Paper Review 8

Generative Adversarial Nets

Authors: Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio

Introduction

In this work, the authors propose adversarial nets as an alternative to deep generative models, which, until that point, didn't have much impact. The latter failed, due to the difficulty to approximate many unmanageable probabilistic computations that arise in maximum likelihood estimation. The adversarial network comprise of a generator G and a discriminator D, both of which are trained simultaneously. The generator model tries to capture the data distribution and the discriminative model tries to estimate the probability that the sample came from training data rather than G. The backpropogation algorithm and dropout regularization technique is used to train each of these models.

The generative adversarial networks have similarity with variational autoencoders in that both pair a differentiable generator network with a second neural network. However, the second network in the latter requires differentiation through the hidden units, while GANs require differentiation through the visible units. The most relevant previous work for the second neural network in GANs is predictability minimization, wherein, each hidden unit is trained to be different from the output of the network. In GANs, one network produces a high dimensional vector that is used as an input to the second network, which in turn is equivalent to an optimization problem, with an objective function to be minimized.

In order to learn the distribution of the data p_{data}, and the generator's distribution p_g such that they both converge to the same value, the authors use multilayer perceptrons to simultaneously train the generator and the discriminator. While the training step can alternate between k-steps of optimizing D and 1 step of optimizing G, the authors choose to do the same by alternating between one step of each optimizations. In Figure 1, the authors plot the simultaneous training update of the discriminator and the generator, in which, the discriminator is unable to differentiate between the two distributions.

Theoretical Results & Experimentation

To separate samples from the data generating distribution, the generative adversarial networks are trained by simultaneously updating the discriminative distribution. The authors list the algorithm to achieve this, by using a minibatch stochastic gradient descent with the number of steps to apply to the discriminator being 1. Furthermore, the authors go on to prove that the global minimum of the training criterion is achieved if and only if the generative model perfectly replicates the data distribution.

The authors trained the adversial net model on a variety of datasets such as MNIST, the Toronto Face Database (TFD), CIFAR-10. They estimated the probability of the test set data under p_g by recording the log-likelihood of a distribution achieved by fitting a Gaussian parzen window on samples generated by G. Through these experiment s and results, the authors prove that the samples produced by adversarial nets are at the very least, competitive with those generated by the existing generative models. In Table 2, the authors summarize the advantages of adversarial models over other generative models.

Strengths & Weaknesses

The authors detail the advantages and disadvantages of this modeling framework relative to previous frameworks used for this purpose. The main advantage of the adversarial nets is that the Markov chains are never needed in the framework. Also, GANs learn the density distributions and internal representations of data, and thus, no inference is needed during learning. One of the biggest advantages of GANs is that they generate data that is similar to real data, thus having many different uses in the real world. One of the major disadvantages, as also noted by the authors is that because of the architecture of the model, the training of both the generators and the discriminators must be synchronized well, which otherwise might lead to the Helvetica scenario. The gradient descent might/might not lead to such an equilibrium in training. Also, it's hard to learn to generate discrete data from the generator models in GANs. This is one of the reasons why GANs have not been able to be employed as successfully for other types of data such as text, speech.

Further Work

The authors demonstrate the possibility of using GANs in an adversarial framework. In the context of images, other models such as deep convolutional generative adversarial networks [1], which learns a hierarchy of representations from the object parts in both the generator and the discriminator. In the context of improving the training of GANs, which was one of the major disadvantages of this architecture, a few training techniques and architectural changes were introduced by Salimans, et al. in [2], using which, they have been able to achieve state-of-the-art results on datasets like CIFAR-10 and SVHN.

References

[1] Radford, Alec, Luke Metz, and Soumith Chintala. "Unsupervised representation learning with deep convolutional generative adversarial networks." arXiv preprint arXiv:1511.06434 (2015).

[2] Salimans, Tim, et al. "Improved techniques for training gans." Advances in neural information processing systems. 2016.