1 Group DRO

In group DRO, assuming the training distribution $\hat{P}$ is a mix of group distributions $\hat{P}_g$ for m groups $g \in \{1, 2, ..., m\}$. The intuitive DRO objective is

$$\hat{\theta}_{\text{DRO}} = \arg\min_{\theta \in \Theta} \sup_{g \in G} \mathbb{E}_{x,y \sim \hat{P}_g}[\ell(x, y, \theta)]$$

The same problem can be rewritten as

$$\hat{\theta}_{\text{DRO}} = \arg\min_{\theta \in \Theta} \sup_{q \in \Delta m} \sum_{g=1}^{m} q_g \mathbb{E}_{x,y \sim \hat{P}_g}[\ell(x, y, \theta)]$$

where q can be viewed as a weight distribution for each group distribution. In realization, q will have 1 on the group with the worst group loss which simulates the original formulation of the problem. Otherwise, the supremum will not be satisfied.

2 Noisy Group Attributes

Now, we are interested in problems where our empirical group distribution is noisy. Consider an empirical distribution $\hat{P}$ made of a mix of two true group distributions $\hat{P}_1$ and $\hat{P}_2$. However, we don't have access to these two true group distributions. Instead, we have access to $\tilde{P}_1$ and $\tilde{P}_2$ where some samples in $\tilde{P}_1$ and $\tilde{P}_2$ are misclassified. Suppose the error rate for the group information is ϵ_1 and ϵ_2. In other words, samples that are truly in group 1 can be categorized into group 2 with probability of ϵ_1 and samples that are truly in group 2 can be categorized into group 1 with probability of ϵ_2. Thus, we can represent

$$\tilde{P}_1 = \frac{\mathbb{P}[G=1] \cdot (1-\epsilon_1)}{\mathbb{P}[G=1] \cdot (1-\epsilon_1) + \mathbb{P}[G=2] \cdot \epsilon_2} \hat{P}_1 + \frac{\mathbb{P}[G=2] \cdot \epsilon_2}{\mathbb{P}[G=1] \cdot (1-\epsilon_1) + \mathbb{P}[G=2] \cdot \epsilon_2} \hat{P}_2$$
$$\tilde{P}_2 = \frac{\mathbb{P}[G=1] \cdot \epsilon_1}{\mathbb{P}[G=1] \cdot \epsilon_1 + \mathbb{P}[G=2] \cdot (1-\epsilon_2)} \hat{P}_1 + \frac{\mathbb{P}[G=2] \cdot (1-\epsilon_2)}{\mathbb{P}[G=1] \cdot \epsilon_1 + \mathbb{P}[G=2] \cdot (1-\epsilon_2)} \hat{P}_2$$

For notation simplicity, we let

$$\alpha_1 = \frac{\mathbb{P}[G=1] \cdot (1-\epsilon_1)}{\mathbb{P}[G=1] \cdot (1-\epsilon_1) + \mathbb{P}[G=2] \cdot \epsilon_2}$$
$$\alpha_2 = \frac{\mathbb{P}[G=1] \cdot \epsilon_1}{\mathbb{P}[G=1] \cdot \epsilon_1 + \mathbb{P}[G=2] \cdot (1-\epsilon_2)}$$

Then,

$$\tilde{P}_1 = \alpha_1 \cdot \hat{P}_1 + (1-\alpha_1) \cdot \hat{P}_2$$
$$\tilde{P}_2 = \alpha_2 \cdot \hat{P}_1 + (1-\alpha_2) \cdot \hat{P}_2$$

In the context of Group DRO, with access to the true group distributions, we want to optimize

$$\hat{\theta}_{\text{DRO}} = \arg\min_{\theta \in \Theta} \sup_{q \in \Delta m} q_1 \mathbb{E}_{x,y \sim \hat{P}_1}[\ell(x, y, \theta)] + q_2 \mathbb{E}_{x,y \sim \hat{P}_2}[\ell(x, y, \theta)]$$

If we naively use Group DRO on the noisy group distributions $\tilde{P}_1, \tilde{P}_2$, we are actually solving

$$\begin{aligned}
\tilde{\theta}_{\text{DRO}} &= \arg\min_{\theta \in \Theta} \sup_{\tilde{q} \in \Delta m} \tilde{q}_1 \mathbb{E}_{x,y \sim \tilde{P}_1}[\ell(x, y, \theta)] + \tilde{q}_2 \mathbb{E}_{x,y \sim \tilde{P}_2}[\ell(x, y, \theta)] \\
&= \arg\min_{\theta \in \Theta} \sup_{\tilde{q} \in \Delta m} (\tilde{q}_1 \alpha_1 + \tilde{q}_2 \alpha_2) \mathbb{E}_{x,y \sim \hat{P}_1}[\ell(x, y, \theta)] + (\tilde{q}_1 (1-\alpha_1) + \tilde{q}_2 (1-\alpha_2)) \mathbb{E}_{x,y \sim \hat{P}_2}[\ell(x, y, \theta)]
\end{aligned}$$

Comment: This above equation seems to make sense but also not really. Ignoring the minimization problem outside, if we focus on the supremum inside, we see that the first one should achieve a $\tilde{q}^*$ s.t. a weight of 1 is put on the maximum between $\mathbb{E}_{x,y\sim\tilde{P}_1}[\ell(x,y,\theta)]$ and $\mathbb{E}_{x,y\sim\tilde{P}_2}[\ell(x,y,\theta)]$. However, taking a look at the second equation, the ideal solution for the supremum should achieve $\tilde{q}^*$ s.t. either $(\tilde{q_1}\alpha_1 + \tilde{q_2}\alpha_2)$ or $(\tilde{q_1}(1-\alpha_1) + \tilde{q_2}(1-\alpha_2))$ achieves is 1 depending on the maximum between $\mathbb{E}_{x,y\sim\hat{P}_1}[\ell(x,y,\theta)]$ and $\mathbb{E}_{x,y\sim\hat{P}_2}[\ell(x,y,\theta)]$

We can make some observations

1. For a given $\tilde{q}$ when optimizing for $\tilde{\theta}_{\text{DRO}}$ using the noisy group distributions, we are actually using a transformed q in optimizing with non-noisy group distributions. For instance, if we at some point $\tilde{q_1} = t$ and $\tilde{q_2} = 1-t$ in the context of noisy group distributions, it's actually representing $q_1 = t\cdot\alpha_1 + (1-t)\cdot\alpha_2$ and $q_2 = t\cdot(1-\alpha_1)\cdot(1-t)\cdot(1-\alpha_2)$ in optimizing with true group distributions.
2. With the transformation, searching in the space of Δm for q and $\tilde{q}$ are different because $\tilde{q}$ in Δm does not correspond to the same space Δm for q.
3. The transformation can be written as
$$\begin{bmatrix} q_1 \\ q_2 \end{bmatrix} = \begin{bmatrix} \alpha_1 & \alpha_2 \\ 1-\alpha_1 & 1-\alpha_2 \end{bmatrix} \begin{bmatrix} \tilde{q_1} \\ \tilde{q_2} \end{bmatrix}$$
So if we know Δm contains the q we are interested in, then
$$\tilde{\Delta m} = \begin{bmatrix} \alpha_1 & \alpha_2 \\ 1-\alpha_1 & 1-\alpha_2 \end{bmatrix}^{-1} \Delta m$$
would contain the corresponding $\tilde{q}$.
4. When $\epsilon_1 = \epsilon_2 = 0$, the objective is the same for original group DRO with the true group distributions because $\alpha_1 = 1$ and $\alpha_2 = 0$ form an identity matrix.

3 Online Group DRO

Algorithm 1: Online optimization algorithm for group DRO

Input: Step sizes η_q, η_θ; P_g for each $g \in \mathcal{G}$
Initialize $\theta^{(0)}$ and $q^{(0)}$
for $t = 1, \ldots, T$ **do**
  $g \sim \text{Uniform}(1, \ldots, m)$ `// Choose a group g at random`
  $x, y \sim P_g$ `// Sample x,y from group g`
  $q' \leftarrow q^{(t-1)}$; $q'_g \leftarrow q'_g \exp(\eta_q \ell(\theta^{(t-1)}; (x,y)))$ `// Update weights for group g`
  $q^{(t)} \leftarrow q' / \sum_{g'} q'_{g'}$ `// Renormalize q`
  $\theta^{(t)} \leftarrow \theta^{(t-1)} - \eta_\theta q_g^{(t)} \nabla \ell(\theta^{(t-1)}; (x,y))$ `// Use q to update θ`
end

4 Bound error rate

5 References