Common Problems

GANs have a number of common failure modes. All of these common problems are areas of active research. While none of these problems have been completely solved, we'll mention some things that people have tried.

Vanishing Gradients

Research (https://arxiv.org/pdf/1701.04862.pdf) has suggested that if your discriminator is too good, then generator training can fail due to vanishing gradients (https://wikipedia.org/wiki/Vanishing_gradient_problem). In effect, an optimal discriminator doesn't provide enough information for the generator to make progress.

Attempts to Remedy

- **Wasserstein loss**: The Wasserstein loss (/machine-learning/gan/loss) is designed to prevent vanishing gradients even when you train the discriminator to optimality.
- **Modified minimax loss**: The original GAN paper (http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf) proposed a modification to minimax loss (/machine-learning/gan/loss) to deal with vanishing gradients.

Mode Collapse

Usually you want your GAN to produce a wide variety of outputs. You want, for example, a different face for every random input to your face generator.

However, if a generator produces an especially plausible output, the generator may learn to produce *only* that output. In fact, the generator is always trying to find the one output that seems most plausible to the discriminator.

If the generator starts producing the same output (or a small set of outputs) over and over again, the discriminator's best strategy is to learn to always reject that output. But if the next generation of discriminator gets stuck in a local minimum and doesn't find the best strategy, then it's too easy for the next generator iteration to find the most plausible output for the current discriminator.

Each iteration of generator over-optimizes for a particular discriminator, and the discriminator never manages to learn its way out of the trap. As a result the generators rotate through a small set of output types. This form of GAN failure is called **mode collapse**.

Attempts to Remedy

The following approaches try to force the generator to broaden its scope by preventing it from optimizing for a single fixed discriminator:

- **Wasserstein loss**: The Wasserstein loss (/machine-learning/gan/loss) alleviates mode collapse by letting you train the discriminator to optimality without worrying about vanishing gradients. If the discriminator doesn't get stuck in local minima, it learns to reject the outputs that the generator stabilizes on. So the generator has to try something new.
- **Unrolled GANs**: Unrolled GANs (https://arxiv.org/pdf/1611.02163.pdf) use a generator loss function that incorporates not only the current discriminator's classifications, but also the outputs of future discriminator versions. So the generator can't over-optimize for a single discriminator.

Failure to Converge

GANs frequently fail to converge, as discussed in the module on training (/machine-learning/gan/training).

Attempts to Remedy

Researchers have tried to use various forms of regularization to improve GAN convergence, including:

- **Adding noise to discriminator inputs**: See, for example, Toward Principled Methods for Training Generative Adversarial Networks (https://arxiv.org/pdf/1701.04862.pdf).
- **Penalizing discriminator weights**: See, for example, Stabilizing Training of Generative Adversarial Networks through Regularization (https://arxiv.org/pdf/1705.09367.pdf).

Previous
← Check Your Understanding (/machine-learning/gan/check)
Next

GAN Variations (/machine-learning/gan/applications) →

Last updated 2020-02-10 UTC.