Appendix: Balanced Self-Paced Learning for Generative Adversarial Clustering Network

Kamran Ghasedi Dizaji[1], Xiaoqian Wang[1], Cheng Deng[2], and Heng Huang[1,3]

[1]Electrical and Computer Engineering Department, University of Pittsburgh, PA, USA
[2]School of Electronic Engineering, Xidian University, Xi'an, Shaanxi, China
[3]JD Digits
kamran.ghasedi@gmail.com, xiaoqian.wang@pitt.edu, chdeng.xd@gmail.com, heng.huang@pitt.edu

The objective function of the adversarial game for *ClusterGAN* is:

$$\min_{\mathcal{G},\mathcal{C}} \max_{\mathcal{D}} \; \mathbf{U}(\mathcal{D},\mathcal{G},\mathcal{C}) = \mathbb{E}_{\mathbf{x}\sim P(\mathbf{x})}\big[\log \mathcal{D}\big(\mathcal{C}(\mathbf{x}),\mathbf{x}\big)\big] + \mathbb{E}_{\mathbf{z}\sim P(\mathbf{z})}\big[\log\big(1-\mathcal{D}\big(\mathbf{z},\mathcal{G}(\mathbf{z})\big)\big)\big]. \quad (1)$$

Lemma 1. *For any fixed $\mathcal{G}$ and $\mathcal{C}$, the optimal $\mathcal{D}$ defined by the utility function $\mathbf{U}(\mathcal{D},\mathcal{G},\mathcal{C})$ is:*

$$\mathcal{D}^*(\mathbf{z},\mathbf{x}) = \frac{P(\mathbf{x})P_{\mathcal{C}}(\mathbf{z}|\mathbf{x})}{P(\mathbf{x})P_{\mathcal{C}}(\mathbf{z}|\mathbf{x}) + P(\mathbf{z})P_{\mathcal{G}}(\mathbf{x}|\mathbf{z})} = \frac{P_{\mathcal{C}}(\mathbf{z},\mathbf{x})}{P_{\mathcal{C}}(\mathbf{z},\mathbf{x}) + P_{\mathcal{G}}(\mathbf{z},\mathbf{x})}$$

Proof. Given the clusterer and generator, the utility function $\mathbf{U}(\mathcal{D},\mathcal{G},\mathcal{C})$ can be rewritten as

$$\begin{aligned}\mathbf{U}(\mathcal{D},\mathcal{G},\mathcal{C}) =& \iint P(\mathbf{x})P_{\mathcal{C}}(\mathbf{z}|\mathbf{x})\log(\mathcal{D}(\mathbf{z},\mathbf{x}))d\mathbf{x}d\mathbf{z} \quad (2)\\ &+ \iint P(\mathbf{z})P_{\mathcal{G}}(\mathbf{x}|\mathbf{z})\log(1-\mathcal{D}(\mathbf{z},\mathbf{x}))d\mathbf{x}d\mathbf{z}\\ =& \iint P_{\mathcal{C}}(\mathbf{z},\mathbf{x})\log(\mathcal{D}(\mathbf{z},\mathbf{x}))d\mathbf{x}d\mathbf{z}\\ &+ \iint P_{\mathcal{G}}(\mathbf{z},\mathbf{x})\log(1-\mathcal{D}(\mathbf{z},\mathbf{x}))d\mathbf{x}d\mathbf{z}\\ =& f(\mathcal{D}(\mathbf{z},\mathbf{x}))\end{aligned}$$

For any $\big(P_{\mathcal{C}}(\mathbf{z},\mathbf{x}), P_{\mathcal{G}}(\mathbf{z},\mathbf{x})\big) \in \mathbb{R}^2 \; \{0,0\}$, the function $f(\mathcal{D}(\mathbf{z},\mathbf{x}))$ achieves its maximum at $\frac{P_{\mathcal{C}}(\mathbf{z},\mathbf{x})}{P_{\mathcal{C}}(\mathbf{z},\mathbf{x})+P_{\mathcal{G}}(\mathbf{z},\mathbf{x})}$. □

Given $\mathcal{D}^*(\mathbf{x},\mathbf{z})$, we can further replace $\mathcal{D}$ in the utility function $\mathbf{U}(\mathcal{D},\mathcal{G},\mathcal{C})$ and reformulate the objective as $\mathbf{V}(\mathcal{G},\mathcal{C}) = \max_{\mathcal{D}} \mathbf{U}(\mathcal{D},\mathcal{G},\mathcal{C})$.

Lemma 2. *The global optimum point of $\mathbf{V}(\mathcal{G},\mathcal{C})$ is achieved if and only if $P(\mathbf{z},\hat{\mathbf{x}}) = P(\hat{\mathbf{z}},\mathbf{x})$.*

Proof. Given $\mathcal{D}^*(\mathbf{x},\mathbf{z})$, the utility function $\mathbf{V}(\mathcal{G},\mathcal{C})$ can be reformulated as:

$$\mathbf{V}(\mathcal{G},\mathcal{C}) = \iint P_{\mathcal{C}}(\mathbf{z},\mathbf{x})\log\Big(\frac{P_{\mathcal{C}}(\mathbf{z},\mathbf{x})}{P_{\mathcal{C}}(\mathbf{z},\mathbf{x})+P_{\mathcal{G}}(\mathbf{z},\mathbf{x})}\Big)d\mathbf{x}d\mathbf{z} + \iint P_{\mathcal{G}}(\mathbf{z},\mathbf{x})\log\Big(\frac{P_{\mathcal{G}}(\mathbf{z},\mathbf{x})}{P_{\mathcal{C}}(\mathbf{z},\mathbf{x})+P_{\mathcal{G}}(\mathbf{z},\mathbf{x})}\Big)d\mathbf{x}d\mathbf{z} \quad (3)$$

Sketching the proof in original *GAN* paper [1], $\mathbf{V}(\mathcal{G},\mathcal{C})$ cab be rewritten as:

$$\mathbf{V}(\mathcal{G},\mathcal{C}) = -\log 4 + 2JSD(P_{\mathcal{C}}(\mathbf{z},\mathbf{x}) \| P_{\mathcal{G}}(\mathbf{z},\mathbf{x})), \quad (4)$$

where JSD represents the Jensen-Shannon divergence, which is always non-negative. Therefore, the unique optimum of $\mathbf{V}(\mathcal{G},\mathcal{C})$ is achieved if and only if $P_{\mathcal{C}}(\mathbf{z},\mathbf{x}) = P_{\mathcal{G}}(\mathbf{z},\mathbf{x})$, or in other words

$$P(\mathbf{z},\hat{\mathbf{x}}) = P(\hat{\mathbf{z}},\mathbf{x})$$

□

The optimization problem for estimating our balanced self-paced learning algorithm is:

$$\min_{\boldsymbol{\nu}} \; \mathbf{L}(\boldsymbol{\nu}) = \sum_{i=1}^{n} \nu_i l_i - \lambda_\nu \|\boldsymbol{\nu}\|_1 + \gamma \|\boldsymbol{\nu}\|_e \;\; s.t. \; \boldsymbol{\nu} \in [0,1]^n. \quad (5)$$

Theorem 1. *For any fixed $\mathcal{C}$, the optimal $\boldsymbol{\nu}$ defined by the objective function $\mathbf{L}(\boldsymbol{\nu})$ is:*

$$\begin{cases} \nu^*_{kq} = 1, & if \;\; l_{kq} < \lambda_\nu - 2\gamma q \\ \nu^*_{kq} = \frac{\lambda_\nu - l_{kq}}{2\gamma} - q, & if \;\; \lambda_\nu - 2\gamma q \le l_{kq} < \lambda_\nu - 2\gamma(q-1) \\ \nu^*_{kq} = 0, & if \;\; l_{kq} \ge \lambda_\nu - 2\gamma(q-1) \end{cases}$$

where $q \in \{1, ..., n_k\}$ is the sorted index of loss values $\{l_{k1}, ..., l_{kn_k}\}$ in the k-th group.

Proof.

$$\min_{\nu} \quad \mathbf{L}(\boldsymbol{\nu}) = \sum_{k=1}^{c} \mathbf{L}(\boldsymbol{\nu}_k) \tag{6}$$

$$= \sum_{k=1}^{c} \Big[\sum_{i=1}^{n_k} \nu_{ki}(l_{ki} - \lambda_\nu) + \gamma \Big(\sum_{i=1}^{n_k} |\nu_{ki}| \Big)^2 \Big], \quad s.t. \ \boldsymbol{\nu} \in [0,1]^n ,$$

We can handle the c groups in Problem (6) separately. Given k, define $\mathbf{b} = [\frac{(l_{k1}-\lambda_\nu)}{\gamma}, \frac{(l_{k2}-\lambda_\nu)}{\gamma}, \ldots, \frac{(l_{kn_k}-\lambda_\nu)}{\gamma}]$, the optimization problem *w.r.t.* the k-th group can be formulated as follows:

$$\min_{\mathbf{u}} \mathbf{b}^T\mathbf{u} + \mathbf{u}^T\mathbf{1}\mathbf{1}^T\mathbf{u}, \ \ s.t. \ \mathbf{0} \leq \mathbf{u} \leq \mathbf{1} , \tag{7}$$

where $\mathbf{u} = [v_{k1}, v_{k2}, \ldots v_{kn_k}]$. The Lagrangian function of Problem (7) is

$$\min_{\mathbf{u}} \mathbf{b}^T\mathbf{u} + \mathbf{u}^T\mathbf{1}\mathbf{1}^T\mathbf{u} - \eta^T\mathbf{u} - \lambda^T(\mathbf{1} - \mathbf{u}) . \tag{8}$$

where $\eta \geq \mathbf{0}$ and $\lambda \geq \mathbf{0}$ are Lagrangian multipliers. Take derivate of Problem (8) *w.r.t.* $\mathbf{u}$ and set it to zero, we get

$$\eta + \lambda - \mathbf{b} = 2m\mathbf{1} . \tag{9}$$

where $m = \mathbf{1}^T\mathbf{u}$. From the KKT condition we can derive $\eta^T\mathbf{u} = 0$ and $\lambda^T(\mathbf{1} - \mathbf{u}) = 0$. Consequently, we can derive

$$\begin{cases} u_q = 0 & \Longrightarrow \eta_q > 0, \lambda_q = 0 \quad \Longrightarrow \frac{b_q}{2} + m > 0 , \\ 0 < u_q < 1 & \Longrightarrow \eta_q = 0, \lambda_q = 0 \quad \Longrightarrow \frac{b_q}{2} + m = 0 , \\ u_q = 1 & \Longrightarrow \eta_q = 0, \lambda_q > 0 \quad \Longrightarrow \frac{b_q}{2} + m < 0 , \end{cases} \tag{10}$$

where $q \in \{1, ..., n_k\}$. Without loss of generality, suppose $\mathbf{b}$ is a sorted vector such that $b_1 < b_2 < \cdots < b_{n_k}$, then according to Eq. (10) we have $1 \geq u_1 \geq u_2 \geq \cdots \geq u_{n_k} \geq 0$, from which we can derive

$$\begin{cases} u_q = 0 & \Longrightarrow u_r = 0, \forall r \geq q & \Longrightarrow m \leq q - 1 , \\ 0 < u_q < 1 & \Longrightarrow u_r = 0, \forall r > q, \text{and } u_r = 1, \forall r < q & \Longrightarrow q - 1 < m < q , \\ u_q = 1 & \Longrightarrow u_r = 1, \forall r \leq q & \Longrightarrow m \geq q . \end{cases} \tag{11}$$

Combining Eq. (10) and Eq. (11) we can derive the solution to Problem (7) as follows:

$$\begin{cases} -\frac{b_q}{2} \leq q - 1 & \Longrightarrow u_q = 0 , \\ q - 1 < -\frac{b_q}{2} < q & \Longrightarrow u_q = -\frac{b_q}{2} - q + 1 , \\ -\frac{b_q}{2} \geq q & \Longrightarrow u_q = 1 , \end{cases}$$

which can be rewritten based on $\boldsymbol{\nu}$ as:

$$\begin{cases} \nu_{kq}^* = 1, & if \quad l_{kq} < \lambda_\nu - 2\gamma q \\ \nu_{kq}^* = \frac{\lambda_\nu - l_{kq}}{2\gamma} - q, & if \quad \lambda_\nu - 2\gamma q \leq l_{kq} < \lambda_\nu - 2\gamma(q-1) \\ \nu_{kq}^* = 0, & if \quad l_{kq} \geq \lambda_\nu - 2\gamma(q-1) \end{cases}$$

□

References

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Advances in neural information processing systems (NIPS)*, pages 2672–2680, 2014. 1