Toward Improving Deep Generative Models on Sparse Data with Sparse Convolutional Nets and Point Nets

Michael Albergo

1 Overview

Deep generative modeling is a class of methods that seeks to describe a probability distribution $P(\mathbf{x})$ over some potentially high-dimensional $\mathbf{x}$. Generative modeling offers some possible advantages over discriminative approaches, such as more principled methods for dealing with missing data [1] and producing multi-modal output where a single input could be mapped to multiple possible outputs [2]. Recent new generative techniques such as variational auto-encoders (VAEs) [3], generative adversarial networks (GANs) [4], and the combination of them have emerged to show great promise for a number of common challenges in the field of machine learning, such as image generation and text modeling.

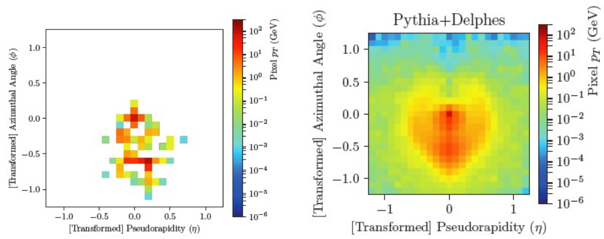

Figure 1: Physics example of sparsely distributed data. 1: Energy from single QCD jet. 2: Average of many jet images [5].

However, limitations remain on convergent learning and sample resolution across the two methods that are discussed below. In particular, these problems may be exacerbated for modeling sparse distributions, and the deep neural networks methods embedded in them, such as the convolutional approaches often taken in image analysis, might break down [6] or the learning process might be volatile and unstable. Sparse learning is essential for dealing with circumstances where signals are small and/or spread across a large domain that may be highly-dimensional, or when only part of the signal is currently available, but inference is desired on it in its entirety. This has clear application in sciences such as particle physics, cosmology, and neuroscience, as well as signal processing in general. See Figure 1 for an example, in which a jet of gluons smears its energy into a detector. The contour, intensities, and spread of this smearing are dependent on a number of physical qualities of the particle, but averaged over many samples can take a denser form. In real collisions, only the lefthand image is available. Recent methods for dealing with thinly distributed data using sparse convolutional nets [6] and point set models [7] could be applied to these novel generative cases. I intend to investigate how these approaches could alleviate the statistical shortcomings in generating sparse data signals as well as develop new methods for dealing with sparse data in the context of novel deep latent generative models (DLGMs).

1.1 GANs

While there are numerous approaches to deep generative modeling already [8, 9, 10], GANs and VAEs have been a center point of attention in research communities for their current success and future prospects. GANs as originally proposed by [4] are made of two competing networks. A generator takes in a noise sample $\mathbf{z}$ from some prior and outputs data $\mathbf{x}_{gen}$, and a discriminator takes in $\mathbf{x}_{real}$ and $\mathbf{x}_{gen}$ and outputs a value between [0,1] that represents the likelihood of the input being a sample from the real distribution. The two networks essentially play a mini-max game, where the generator tries to minimize and the discriminator tries to maximize the value function:

$$\min_G \max_D V(G, D) = \mathbb{E}_x[\ln(D(\mathbf{x})] + \mathbb{E}_z[\ln(1 - D(G(\mathbf{z}))] \tag{1}$$

The training process is shown to be equivalent to minimizing the Jensen-Shannon divergence given by:

$$D_{\mathrm{JS}} = \frac{1}{2} D_{\mathrm{KL}}(P_{real}||\frac{1}{2}(P_{real} + P_{gen})) + D_{\mathrm{KL}}(P_{gen}||\frac{1}{2}(P_{real} + P_{gen})) \tag{2}$$

for Kullback-Leibler divergence $D_{\mathrm{KL}}(P||Q) = -\sum_i P(i) \ln \frac{Q(i)}{P(i)}$.

By such, there is no explicit computation of likelihoods. The reliance on f-divergence that the original GAN structure (and some variations of it) employs causes limitations that could arise from a number of theoretical problems as detailed in [11]: overfitting, density misspecification, or dimensional misspecification. GANs are thus susceptible to collapsing on one mode of the data or computing non-finite gradient updates during backpropagation. There are improvements to these shortcomings by changing the objective metric to integral probability metrics [12], applying regularization on said metric [11] or on missing data modes [13], combining regularization with a new objective function [14], or by adding noise to the discriminator input to create a smoother probability distribution [15] to prevent vanishing gradients. These methods must be theoretically considered in the instance of sparse data, as it is not immediately clear that they will hold as solutions. For example, a continuous noise term might dominate the desired feature information if the signal is sparse enough, like a 2D image with only a few scattered active pixels of low scaled intensity.

1.2 VAEs

Rooted in Bayesian inference, variational auto-encoders generate approximate maximum likelihood estimations on $P(\mathbf{x})$ by using variational inference to avoid the intractable calculation of the marginal likelihood. That is, $P(\mathbf{z})$ in the MLE equation $P(\mathbf{x}) = \int P(\mathbf{x}|\mathbf{z})P(\mathbf{z})$ is estimated by making inference on $P(\mathbf{z}|\mathbf{x})$ via maximizing the first term, known as the evidence lower bound (ELBO), in

$$\ln p(\mathbf{x}) = \mathbb{E}_{q(\mathbf{z}|\mathbf{x})}[\frac{\ln p(\mathbf{x},\mathbf{z})}{\ln q(\mathbf{z}|\mathbf{x})}] + D_{\mathrm{KL}}(q(\mathbf{z}|\mathbf{x})||p(\mathbf{z}|\mathbf{x})) \qquad \mathcal{L}_{rec} = \mathbb{E}_{q(\mathbf{z}|\mathbf{x})}[\ln p(\mathbf{x},\mathbf{z}) - \ln p(\mathbf{z}|\mathbf{x})] \tag{3}$$

for some learned encoding $q(\mathbf{z}|\mathbf{x})$ of the posterior $p(\mathbf{z}|\mathbf{x})$ [3, 16]. Learning is done through gradient descent over the parameters of q and p jointly to optimize this variational lower bound. This is essentially minimizing reconstruction error between the original and decoded input. This reconstruction error has been argued to poorly encode meaningful information into the latent space [17] which limits image quality and variety, causing, for example, blurry reconstructed images. While this problem has been addressed by changing the regularization process [17], embedding adversarial learning into the loss function [18], or by replacing the normally Gaussian $q(\mathbf{z}|\mathbf{x})$ with a black-box inference and adversarial training [19], there is no evidence that this training will be convergent on sparse data, especially using those remedies which depend on adversarial learning mechanisms that are detailed as unstable above.

2 Sparse Neural Nets and Point Nets

Current approaches, like that of [20] which employs VAEs for compressed sensing that may require learning sparse representations of data or small signals from otherwise noisy data, still uphold the caveat that theoretical convergence is only reached if gradient descent doesn't collapse. The issue of unstable gradients, as described by the three possible problems above for adversarial and variational learning, is precisely a problem I would like to work toward solving or simplifying, especially in the contexts of sparse data. Novel neural network architectures that deal with sparse images [6, 21] and point cloud data representation models [7, 22] could be productive avenues for avoiding some of the increased instability that comes from modeling small and potentially highly variable signals. Point cloud methods, though implemented in different ways (Qi et al. (2017) avoids 2D convolution altogether, whereas Achlioptas et al. (2017) use ShapeNet [23] to do convolution in neighboring points), help avoid the grid consolidation that occurs in image construction with a compact characterization of the data. Point nets have also been implemented to project between 2D and 3D [24], allowing for manipulation of dimensionality to reconstruct or maintain information. In addition or in tandem to this, sparse convolutional nets could be implemented where normal convolutional filtering is not well defined – when any input in a filter may appear or disappear between samples. Efforts had have been made to incorporate sparsity into the architecture, retaining information about which hidden variables are active for certain inputs [21]. Other approaches detail adding sparse convolutional layers that only operate on observed pixels with binary max-pooling [6]. This method shows invariance to the level of sparsity. In the context of generative signal processing, it is of interest to combine these sparsity-invariant models with translation and rotation-invariant models like that of [25], which is useful for sparse signals that maintain the same structure, but may flip or rotate depending on the variables of what created the signal.

Constructing deep generative models out of these techniques and exploring the theoretical benefits of these implementations, especially in the context of GANs, VAEs, and perhaps any new learning structure created or that we create as a byproduct of this research, is of interest for this project. In addition, empirical comparisons of outputs across different methods will be made, with sparse consideration and without, to elucidate potential benefits. Point nets and the aforementioned approaches of embedding sparse qualities into network architectures are two promising starting points to examine the production of sparse generative models. Lastly, these noted improvements for dealing with thinly distributed signals are not necessarily plenary solutions for the problem, and it is of interest in this project to further consider the development of new approaches to representing sparse information or to draw on inspiration from past approaches to sparse vector estimation in context of deep networks [26].

References

[1] I. Goodfellow, "NIPS 2016 Tutorial: Generative Adversarial Networks," 2016.

[2] J. A. Lasserre, C. M. Bishop, and T. P. Minka, "Principled hybrids of generative and discriminative models," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, no. 6, pp. 87–94, 2006.

[3] D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," no. Ml, pp. 1–14, 2013.

[4] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Networks," 2014.

[5] S. Sivarajah and C. Lester, "Machine Learning Based Simulation of Particle Physics Detectors," Tech. Rep. May, University of Cambridge, 2017.

[6] J. Uhrig, N. Schneider, L. Schneider, U. Franke, T. Brox, and A. Geiger, "Sparsity Invariant CNNs," 2017.

[7] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "PointNet:Deep Learning on Point Sets for 3D Classification and Segmentation," pp. 601–610, IEEE, oct 2016.

[8] G. Hinton, P. Dayan, B. Frey, and R. Neal, "The "wake-sleep" algorithm for unsupervised neural networks," *Science*, vol. 268, no. 5214, pp. 1158–1161, 1995.

[9] L. Dinh, J. Sohl-Dickstein, and S. Bengio, "Density estimation using Real NVP," 2016.

[10] Y. Bengio, É. Thibodeau-Laufer, G. Alain, and J. Yosinski, "Deep Generative Stochastic Networks Trainable by Backprop," 2013.

[11] K. Roth, A. Lucchi, S. Nowozin, and T. Hofmann, "Stabilizing Training of Generative Adversarial Networks through Regularization," vol. 1, no. 2, pp. 1–16, 2017.

[12] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," 2017.

[13] T. Che, Y. Li, A. P. Jacob, Y. Bengio, and W. Li, "Mode Regularized Generative Adversarial Networks," pp. 1–13, 2016.

[14] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, "Improved Training of Wasserstein GANs," pp. 1–19, 2017.

[15] M. Arjovsky and L. Bottou, "Towards Principled Methods for Training Generative Adversarial Networks," pp. 1–17, 2017.

[16] D. P. Kingma, *Variational Inference & Deep Learning: A New Synthesis.* PhD thesis, University of Amsterdam, 2017.

[17] S. Zhao, J. Song, and S. Ermon, "Towards Deeper Understanding of Variational Autoencoding Models," 2017.

[18] A. Makhzani, J. Shlens, N. Jaitly, I. Goodfellow, and B. Frey, "Adversarial Autoencoders," 2015.

[19] L. Mescheder, S. Nowozin, and A. Geiger, "Adversarial Variational Bayes: Unifying Variational Autoencoders and Generative Adversarial Networks," 2017.

[20] A. Bora, A. Jalal, E. Price, and A. G. Dimakis, "Compressed Sensing using Generative Models," 2017.

[21] B. Graham, "Spatially-sparse convolutional neural networks," pp. 1–13, 2014.

[22] P. Achlioptas, O. Diamanti, I. Mitliagkas, and L. Guibas, “Representation Learning and Adversarial Generation of 3D Point Clouds,” pp. 1–20, 2017.

[23] A. X. Chang, T. Funkhouser, L. Guibas, P. Hanrahan, Q. Huang, Z. Li, S. Savarese, M. Savva, S. Song, H. Su, J. Xiao, L. Yi, and F. Yu, “ShapeNet: An Information-Rich 3D Model Repository,” 2015.

[24] H. Fan, H. Su, and L. Guibas, “A Point Set Generation Network for 3D Object Reconstruction from a Single Image,” 2016.

[25] D. Tao, “Transform-Invariant Convolutional Neural Networks for Image Classification and Search,” *Acm Mm*, pp. 1345–1354, 2016.

[26] B. Xin, Y. Wang, W. Gao, and D. Wipf, “Maximal Sparsity with Deep Networks?,” no. 1, pp. 1–31, 2016.