TTIC 31230, Fundamentals of Deep Learning

David McAllester, Winter 2020

Contrastive Gans

GANs

The generator tries to fool the discriminator.

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

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

Contrastive GANs

A GAN can be built with a "contrastive" discriminator. Rather than estimate the probability that y is from the population, the discriminator must select which of $y_1, \ldots, y_N$ is from the population.

More formally, for $N \geq 2$ let $\tilde{p}_\Phi^{(N)}$ be the distribution defined by drawing one "positive" from pop and $N - 1$ IID negatives from p_Φ; then inserting the positive at a random position among the negatives; and returning $(i, y_1, \ldots, y_N)$ where i is the index of the positive.

Contrastive GANs

$$\Psi^*(\Phi) = \operatorname*{argmin}_{\Psi} \; E_{(i,y_1,\ldots,y_N)\sim\tilde{p}_\Phi^{(N)}} \; -\ln P_\Psi(i|y_1,\ldots,y_N)$$

$$\Phi^* = \operatorname*{argmax}_{\Phi} \; E_{(i,y_1,\ldots,y_N)\sim\tilde{p}_\Phi^{(N)}} \; -\ln P_{\Psi^*(\Phi)}(i|y_1,\ldots,y_N)$$

$\tilde{p}_\Phi^{(2)}(i|y_1,y_2)$ requires a choice between two y's while $\tilde{p}_\Phi(i|y)$ classifies a single y — these are different.

The discrimination gets more difficult as N gets larger.

Contrastive GANs

$$\Psi^*(\Phi) = \operatorname*{argmin}_{\Psi} E_{(i,y_1,\ldots,y_N)\sim\tilde{p}_{\Phi}^{(N)}} - \ln P_{\Psi}(i|y_1,\ldots,y_N)$$

$$\Phi^* = \operatorname*{argmax}_{\Phi} E_{(i,y_1,\ldots,y_N)\sim\tilde{p}_{\Phi}^{(N)}} - \ln P_{\Psi^*(\Phi)}(i|y_1,\ldots,y_N)$$

Assuming universality

$$\mathcal{L}(\Psi^*(\Phi)) = H_{\Phi}(i|y_1,\ldots y_N)$$

$$p_{\Phi^*} = \mathrm{pop} \qquad H_{\Phi^*}(i|y_1,\ldots,y_N) = \ln N$$

A Theorem for Universal Ψ

$$E_{(i,y_1,\ldots,y_N)\sim\tilde{p}_\Phi^{(N)}} \;\; -\ln \tilde{p}_\Phi^{(N)}(i|y_1,\ldots,y_N)$$

$$\geq \ln N - \frac{N-1}{N}(KL(\mathrm{pop},p_\Phi) + KL(p_\Phi,\mathrm{pop}))$$

Note that the bound holds with equality for $p_\Phi = \mathrm{pop}$.

This is analogous to the JSD expression for the optimal discriminator.

Proof Part A.

$$
\begin{aligned}
& E_{(i,y_1,\ldots,y_N)\sim\tilde{p}_\Phi^{(N)}} \;\ln p_{\Psi^*}(i|y_1,\ldots,y_N) \\
= \;& E_{(i,y_1,\ldots,y_N)\sim\tilde{p}_\Phi^{(N)}} \;\ln\left(\operatorname*{softmax}_i \;\ln\frac{\mathrm{pop}(y_i)}{p_\Phi(y_i)}\right)[i] \\
= \;& E_{(i,y_1,\ldots,y_N)\sim\tilde{p}_\Phi^{(N)}} \;\ln\frac{\mathrm{pop}(y_i)}{p_\Phi(y_i)} - \ln\left(\sum_j \frac{\mathrm{pop}(y_j)}{p_\Phi(y_j)}\right) \\
= \;& E_{(i,y_1,\ldots,y_N)\sim\tilde{p}_\Phi^{(N)}} \;\ln\frac{\mathrm{pop}(y_i)}{p_\Phi(y_i)} - \ln\left(\frac{1}{N}\sum_j \frac{\mathrm{pop}(y_j)}{p_\Phi(y_j)}\right) - \ln\;N
\end{aligned}
$$

Proof Part B.

$$E_{(i,y_1,\ldots,y_N)\sim\tilde{p}_\Phi^{(N)}} \;\ln \frac{\mathrm{pop}(y_i)}{p_\Phi(y_i)} - \ln\left(\frac{1}{N}\sum_j \frac{\mathrm{pop}(y_j)}{p_\Phi(y_j)}\right) - \ln\; N$$

$$\leq\; E_{(i,y_1,\ldots,y_N)\sim\tilde{p}_\Phi^{(N)}} \;\ln \frac{\mathrm{pop}(y_i)}{p_\Phi(y_i)} - \frac{1}{N}\sum_j \ln \frac{\mathrm{pop}(y_j)}{p_\Phi(y_j)} - \ln\; N$$

$$=\; E_{y\sim\mathrm{pop}} \ln \frac{\mathrm{pop}(y)}{p_\Phi(y)} - E_{(i,y_1,\ldots,y_N)\sim\tilde{p}_\Phi^{(N)}} \frac{1}{N}\sum_j \ln \frac{\mathrm{pop}(y_j)}{p_\Phi(y_j)} - \ln N$$

$$=\; \frac{N-1}{N}(KL(\mathrm{pop}, p_\Phi) + KL(p_\Phi, \mathrm{pop})) - \ln\; N$$

END