbol{z}))}_{\text{Evidence lower bound in feature space}} - \ln\left|\det\frac{\partial f_\psi}{\partial \boldsymbol{x}}\right|$$

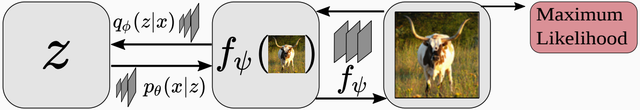
Maximum likelihood estimation with feature targets



Amortized Variational inference in feature space:

$$\mathcal{L}_{\theta,\phi,\psi}(\boldsymbol{x}) = -\mathbb{E}_{q_\phi(\boldsymbol{z}|\boldsymbol{x})}\left[\ln(p_\theta(f_\psi(\boldsymbol{x})|\boldsymbol{z}))\right] + D_{KL}(q_\phi(\boldsymbol{z}|\boldsymbol{x})||p_\theta(\boldsymbol{z})) \underbrace{- \ln\left|\det \frac{\partial f_\psi}{\partial \boldsymbol{x}}\right|}_{\text{Ch. of Var.}}$$

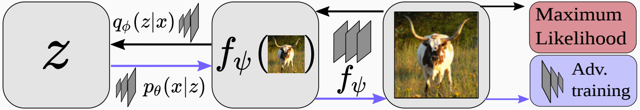
Maximum likelihood estimation with feature targets



Amortized Variational inference in feature space:

$$\mathcal{L}_{\theta,\phi,\psi}(\boldsymbol{x}) = -\mathbb{E}_{q_\phi(\boldsymbol{z}|\boldsymbol{x})}\left[\ln(p_\theta(f_\psi(\boldsymbol{x})|\boldsymbol{z}))\right] + D_{KL}(q_\phi(\boldsymbol{z}|\boldsymbol{x})||p_\theta(\boldsymbol{z})) - \ln\left|\det\frac{\partial f_\psi}{\partial \boldsymbol{x}}\right|$$

Maximum likelihood estimation with feature targets



Amortized Variational inference in feature space:

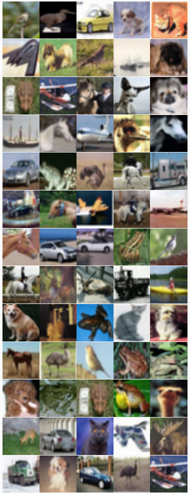
$$\mathcal{L}_{\theta,\phi,\psi}(\boldsymbol{x}) = -\mathbb{E}_{q_\phi(\boldsymbol{z}|\boldsymbol{x})}\left[\ln(p_\theta(f_\psi(\boldsymbol{x})|\boldsymbol{z}))\right] + D_{KL}(q_\phi(\boldsymbol{z}|\boldsymbol{x})||p_\theta(\boldsymbol{z})) - \ln\left|\det \frac{\partial f_\psi}{\partial \boldsymbol{x}}\right|$$

Adversarial training with Adaptive Density Estimation:

$$\mathcal{L}_{\text{adv}}(p_{\theta,\psi}) = \underbrace{-\mathbb{E}_{p_\theta(z)}\left[\ln \frac{D(f_\psi^{-1}(\mu_\theta(z)))}{1 - D(f_\psi^{-1}(\mu_\theta(z)))}\right]}_{\text{Adv. update using log ratio loss}}$$

Experiments on CIFAR10

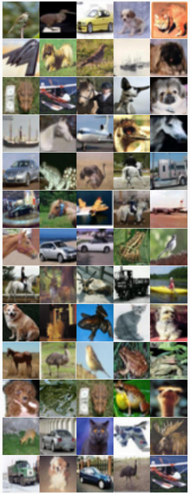
Samples



Real images

Experiments on CIFAR10

Samples

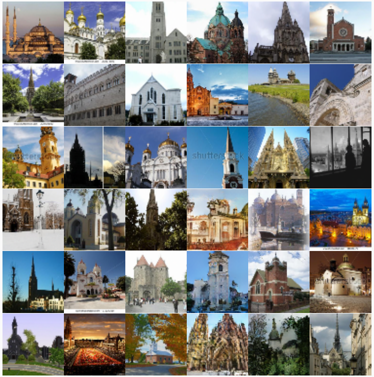


Real images

Model	BPD ↓	IS ↑	FID ↓
GAN			
WGAN-GP		7.9	
SNGAN		7.4	29.3
$\text{SNGAN}_{(R,H)}$		8.2	21.7
MLE			
VAE-IAF	3.1	$3.8^{\dagger}$	$73.5^{\dagger}$
NVP	3.5	$4.5^{\dagger}$	$56.8^{\dagger}$
Hybrid			
Ours (v1)	3.8	8.2	17.2
Ours (v2)	3.5	6.9	28.9
FlowGan	4.2	3.9	

Samples and real images (LSUN churches, 64×64)

Samples @ 4.3 BPD



Real images

Thank you for listening. Come see us at poster 71 :)