Problem Reduction

Initial Problem

$$\min_{\theta,\alpha_A,\alpha_P,\{\beta_i\},\{\beta'_j\}} \alpha_A r_A + \alpha_P r_P + \frac{1}{n_A}\sum_{i\in[n_\alpha]}\beta_i + \frac{1}{n_P}\sum_{j\in[n_P]}\beta'_j$$

$$\text{s.t. } \max_{y\in\{\pm1\}}\sup_{(x,a)}\ell(\theta,(x,a,y)) - \alpha_A d^i_A(x,a) - \alpha_P d^j_P(x,y) \le \beta_i + \beta'_j \qquad \forall i,j$$

Constraint Reduction

1. Focus on the constraint because there is a sup term that makes it hard to optimize (analyze $y = +1$ for simplicity)

$$\sup_{(x,a)}\ell(\theta,(x,a,y)) - \alpha_A d^i_A(x,a) - \alpha_P d^j_P(x,y)$$

2. Focus on the term depending on (x,a) and ignoring dependence on y

$$= \kappa_P\alpha_P|y^P_j - y| + \left(\sup_{(x,a)} h(y\theta,(x,a)) - \alpha_A||x^A_i - x||_p - \alpha_P||x^P_j - x||_p - \alpha_A\kappa_A||a^A_i - a||_{p'}\right)$$

3. Upperbound the new sup term inside

$$\sup_{(x,a)} h(y\theta,(x,a)) - \alpha_A||x^A_i - x||_p - \alpha_P||x^P_j - x||_p - \alpha_A\kappa_A||a^A_i - a||_{p'}$$
$$\le f\left(\frac{\min(\alpha_A,\alpha_P)||\theta_1||_*||x^A_i - x^P_j||}{\alpha_A+\alpha_P} + \frac{\langle y\theta_1, \alpha_A x^A_i + \alpha_P x^P_j\rangle}{\alpha_A+\alpha_P} + \langle y\theta_2, a^A_i\rangle\right) - \min(\alpha_A,\alpha_P)||x^A_i - x^P_j||_p$$

required

$$||\theta_1||_{p,*} \le \alpha_A + \alpha_P \textit{ and } ||\theta_2||_{p',*} \le \kappa_A\alpha_A$$

Otherwise, the sup term evaluates to ∞. *They also showed the gap is not signifiant*

4. They also showed an approximation upperbound with $f(\langle y\theta_1,(\hat{x}_{i,j},a^A_i)\rangle)$

$$\hat{x}_{i,j} = \begin{cases} x^P_j & \text{if } \alpha_A < \alpha_P \\ x^A_i & \end{cases}$$

$$f\left(\frac{\min(\alpha_A,\alpha_P)||\theta_1||_*||x^A_i - x^P_j||}{\alpha_A+\alpha_P} + \frac{\langle y\theta_1, \alpha_A x^A_i + \alpha_P x^P_j\rangle}{\alpha_A+\alpha_P} + \langle y\theta_2, a^A_i\rangle\right) - f(\langle y\theta_1,(\hat{x}_{i,j},a^A_i)\rangle) \le 2\hat{\alpha}||x^A_i - x^P_j||$$

5. Use the bounds

$$\left(\max_{y\in\{\pm1\}}\sup_{(x,a)}\ell(\theta,(x,a,y)) - \alpha_A d^i_A(x,a) - \alpha_P d^j_P(x,y)\right) + \left(\hat{\alpha}||x^A_i - x^P_j||_p\right)$$
$$- f(\langle y^P_j\theta,(\hat{x}_{i,j},a^A_i)\rangle) - \max(y^P_j\langle\theta,(\hat{x}_{i,j},a^A_i)\rangle - \alpha_P\kappa_P, 0) \le 2\hat{\alpha}||x^A_i - x^P_j||_p$$

6. To formulate it better to replace the max/sup term

$$\left(\max_{y\in\{\pm1\}}\sup_{(x,a)}\ell(\theta,(x,a,y)) - \alpha_A d^i_A(x,a) - \alpha_P d^j_P(x,y)\right)$$
$$\le \left(f(\langle y^P_j\theta,(\hat{x}_{i,j},a^A_i)\rangle) + \max(y^P_j\langle\theta,(\hat{x}_{i,j},a^A_i)\rangle - \alpha_P\kappa_P, 0) - \hat{\alpha}||x^A_i - x^P_j||_p\right) - 2\hat{\alpha}||x^A_i - x^P_j||_p$$

7. We can replace the max/sup term and re-formulate the problem

$$\min_{\theta,\alpha_A,\alpha_P,\{\beta_i\},\{\beta'_j\}} \alpha_A r_A + \alpha_P r_P + \frac{1}{n_A}\sum_{i\in[n_\alpha]}\beta_i + \frac{1}{n_P}\sum_{j\in[n_P]}\beta'_j$$

$$\text{s.t.}\ \ \big(f(\langle y_j^P\theta,(\hat{x}_{i,j},a_i^A)\rangle) + \max(y_j^P\langle\theta,(\hat{x}_{i,j},a_i^A)\rangle - \alpha_P\kappa_P,0) - \hat{\alpha}||x_i^A - x_j^P||_p\big) \le \beta_i + \beta'_j \qquad \forall i,j$$

Optimization on Reduced Problem

1. Approximation bounds reduces to solve this problem

$$\min_{\alpha_A\alpha_P\theta_1\theta_2}(\alpha_A r_A + \alpha_P r_P) + \frac{1}{n_A n_P}\sum_{(i,j)\in M}(f(y_j^P\langle\theta,(\hat{x}_{i,j},a_i^A)\rangle) + \max(y_j^P\langle\theta,\hat{x}_{i,j},a_i^A\rangle - \alpha_P\kappa_P,0) - \hat{\alpha}||x_i^A - x_j^P||)$$

s.t. $||\theta_1||_* \le \alpha_A + \alpha_P, ||\theta_2||_* \le \kappa_A\alpha_A$

2. Projected gradient descent on both cases $\alpha_A < \alpha_P$ and the other way around When $\alpha_A < \alpha_P$, $\hat{x}_{i,j} = x_j^P$ and $\hat{\alpha} = \alpha_A$

$$\begin{aligned}&\Omega^A(\alpha_A,\alpha_P,\theta)\\ &= \min_{\alpha_A\alpha_P\theta_1\theta_2}(\alpha_A r_A + \alpha_P r_P) + \frac{1}{n_A n_P}\sum_{(i,j)\in M}(f(y_j^P\langle\theta,(x_j^P,a_i^A)\rangle) + \max(y_j^P\langle\theta,(x_j^P,a_i^A)\rangle - \alpha_P\kappa_P,0) - \alpha_A||x_i^A - x_j^P||)\end{aligned}$$

3. The paper derives projected gradient descent

$$\begin{gathered}||\theta_1||_* \le \alpha_A + \alpha_P\\ ||\theta_2||_* \le \kappa_A\alpha_A\\ \alpha_A \le \alpha_P \text{ or } \alpha_P \le \alpha_A\end{gathered}$$

4. The projection is solved by a Lagrangian

Constrained Optimization

1. Lagrangian
2. Projected Gradient Descent (In the paper)
3. Reparameterization