TTIC 31230, Fundamentals of Deep Learning

David McAllester, Autumn 2020

Generative Adversarial Networks (GANs)

Modeling Probability Distributions on Images

Suppose we want to train a model of the probability distribution of natural images using cross-entropy loss.

$$\Phi^* = \operatorname*{argmin}_{\Phi} \; E_{y\sim\mathrm{pop}} \; -\ln p_\Phi(y)$$

Images are continuous stuctured objects — a continuous value at every pixel.

It is difficult to build probability models for images (or other continuous structured values) that both accurately model the distribution and also allow us to calculate $p_\Phi(y)$.

Generative Adversarial Networks (GANs)

GANs represent $p_\Phi(y)$ implicitly by constructing an image generator and abandon the ability to compute $p_\Phi(y)$.

The cross-entropy loss is replaced by an adversarial discriminator which tries to distinguish between generated images and real images.

Representing a Distribution with a Generator

z $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $y_\Phi(z)$

The random input z defines a probability density on images $y_\Phi(z)$. We will write this as $p_\Phi(y)$ for the image y.

Representing a Distribution with a Generator

z $\qquad\qquad\qquad\qquad\qquad\qquad$ $y_\Phi(z)$

We want $p_\Phi(y)$ to model a natural image distribution such as the distribution over human faces.

Representing a Distribution with a Generator

z $\qquad\qquad\qquad\qquad\qquad\qquad$ $y_\Phi(z)$

We can sample from $p_\Phi(y)$ by sampling z. But we cannot compute $p_\Phi(y)$ for y sampled from the population.

Increasing Spatial Dimension
(ConvTranspose in PyTorch)

To increase spatial dimension we use 4 times the desired number of output features.

$$L'_{\ell+1}[x, y, i] = \sigma\left(W[\Delta X, \Delta Y, J, i]\; L'_{\ell}[x + \Delta X, y + \Delta Y, J]\right)$$

We then convert $L'_{\ell+1}[X, Y, I]$ to $L'_{\ell+1}[2X, 2Y, I/4]$.

Generative Adversarial Networks (GANs)

Let y range over images. We have a generator p_Φ. For $i \in \{-1, 1\}$ we define a probability distribution over pairs $\langle i, y\rangle$ by

$$\begin{aligned} \tilde{p}_\Phi(i = 1) &= 1/2 \\ \tilde{p}_\Phi(y|i = 1) &= \mathrm{pop}(y) \\ \tilde{p}_\Phi(y|i = -1) &= p_\Phi(y) \end{aligned}$$

We also have a discriminator $P_\Psi(i|y)$ that tries to determine the source i given the image y.

The generator tries to fool the discriminator.

$$\Phi^* = \operatorname*{argmax}_{\Phi} \; \min_{\Psi} \; E_{\langle i, y\rangle \sim \tilde{p}_\Phi} \; - \ln P_\Psi(i|y)$$

GANs

The generator tries to fool the discriminator.

$$\Phi^* = \operatorname*{argmax}_{\Phi} \ \min_{\Psi} \ E_{\langle i, y\rangle \sim \tilde{p}_\Phi} \ -\ln P_\Psi(i|y)$$

Assuming universality of both the generator p_Φ and the discriminator P_Ψ we have $p_{\Phi^*} = \mathrm{pop}$.

Note that this involves only discrete cross-entropy.

GANs

To make the gradient descent clearer we write

$$E_{\langle i,y\rangle\sim\tilde{p}_\Phi}\; -\ln P_\Psi(i|y)$$

as

$$\frac{1}{2}E_{y\sim\mathrm{pop}}\; -\ln P_\Psi(1|y) \quad + \quad \frac{1}{2}E_z\; -\ln P_\Psi(-1|y_\Phi(z))$$

Generative Adversarial Nets

Goodfellow et al., June 2014

The rightmost column (yellow boarders) gives the nearest neighbor in the training data to the adjacent column.

GAN Mode Collapse

A major concern is "mode collapse" where the learned distribution omits a significant fraction of the population distribution.

There is no quantitative performance measure that provides a meaningful guarantee against mode collapse.

The Frénchet Inception Score (FID)

The main problem with GANs is the lack of a meaningful quantitative evaluation metric.

A standard quantitative performance measure is Frenchét Inception Distance (FID).

This measures statistics of the features of the inception image classification model (trained on imagenet) for images generated by the generator.

It then compares those statistics to the same statistics for images drawn from the population.

But the FID score provides no guarantees against mode collapse.

END