

# Generative Adversarial Nets

Cheng Guan

July 22, 2018

## Abstract

*The authors propose a new framework for estimating generative models via an adversarial process, in which they simultaneously train two models: a generative model  $G$  that captures the data distribution, and a discriminative model  $D$  that estimates the probability that a sample came from the training data rather than  $G$ . The training procedure for  $G$  is to maximize the probability of  $D$  making a mistake. This framework corresponds to a minimax two-player game. In the space of arbitrary functions  $G$  and  $D$ , a unique solution exists, with  $G$  recovering the training data distribution and  $D$  equal to  $\frac{1}{2}$  everywhere.*

## 1. Introduction

The promise of deep learning is to discover rich, hierarchical models that represent probability distributions over the kinds of data encountered in artificial intelligence applications, such as natural images, audio waveforms containing speech, and symbols in natural language corpora. So far, the most striking successes in deep learning have involved discriminative models, usually those that map a high-dimensional, rich sensory input to a class label [3, 5]. These striking successes have primarily been based on the backpropagation and dropout algorithms, using piecewise linear units [2, 4] which have a particularly well-behaved gradient. Deep generative models have had less of an impact, due to the difficulty of approximating many intractable probabilistic computations that arise in maximum likelihood estimation and related strategies, and due to difficulty of leveraging the benefits of piecewise linear units in the generative context.

In the proposed adversarial nets framework, the generative model is pitted against an adversary: a discriminative model that learns to determine whether a sample is from the model distribution or the data distribution. The generative model can be thought of as analogous to a team of counterfeiters, trying to produce fake currency and use it without detection, while the discriminative model is analogous to the police, trying to detect the counterfeit currency. Compe-

tition in this game drives both teams to improve their methods until the counterfeits are indistinguishable from the genuine articles. In this article, the authors explore the special case when the generative model generates samples by passing random noise through a multilayer perceptron, and the discriminative model is also a multilayer perceptron. They refer to this special case as adversarial nets.

## 2. Related work

Until recently, most work on deep generative models focused on models that provided a parametric specification of a probability distribution function. The model can then be trained by maximizing the log likelihood. In this family of model, perhaps the most successful is the deep Boltzmann machine. Generative stochastic networks [1] are an example of a generative machine that can be trained with exact backpropagation rather than the numerous approximations required for Boltzmann machines. This work extends the idea of a generative machine by eliminating the Markov chains used in generative stochastic networks.

Their work backpropagates derivatives through generative processes by using the observation that

$$\lim_{\sigma \rightarrow 0} \nabla_x \mathbb{E}_{\epsilon \sim \mathcal{N}(0, \sigma^2 I)} f(x + \epsilon) = \nabla_x f(x) \quad (1)$$

Previous work has also taken the approach of using a discriminative criterion to train a generative model [6]. These approaches use criteria that are intractable for deep generative models. These methods are difficult even to approximate for deep models because they involve ratios of probabilities which cannot be approximated using variational approximations that lower bound the probability.

## References

- [1] Y. Bengio, E. Laufer, G. Alain, and J. Yosinski. Deep generative stochastic networks trainable by backprop. In *ICML*, 2014. 1
- [2] I. J. Goodfellow, D. Warde-Farley, M. Mirza, A. Courville, and Y. Bengio. Maxout networks. *arXiv preprint arXiv:1302.4389*, 2013. 1

- [3] G. Hinton, L. Deng, D. Yu, G. E. Dahl, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, and T. N. Sainath. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, 2012. 1
- [4] K. Jarrett, K. Kavukcuoglu, and Y. LeCun. What is the best multi-stage architecture for object recognition? In *ICCV*, pages 2146–2153, 2009. 1
- [5] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet classification with deep convolutional neural networks. In *NIPS*, 2012. 1
- [6] Z. Tu. Learning generative models via discriminative approaches. In *CVPR*, 2007. 1