

# 1 AsynDGAN learns the correct distribution

In this section we present an analysis of AsynDGAN and discuss the implications of the results. We show that the AsynDGAN is able to aggregates multiple separated data set and learn generative distribution in an *all important* fashion. We first begin with a technical lemma describing the optimal strategy of the discriminator.

**Lemma 1.** *When generator  $G$  is fixed, the optimal discriminator  $D_j(y, x)$  is :*

$$D_j(y, x) = \frac{p(y|x)}{p(y|x) + q(y|x)} \quad (1)$$

**Proof:**

$$\begin{aligned} \max_D V(D) &= \max_{D_1 \sim D_N} \sum \pi_j \int_x s_j(x) \int_y p(y|x) \log D_j(y, x) + q(y|x) \log(1 - D_j(y, x)) dy dx \\ &\leq \sum \pi_j \int_x s_j(x) \int_y \max_{D_j} \{p(y|x) \log D_j(y, x) + q(y|x) \log(1 - D_j(y, x))\} dy dx \end{aligned}$$

by setting  $D_j(y, x) = \frac{p(y|x)}{p(y|x) + q(y|x)}$  we can maximize each component in the integral thus make the inequality hold with equality.  $\square$

Suppose in each training step the discriminator achieves its maxima criterion in Lemma 1, the loss function for the generator becomes:

$$\begin{aligned} \min_G V(G) &= \mathbb{E}_y \mathbb{E}_{x \sim p_{data}(x|y)} [\log D(y, x)] + \mathbb{E}_{z \sim p_{\hat{y}}(\hat{y}|x)} [\log(1 - D(z, y))] \\ &= \sum_{j \in [N]} \pi_j \int_x s_j(x) \int_y \underbrace{p(y|x) \log \frac{p(y|x)}{p(y|x) + q(y|x)} + q(y|x) \log \frac{q(y|x)}{p(y|x) + q(y|x)}}_{\text{To be analyzed in Lemma 2}} dy dx \end{aligned}$$

**Lemma 2.** *Let  $a(y)$  and  $b(y)$  be two probability distributions s.t. support  $\Omega(a) \subset \Omega(b)$ , the loss function  $L(a) = \int_y a(y) \log \frac{a(y)}{a(y) + b(y)} + b(y) \log \frac{b(y)}{a(y) + b(y)} dy \geq -2 \log 2$ .*

*The inequality holds iff  $b(y) = a(y)$ .*

**Proof:**

Let  $\lambda$  be the Lagrangian multiplier.

$$L(a, \lambda) = \int_y a(y) \log \frac{a(y)}{a(y) + b(y)} + b(y) \log \frac{b(y)}{a(y) + b(y)} + \lambda a(y) dy - \lambda \quad (2)$$

By setting  $\frac{\partial L}{\partial a} = 0$  we have  $\log \frac{a(y)}{a(y) + b(y)} = \lambda$  holds for all  $x$ . The fact that  $\log \frac{a(y)}{a(y) + b(y)}$  is a constant enforces  $a(y) = b(y)$ . Plugging  $a(y) = b(y)$  into loss function we have  $L(a)_{|a(y)=b(y)} = -2 \log(2)$ .

Now we are ready to prove our main result that AsynDGAN learns the correct distribution. Assuming in each step, the discriminator always perform optimally, we show

indeed the generative distribution  $G$  seeks to minimize the loss by approximating underlying generative distribution of data.

**Theorem 1.** *Suppose the discriminators  $D_{1 \sim N}$  always behaves optimally (denoted as  $D_{1 \sim N}^*$ ), the loss function of generator is global optimal iff  $q(y, x) = p(y, x)$  where the optimal value of  $V(G, D_{1 \sim N}^*)$  is  $-\log 4$ .*

**Proof:**

$$\begin{aligned} & \min_{\substack{q(y,x) > 0, \\ \int_y q(y,x) = s(x)}} \sum_{j \in [N]} \pi_j \int_x s_j(x) \int_y p(y|x) \log \frac{p(y|x)}{p(y|x) + q(y|x)} + q(y|x) \log \frac{q(y|x)}{p(y|x) + q(y|x)} dy dx \\ & \geq \sum_{j \in [N]} \pi_j \int_x s_j(x) \min_{\substack{q(y|x) > 0, \\ \int_y q(y|x) = 1}} \int_y p(y|x) \log \frac{p(y|x)}{p(y|x) + q(y|x)} + q(y|x) \log \frac{q(y|x)}{p(y|x) + q(y|x)} dy dx \end{aligned}$$

By Lemma 2, the optimal condition for minimizing:

$$\min_{\substack{q(y|x) > 0, \\ \int_y q(y|x) = 1}} \int_y p(y|x) \log \frac{p(y|x)}{p(y|x) + q(y|x)} + q(y|x) \log \frac{q(y|x)}{p(y|x) + q(y|x)} dy dx$$

is by setting  $q(y|x) = p(y|x), \forall x$ . Such choice of  $q(y|x)$  makes the inequality holds as an equality. Meanwhile  $q(y|x) = p(y|x)$  implies  $q(y, x) = p(y, x)$  given the fact that  $p, q$  has the same marginal distribution on  $y$ . By plugging in  $q(y|x) = p(y|x)$  we can derive the optimal value of  $V(G, D_{1 \sim N}^*)$  to be  $-\log 4$ .

**Remark 1.** While analysis of AsynDGAN loss shares similar spirit with [1], it has different implications. In the distributed learning setting, data from different nodes are often dissimilar. Consider the case where  $\Omega(s_j(x)) \cap \Omega(s_k(x)) = \emptyset, \text{ for } k \neq j$ , the information for  $p(y|x), x \in \Omega(s_j(x))$  will be missing if we lose the  $j$ -th node. The behavior of trained generative model is unpredictable when receiving auxiliary variables from unobserved distribution  $s_j(x)$ . The AsynDGAN framework provides a solution for unifying different datasets by collaborating multiple discriminators thus can aggregate separated datasets in an *all important* fashion.

## References

- [1] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.