Generalized Energy Based Models

Michael Arbel[1] Liang Zhou[1]
Arthur Gretton[1]

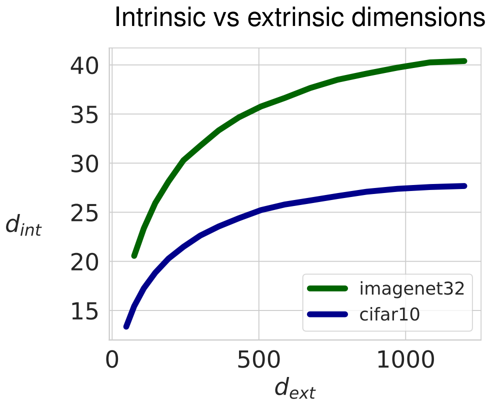
[1]Gatsby Computational Neuroscience Unit
University College London

March 25, 2021

Motivation

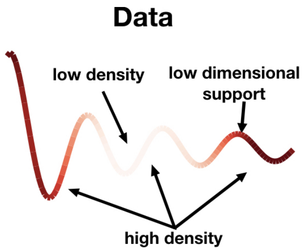
Problem

- ✓ Setting: data distributions with *small intrinsic dimension* embedded in a space with *high extrinsic dimension*.
- ✓ Example: Includes data such as natural images (Thiry et al. 2021).
- ✓ Goal: Flexible models exploiting low intrinsic dimensionality.



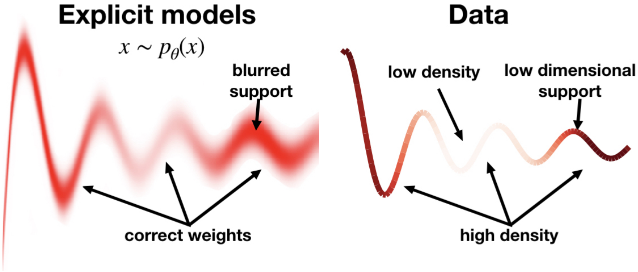
Motivation

Modeling data with low dimensional support and multiple modes.



Motivation

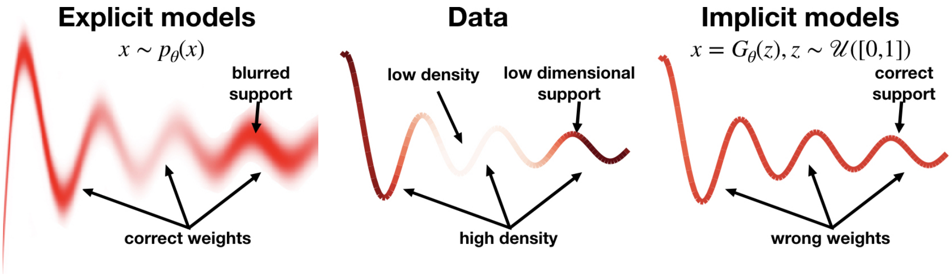
Modeling data with low dimensional support and multiple modes.



- An *explicit model* puts mass on the whole space: it blurs the samples.

Motivation

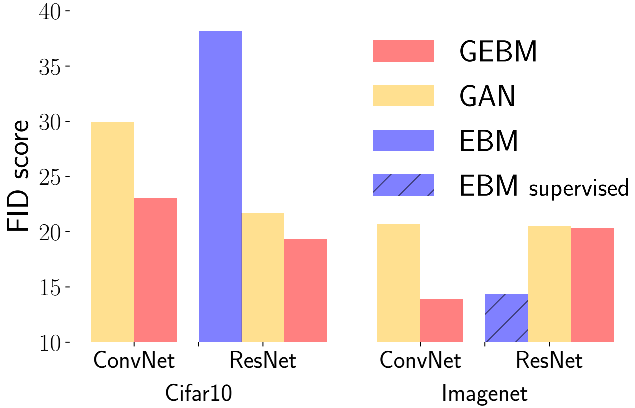
Modeling data with low dimensional support and multiple modes.



- *Explicit models* put mass on the whole space: blurring effect.
- *Implicit models* are wasteful: it throws the critic away (Azadi et al. 2019).

GEBMs: A new class of models for data with low intrinsic dimension

- ✓ Combines Implicit and Explicit models.
- ✓ Improves over GANs by using the critic information for sampling.
- ✓ Improves over EBMs by allowing their support to be learnable and low-dimensional.



Generalized energy-based models

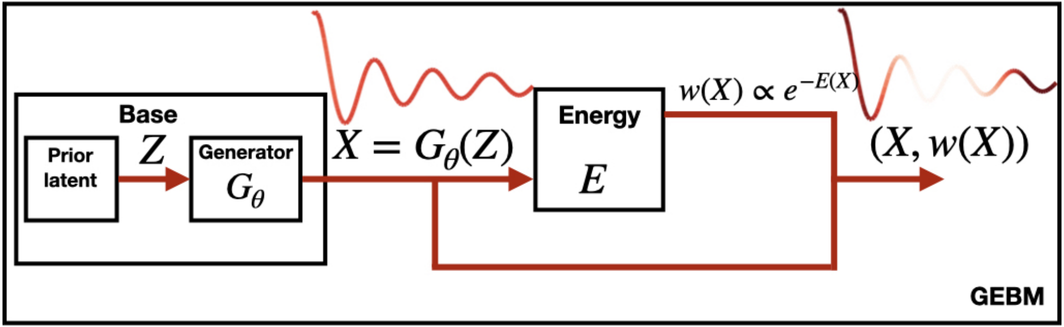
GEBMs are defined by a combination of the two components: *energy* and *base*

- The base $\mathbb{G}_\theta$ maps a prior latent noise $Z \sim \eta$ using a generator G_θ

$$X \sim \mathbb{G}_\theta \iff X = G_\theta(Z), \quad Z \sim \eta$$

- The energy E defines importance weights on the support of $\mathbb{G}_\theta$

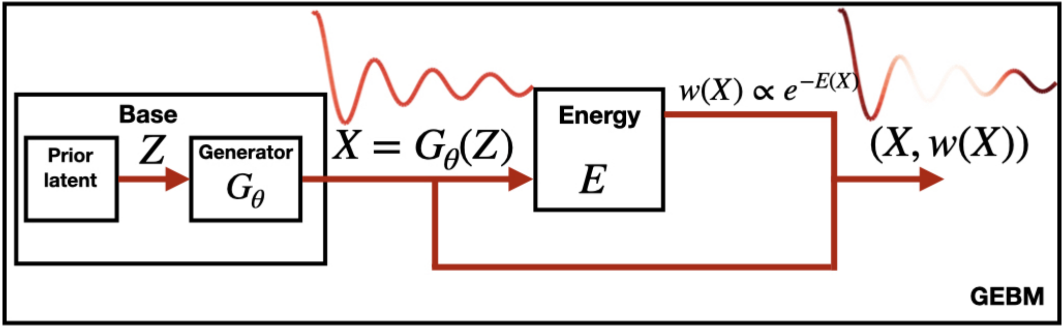
$$w(X) \propto \exp(-E(X))$$



Generalized energy-based models

GEBMs are defined by a combination of the two components: *energy* and *base*

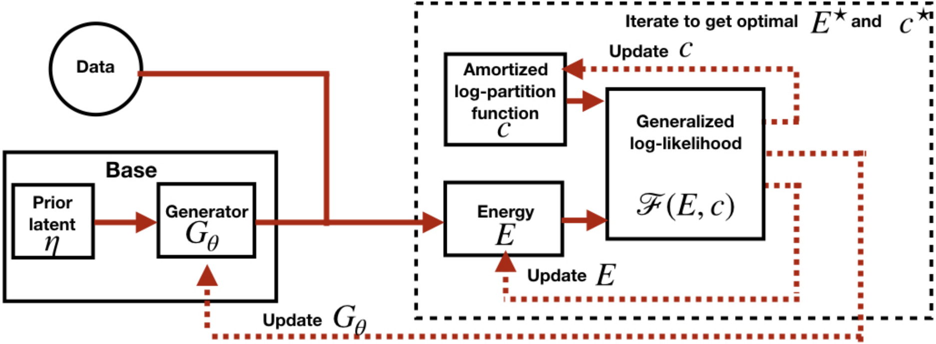
- The base learns the low-dimensional support of the data.
- The energy *refines* the mass on the low-dimensional support of the base.



Adversarial Training for GEBMs

Training with computational cost as GANs Goodfellow et al. 2014 and alternates between two stages

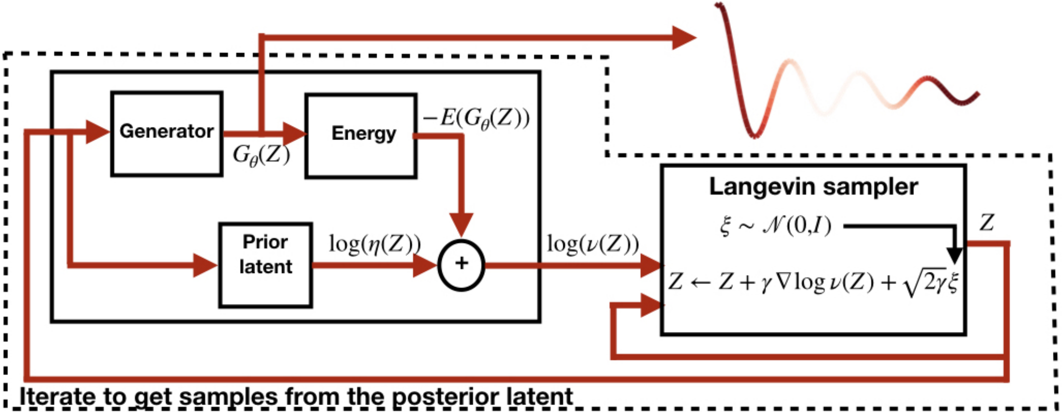
- Energy E: analogous to a critic in an f-GAN (Nowozin et al. 2016) and is learned by maximum likelihood with amortized estimation of the log-partition c of the model: $\max_{E,c} \mathcal{F}(E,c)$.
- Base $\mathbb{G}_\theta$: analogous to a generator in an f-GAN and is learned by minimizing the KL Approximate Lower-bound Estimator $KALE(\mathbb{P}|\mathbb{G}_\theta) := \max_{E,c} \mathcal{F}(E,c)$.



Sampling from GEBMs using Latent MCMC

Sampling from GEBM requires:

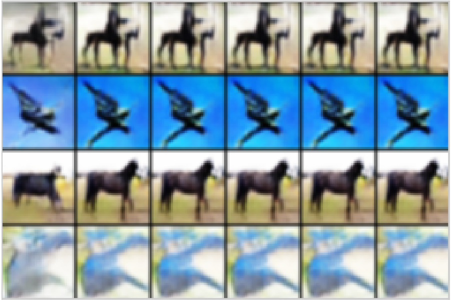
- Producing a posterior latent noise $Z \sim \nu$ using MCMC with
$$\nu(Z) \propto \eta(Z) e^{-E(G_\theta(Z))}$$
- Mapping the posterior noise using the generator $X = G_\theta(Z)$.

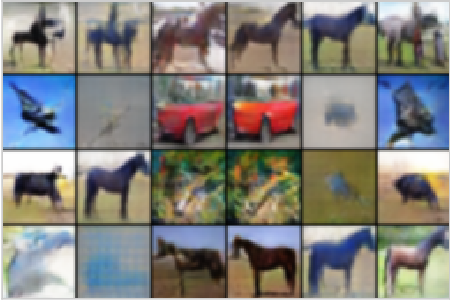


Sampling from GEBMs using Latent MCMC

Various MCMC sampler can be used and yield different behaviors.

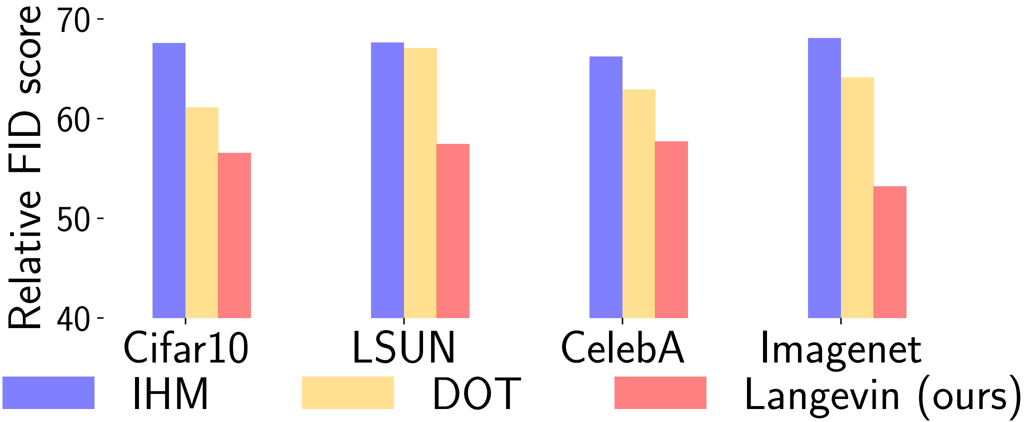
- Overdamped samplers (like ULA) stick to a particular mode within each chain.
- Kinetic samplers (like HMC) explore multiple modes within the same chain.





Sampling from GEBMs using Latent MCMC

✓ Sampling using MCMC in latent space exploits gradient information of the discriminator to accelerate mixing.



Conclusion

GEBM: A new family of models bridging the gap between GANs and EBMs.

✓ Modeling data with small intrinsic dimension:
  - The base learns the low dimensional support of data
  - The energy refines the distribution of mass on the base.

✓ Adversarial training

✓ Latent MCMC sampling

✓ Improves over GANs and EBMs

Code:

`https://github.com/MichaelArbel/GeneralizedEBM`

Thank you !

Azadi, Samaneh et al. (2019). "Discriminator Rejection Sampling". In: arXiv:1810.06758 [cs, stat]. arXiv: 1810.06758. URL: http://arxiv.org/abs/1810.06758 (visited on 02/06/2020).

Goodfellow, Ian J. et al. (2014). "Generative Adversarial Networks". In: arXiv:1406.2661 [cs, stat]. arXiv: 1406.2661. URL: http://arxiv.org/abs/1406.2661 (visited on 03/18/2019).

Nowozin, Sebastian, Botond Cseke, and Ryota Tomioka (2016). "f-GAN: Training Generative Neural Samplers using Variational Divergence Minimization". In: eprint: 1606.00709. URL: https://arxiv.org/abs/1606.00709.

Thiry, Louis et al. (2021). "The Unreasonable Effectiveness of Patches in Deep Convolutional Kernels Methods". In: ICLR.