TTIC 31230, Fundamentals of Deep Learning

David McAllester, Winter 2020

Generative Adversarial Networks (GANs)

The Fundamental Equation for Continuous y

If y is continuous then the fundamental equation for estimating the distribution on y (cross entropy) involves continuous probability densities.

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

This occurs in unsupervised pretraining for sounds and images.

But differential entropy and differential cross-entropy are conceptually problematic.

Generative Adversarial Networks (GANs)

GANs avoid the differential cross-entropy loss function.

The model distribution $p_\Phi(y)$ is represented by a generator and the objective function involves a discriminator in place of cross-entropy.

Representing a Distribution with a Generator

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

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

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 reshape $L'_{\ell+1}[X, Y, I]$ 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 take gradients with respect to Φ 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.

Unsupervised Representation Learning ... (DC GANS)

Radford et al., Nov. 2015

Unsupervised Representation Learning ... (DC GANS)

Radford et al., Nov. 2015

Interpolated Faces

[Ayan Chakrabarti, January 2017]

Image-to-Image Translation (Pix2Pix)

Isola et al., Nov. 2016

We assume a corpus of “image translation pairs” such as images paired with semantic segmentations.

Conditional GANS

In the conditional case we have a population distribution over pairs $\langle x, y\rangle$. For conditional GANs we have a generator $p_\Phi(y|x)$ and a discriminator $P_\Psi(i|x,y)$ For $i \in \{-1, 1\}$ we define a probability distribution over triples $\langle x, y, i\rangle$ by

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

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

Adversarial Discrimination as an Additional Loss

$$\Phi^* = \operatorname*{argmin}_{\Phi} \; E_{(x,y)\sim \text{pop}} \; ||y - y_\Phi(x)||^2 \; + \; \lambda \, \mathcal{L}_{\text{Discr}}(\Phi)$$

$$\mathcal{L}_{\text{Discr}}(\Phi) = \max_{\Psi} \; E_{x,y,i\sim \tilde{p}_\Phi} \; \ln P_\Psi(i|y,x)$$

Discrimination as an Additional Loss

$$\text{L1}: \quad \Phi^* = \text{argmin}_\Phi \; E_{(x,y)\sim\text{pop}} \; ||y - y_\Phi(x)||_1$$

$$\text{cGAN}: \quad \Phi^* = \text{argmin}_\Phi \; \mathcal{L}_{\text{Discr}}(\Phi)$$

$$\text{L1} + \text{cGAN}: \quad \Phi^* = \text{argmin}_\Phi \; E_{(x,y)\sim\text{pop}} \; ||y - y_\Phi(x)||_1 \; + \; \lambda \, \mathcal{L}_{\text{Discr}}(\Phi)$$

Image-to-Image Translation (Pix2Pix)

Isola et al., Nov. 2016

Arial Photo to Map and Back

Unpaired Image-to-Image Translation (Cycle GANs)

Zhu et al., March 2017

We have two corpora of images, say images of zebras and unrelated images of horses, or photographs and unrelated paintings by Monet.

We want to construct translations between the two classes.

Cycle Gans

Cycle Gans

Unsupervised Machine Translation (UMT)

Lample et al, Oct. 2017, also Artetxe et al., Oct. 2017

In unsupervised machine translation the cycle loss is called **back-translation**.

Feature Alignment by Discrimination
Text to Speech (Saito et al. Sept. 2017)

Minimum Generation Error (MGE) uses perceptual distortion — a distance between the feature vector of the generated sound wave and the feature vector of the original.

Perceptual Naturalness can be enforced by a feature discrimination loss.

Adversarial Discriminative Domain Adaptation

Tzeng et al. Feb. 2017

A feature discrimination loss can be used to align source and target features.

Progressive GANs

Progressive Growing of GANs, Karras et al., Oct. 2017

Progressive GANs

Early GANs on ImageNet

BigGans

Large Scale GAN Training, Brock et al., Sept. 2018

This is a class-conditional GAN — it is conditioned on the imagenet class label.

This generates 512 X 512 images without using progressive training.

StyleGANs

A Style-Based Generator Architecture for Generative Adversarial Networks, Karras et al., Dec. 2018

StyleGans: Architecture

StyleGans: Style Transfer

StyleGans: Noise Variation

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.

GANs for Pretraining

A main motivation for distribution modeling is to provide pretrainined models tht can be used in downstream tasks.

This has proved very effective in natural language processing.

To date GANs have not proved useful for pretraining downstream applications.

END