Generative Adversarial Nets

Ian J. Goodfellow, Jean Pouget-Abadie[*]**, Mehdi Mirza,
Bing Xu, David Warde-Farley, Sherjil Ozair**[†]**, Yoshua
Bengio**[‡]
Département d'informatique et de recherche opérationnelle
Université de Montréal
Montréal, QC H3C 3J7

Abstract

We propose a new framework for estimating generative models
via an adversarial process, in which we simultaneously train two
models: a generative model G that captures the data distribu-
tion, and a discriminative model D that estimates the probabil-
ity 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. In the case where
G and D are defined by multilayer perceptrons, the entire sys-
tem can be trained with backpropagation. There is no need for
any Markov chains or unrolled approximate inference networks
during either training or generation of samples. Experiments
demonstrate the potential of the framework through qualitative
and quantitative evaluation of the generated samples.

1   Introduction

The promise of deep learning is to discover rich, hierarchical models [1] that rep-
resent probability distributions over the kinds of data encountered in artificial

---

[*]Jean Pouget-Abadie is visiting Université de Montréal from Ecole Polytechnique.
[†]Sherjil Ozair is visiting Université de Montréal from Indian Institute of Technology Delhi.
[‡]Yoshua Bengio is a CIFAR Senior Fellow.

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 [4, 7]. These striking successes have primarily been based on the backpropagation and dropout algorithms, using piecewise linear units [6, 2, 3] 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 nits in the generative context. We propose a new generative model estimation procedure that sidesteps these difficulties.[1]

In the proposed *adversarial nets* framework, the generative model is pitted against an adversary: a discriminate model that leans 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. Competition in this game drives both teams to improve their methods until the counterfeits are indistinguishable from the genuine articles.

This framework can yield specific training algorithms for many kinds of model and optimization algorithm. In this article, we explore the special case when the generative model generates sample by passing random noise through a multilayer perceptron, and the discriminative model is also a multilayer perceptron. We refer to this special case as *adversarial nets*. In this case, we can train both models using only the highly successful backpropagation and dropout algorithms [5] and sample from the generative model using only forward propagation. No approximate inference or Markov chains are necessary.

References

[1] Yoshua Bengio. Learning deep architectures for ai. *Foundations and Trends® in Machine Learning*, 2(1):1–127, 2009.

[2] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Deep sparse rectifier neural networks. In Geoffrey Gordon, David Dunson, and Miroslav Dudík, editors, *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, volume 15 of *Proceedings of Machine Learning Research*, pages 315–323, Fort Lauderdale, FL, USA, 11–13 Apr 2011. PMLR.

[3] Ian J. Goodfellow, David Warde-Farley, Mehdi Mirza, Aaron Courville, and Yoshua Bengio. Maxout networks, 2013.

---

[1]All code and hyperparameters available at `http://www.github.com/goodfeli/adversarial`

[4] Geoffrey Hinton, Li Deng, Dong Yu, George E. Dahl, Abdel-rahman Mohamed, Navdeep Jaitly, Andrew Senior, Vincent Vanhoucke, Patrick Nguyen, Tara N. Sainath, and Brian Kingsbury. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, 29(6):82–97, 2012.

[5] Geoffrey E. Hinton, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, and Ruslan R. Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors, 2012.

[6] Kevin Jarrett, Koray Kavukcuoglu, Marc'Aurelio Ranzato, and Yann Le-Cun. What is the best multi-stage architecture for object recognition? In *2009 IEEE 12th International Conference on Computer Vision*, pages 2146–2153, 2009.

[7] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C.J. Burges, L. Bottou, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012.