

$$\mathcal{L}(\theta) = \mathbb{E}_{(x,y) \sim \mathcal{D}} [\ell(f_\theta(x), y)] + \lambda \mathbb{E}_{x \sim \mathcal{D}} [\max(0, C - F(A f_\theta(x) - b))] \quad (1)$$

$$\min_{\theta} \mathbb{E}_{x \sim \mathcal{D}} [\ell(f_\theta(x), y)] \quad \text{s.t.} \quad \varphi(\text{sign}(W f_\theta(x))) = \text{True} \quad (2)$$

$$\mathcal{R}(\theta) = \sum_{k=1}^m \log (1 + \exp(-\alpha s_k(f_\theta(x)))), \quad s_k \in \mathcal{S}_{\text{CNF}} \quad (3)$$

$$\min_{\theta} \mathbb{E}_{x \sim \mathcal{D}} [\ell(f_\theta(x), y)] + \mu \|\nabla_z^2 \ell(f_\theta(x), y)\|_F^2 \quad (4)$$