COMS 4995 Methods is Unsupervised Machine Learning (Fall 2018) Project Proposal

Alexandre Lamy, Trung Vu, Sean Lee - {all2187, ttv2107, sl3794}@columbia.edu

October 15, 2018

Background

The seminal work done in [1] proposes a novel architecture, GANs, for estimating generative models. In addition to describing the model and giving some experimental performances, this work shows that the global optima in the space of arbitrary generative and discriminative functions is that where the generator perfectly learns the data distribution and the discriminator is the naive 0.5 random guesser. However, this leaves unanswered two main more important questions. First, the results only apply when the discriminative and generative functions are arbitrarily powerful (i.e. as stated in [1] it applies only in the "non-parametric limit"). What happens when we use depth and size limited neural networks instead? Second, this paper looks at the theoretical global optima and gives no explanation as to what happens when the model is trained on finitely many samples. Some sort of generalization guarantee, i.e. how close the generator's distribution comes to the true data distribution as a function of sample size, is dearly needed.

Objective

In short our objective will be to look at these two questions, i.e. *what are the generalization guarantees of GANs when trained on finite samples*, and *how does restricting the class of generative and/or descrimative functions affect those guarantees*. We will seek first to understand the partial answers that currently exists to these questions and possibly to extend and improve upon those answers.

Recent advancements

As mentioned, we will start by understanding the partial answers that some recent (2017 and 2018) papers have given to our two core questions. In particular there are five papers that we will look at to start our project.

The authors of [2] consider the convergence of real life GANs training. As mentioned before, in practice it is impossible to have arbitrarily complex generative and discriminative functions. This means that many local Nash equilibriums could exist. Furthermore, since GANs is formulated as a two player game in which the players have competing goals, techniques like gradient descent could fail to converge at all. The authors of [2] show that by using a two time scale update rule (TTUR), where the discriminator is updated more often than the generator, one can prove that the training of GANs with TTUR and Adam will converge to a stationary local Nash equilibrium. The authors

also propose the "Frechet Inception Distance" (FID) as a new and more consistent way to evaluate GANs.

The authors of [3] pose two questions. First, they tackle the finite sample case. How many examples do we need in order for the empirical distance between the generated and true distributions to converge to the population distance, and under what distance does convergence hold? The answer is that there is a notion of distance, defined as neural network distance, that allows convergence, but this notion of distance is so weak that even when it is small, it is hard to conclude that the generated and true distributions are actually close. Second, they tackle the assumption of infinite power in [1]. They show that a generator with finite capacity can still win against a discriminator with finite capacitor – thus explaining why optimization halts in practice.

[4] addresses an issue raised by [3]: the neural network distance converges even if the generated distribution is not close to the true distribution. Specifically, it might converge even if the generated distribution has small support, leading to a phenomenon called mode collapse and a lack of diversity in the generated examples. Thus it proposes a test based on the Birthday Paradox: if a distribution has small support then a random batch will most likely contain duplicates.

The authors of [5] render GANs' performance into a trade-off between the discriminative and generative power of the network, and provide theoretical guarantees on the discriminative powers of the network in case the discriminators are dense enough span the set of bounded continuous functions; on the other hand, the generative powers are guaranteed when the discriminator set is small enough. They provide a metric called *neural distance* as a way to quantify such an upper bound.

The authors of [6] elaborate on the convergence rate of GANs given characteristics of the objective function: if the objective function satisfies *strict adversarial divergence*, then there is a weak convergence of the GAN output distribution to the objective function.

Persisting issues

As seen in the above, although our questions have been partially answered in various restricted settings, the convergence results are quite weak and generalization guarantees are almost non-existent. In addition to getting an understanding of these partial results, we want to at least investigate why stronger results haven't been found, or in other words find out what causes the difficulty of this problem.

References

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems 27* (Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, eds.), pp. 2672–2680, Curran Associates, Inc., 2014.

[2] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "Gans trained by a two time-scale update rule converge to a local nash equilibrium," in *Advances in Neural Information Processing Systems 30* (I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, eds.), pp. 6626–6637, Curran Associates, Inc., 2017.

[3] S. Arora, R. Ge, Y. Liang, T. Ma, and Y. Zhang, "Generalization and equilibrium in generative adversarial nets (gans)," in *International Conference on Machine Learning*, pp. 224–232, 2017.

[4] S. Arora, A. Risteski, and Y. Zhang, "Do GANs learn the distribution? some theory and empirics," in *International Conference on Learning Representations*, 2018.

[5] P. Zhang, Q. Liu, D. Zhou, T. Xu, and X. He, "On the discrimination-generalization tradeoff in GANs," in *International Conference on Learning Representations*, 2018.

[6] S. Liu, O. Bousquet, and K. Chaudhuri, "Approximation and convergence properties of generative adversarial learning," in *Advances in Neural Information Processing Systems*, pp. 5545–5553, 2017.