Stabilizing GAN Training with Multiple Random Projections

Behnam Neyshabur **Srinadh Bhojanapalli** **Ayan Chakrabarti**
Toyota Technological Institute at Chicago
6045 S. Kenwood Ave., Chicago, IL 60637
{bneyshabur,srinadh,ayanc}@ttic.edu

Abstract

Training generative adversarial networks is unstable in high-dimensions when the true data distribution lies on a lower-dimensional manifold. The discriminator is then easily able to separate nearly all generated samples leaving the generator without meaningful gradients. We propose training a single generator simultaneously against an array of discriminators, each of which looks at a different random low-dimensional projection of the data. We show that individual discriminators then provide stable gradients to the generator, and that the generator learns to produce samples consistent with the full data distribution to satisfy all discriminators. We demonstrate the practical utility of this approach experimentally, and show that it is able to produce image samples with higher quality than traditional training with a single discriminator.

1 Introduction

Generative adversarial networks (GANs), introduced by [1], endow neural networks with the ability to express distributional outputs. The framework includes a generator network that is tasked with producing samples from some target distribution, given as input a (typically low dimensional) noise vector drawn from a simple known distribution, and possibly conditional side information. The generator learns to generate such samples, not by directly looking at the data, but through adversarial training with a discriminator network that seeks to differentiate real data from those generated by the generator. To satisfy the objective of "fooling" the discriminator, the generator eventually learns to produce samples with statistics that match those of real data.

In regression tasks where the true output is ambiguous, GANs provide a means to simply produce an output that is plausible (with a single sample), or to explicitly model that ambiguity (through multiple samples). In the latter case, they provide an attractive alternative to fitting distributions to parametric forms during training, and employing expensive sampling techniques at the test time. In particular, conditional variants of GANs have shown to be useful for tasks such as in-painting [2], and super-resolution [3]. Recently, [4] demonstrated that GANs can be used to produce plausible mappings between a variety of domains—including sketches and photographs, maps and aerial views, segmentation masks and images, *etc*. GANs have also found uses as a means of un-supervised learning, with latent noise vectors and hidden-layer activations of the discriminators proving to be useful features for various tasks [2, 5, 6].

Despite their success, training GANs to generate high-dimensional data (such as large images) is challenging. Adversarial training between the generator and discriminator involves optimizing a min-max objective. This is typically carried out by gradient-based updates to both networks, and the generator is prone to divergence and mode-collapse as the discriminator begins to successfully distinguish real data from generated samples with high confidence. Researchers have tried to address this instability and train better generators through several techniques. [7] proposed explicitly factorizing generating an image into a sequence of conditional generations of levels of a Laplacian