

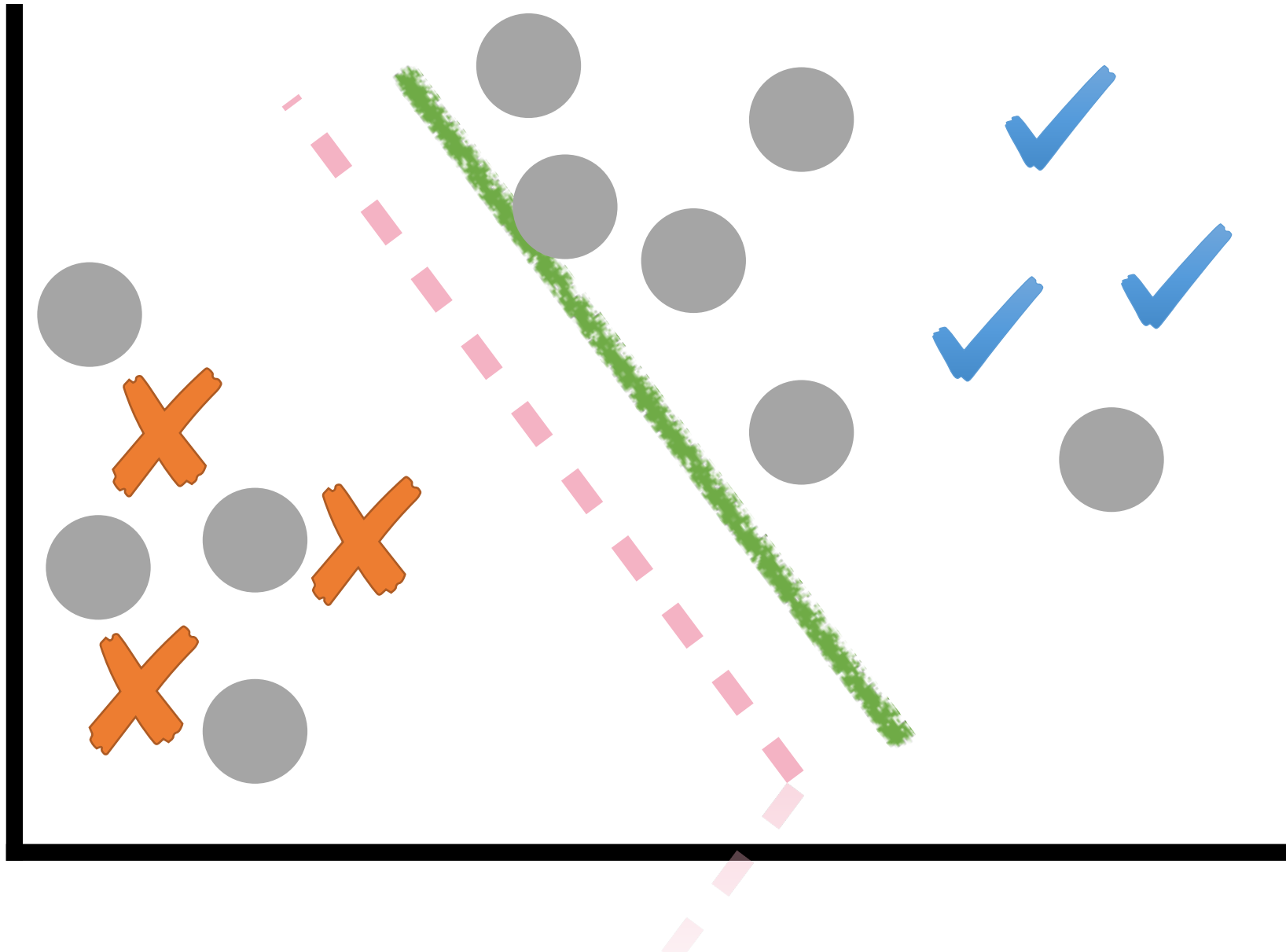
Not All Unlabeled Data are Equal: Learning to Weight in Semi-supervised Learning



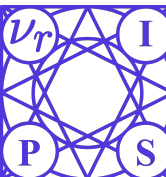
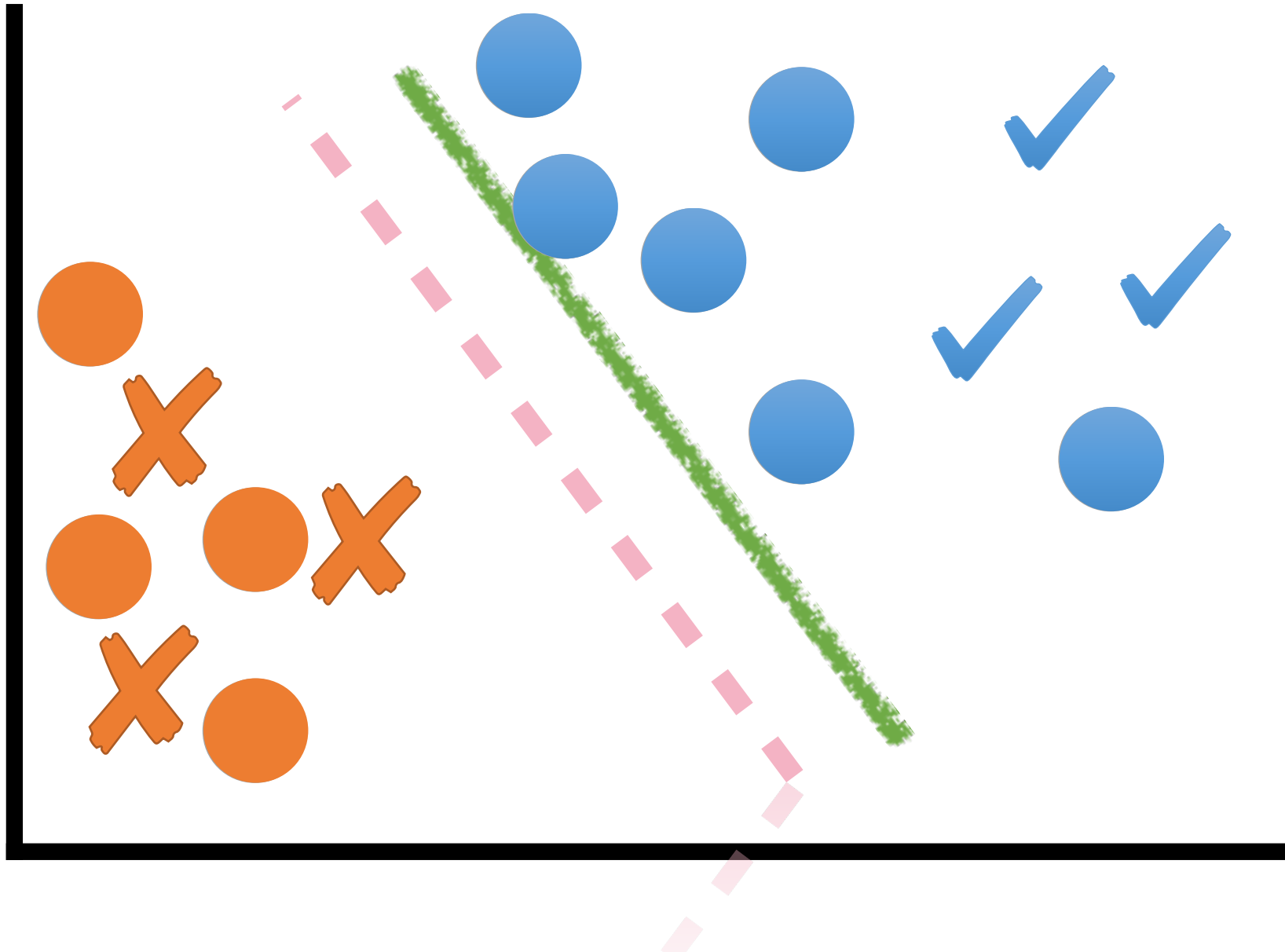
Zhongzheng Ren*, Raymond A. Yeh*, Alexander G. Schwing
University of Illinois at Urbana-Champaign (UIUC)
NeurIPS, 2020



Semi-Supervised Learning (SSL)



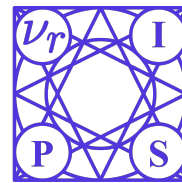
Pseudo-Label Algorithm



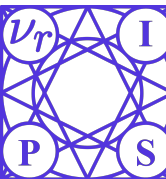
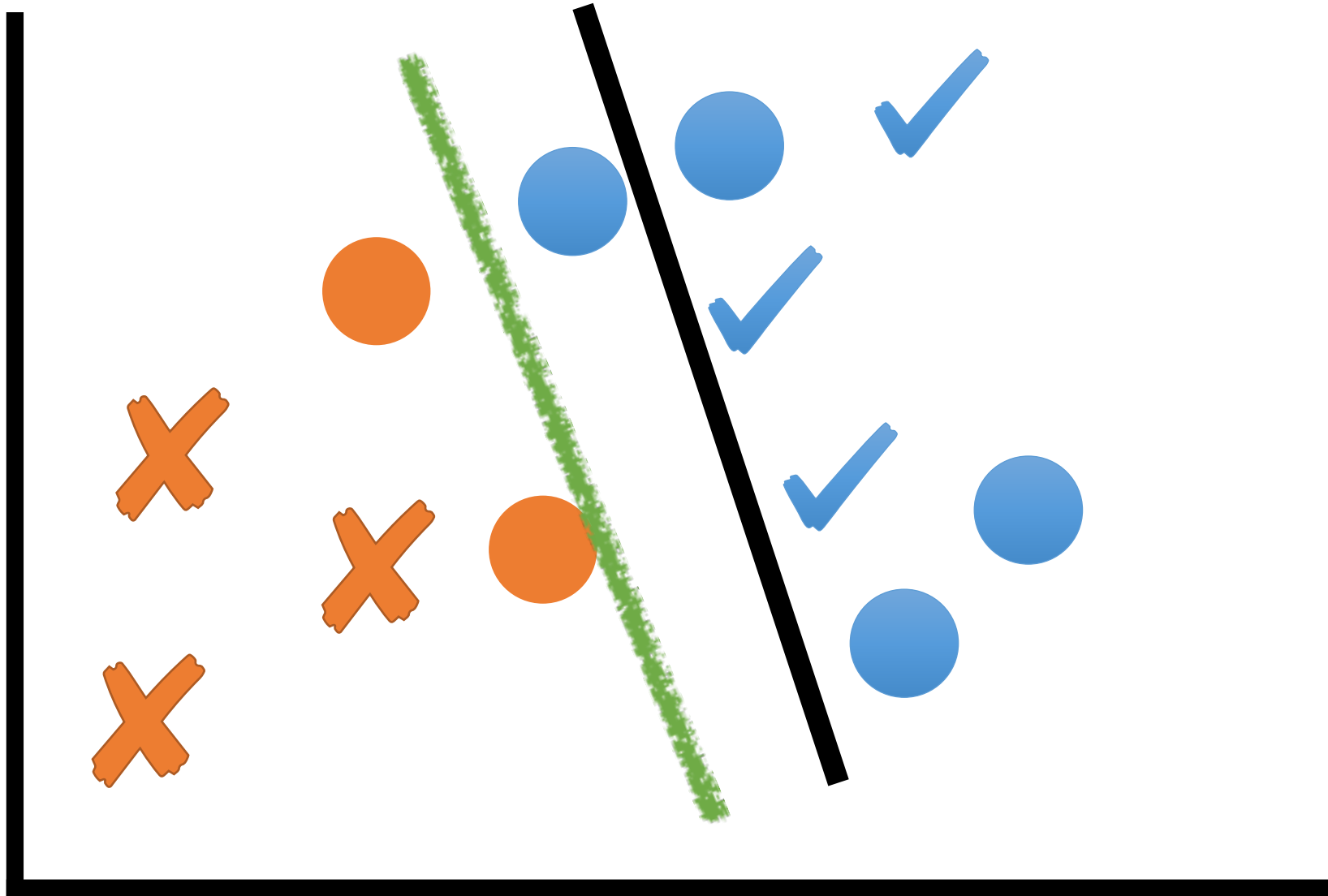
Training Objective:

$$\min_{\theta} \sum_{(x,y) \in \mathcal{D}} \ell_S(x, y, \theta) + \lambda \sum_{u \in \mathcal{U}} \ell_U(u, \theta)$$

- Model parameters θ
- Labeled set $\mathcal{D}$
- Unlabeled set $\mathcal{U}$
- Supervised loss ℓ_S , unsupervised loss ℓ_U
- Scalar weight λ is tuned manually using validation set $\mathcal{V}$



Single weight λ is suboptimal



Tuning Per-example Weight

$$\min_{\theta} \sum_{(x,y) \in \mathcal{D}} \ell_S(x, y, \theta) + \lambda \sum_{u \in \mathcal{U}} \ell_U(u, \theta)$$



$$\min_{\Lambda = \{\lambda_1, \dots, \lambda_{|\mathcal{U}|}\}} L_S(\mathcal{V}, \theta^*(\Lambda))$$

$$s.t. \theta^*(\Lambda) = \operatorname{argmin}_{\theta} \sum_{(x,y) \in \mathcal{D}} \ell_S(x, y, \theta) + \sum_{u \in \mathcal{U}} \lambda_u \cdot \ell_U(u, \theta)$$



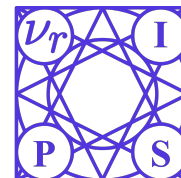
Tuning Per-example Weight

$$\min_{\theta} \sum_{(x,y) \in \mathcal{D}} \ell_S(x, y, \theta) + \sum_{u \in \mathcal{U}} \lambda_u \ell_U(u, \theta)$$



$$\min_{\Lambda = \{\lambda_1, \dots, \lambda_{|\mathcal{U}|}\}} L_S(\mathcal{V}, \theta^*(\Lambda))$$

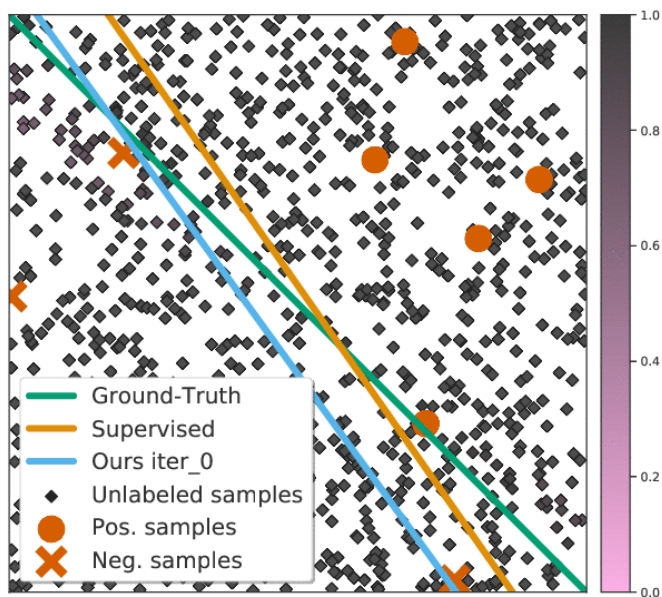
$$s.t. \theta^*(\Lambda) = \operatorname{argmin}_{\theta} \sum_{(x,y) \in \mathcal{D}} \ell_S(x, y, \theta) + \sum_{u \in \mathcal{U}} \lambda_u \cdot \ell_U(u, \theta)$$



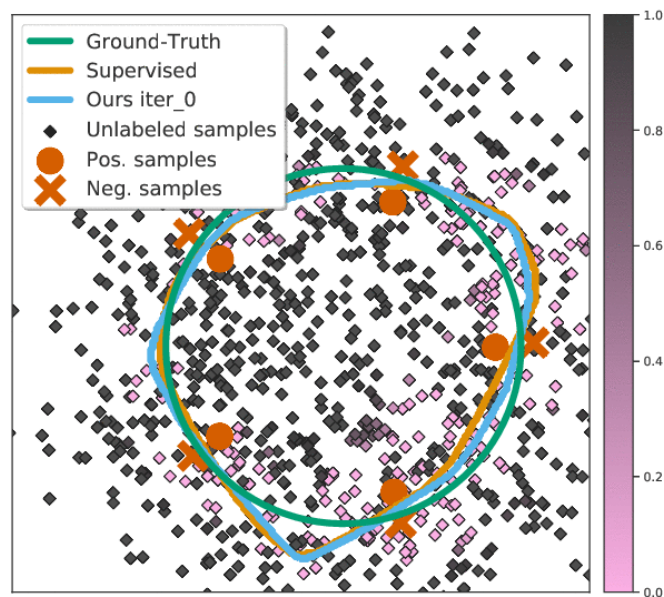
Synthetic Data

- 10 labeled, 1000 unlabeled, 30 validation
- Linear model and a two layer deep-net

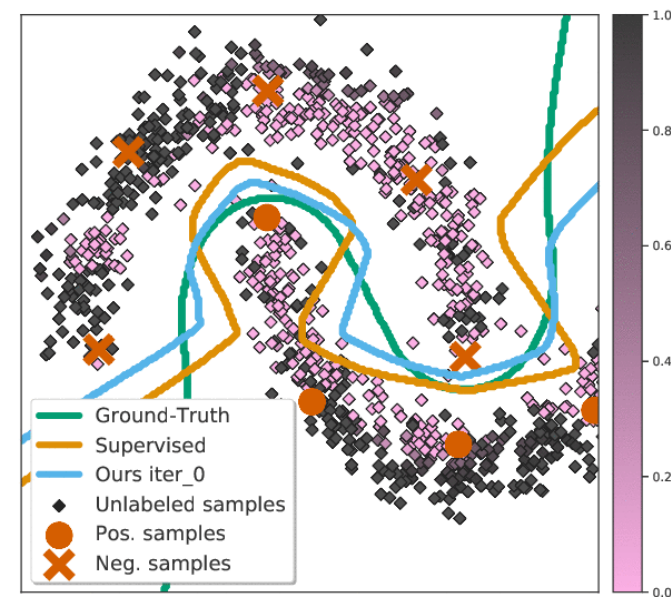
Linearly Separable Data



Circular Data



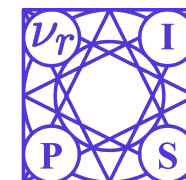
Two Half Moons Data



Comparisons to SOTA methods

Dataset # Labeled	CIFAR-10					SVHN				
	250	500	1000	2000	4000	250	500	1000	2000	4000
Pseudo-Label	49.98 $\pm$ 1.17	40.55 $\pm$ 1.70	30.91 $\pm$ 1.73	21.96 $\pm$ 0.42	16.21 $\pm$ 0.11	21.16 $\pm$ 0.88	14.35 $\pm$ 0.37	10.19 $\pm$ 0.41	7.54 $\pm$ 0.27	5.71 $\pm$ 0.07
VAT	36.03 $\pm$ 2.82	26.11 $\pm$ 1.52	18.64 $\pm$ 0.40	14.40 $\pm$ 0.15	11.05 $\pm$ 0.31	8.41 $\pm$ 1.01	7.44 $\pm$ 0.79	5.98 $\pm$ 0.21	4.85 $\pm$ 0.23	4.20 $\pm$ 0.15
Mean-Teacher	47.32 $\pm$ 4.71	42.01 $\pm$ 5.86	17.32 $\pm$ 4.00	12.17 $\pm$ 0.22	10.36 $\pm$ 0.25	6.45 $\pm$ 2.43	3.82 $\pm$ 0.17	3.75 $\pm$ 0.10	3.51 $\pm$ 0.09	3.39 $\pm$ 0.11
MixMatch	11.08 $\pm$ 0.87	9.65 $\pm$ 0.94	7.75 $\pm$ 0.32	7.03 $\pm$ 0.15	6.24 $\pm$ 0.06	3.78 $\pm$ 0.26	3.64 $\pm$ 0.46	3.27 $\pm$ 0.31	3.04 $\pm$ 0.13	2.89 $\pm$ 0.06
UDA	8.76 $\pm$ 0.90	6.68 $\pm$ 0.24	5.87 $\pm$ 0.13	5.51 $\pm$ 0.21	5.29 $\pm$ 0.25	2.76 $\pm$ 0.17	2.70 $\pm$ 0.09	2.55 $\pm$ 0.09	2.57 $\pm$ 0.09	2.47 $\pm$ 0.15
Re-MixMatch	6.27 $\pm$ 0.34	-	5.73 $\pm$ 0.16	-	5.14 $\pm$ 0.04	3.10 $\pm$ 0.50	-	2.83 $\pm$ 0.30	-	2.42 $\pm$ 0.09
FixMatch (CTA)	5.07 $\pm$ 0.33	-	-	-	4.31$\pm$0.15	2.64 $\pm$ 0.64	-	-	-	2.36 $\pm$ 0.19
FixMatch* (CTA)	5.23 $\pm$ 0.28	-	4.82 $\pm$ 0.09	-	4.48 $\pm$ 0.15	2.77 $\pm$ 0.73	-	2.41 $\pm$ 0.14	-	2.17 $\pm$ 0.08
Ours (UDA)	5.53$\pm$0.17	5.38$\pm$0.23	5.17$\pm$0.16	5.14$\pm$0.17	4.75$\pm$0.28	2.45$\pm$0.08	2.39$\pm$0.04	2.33$\pm$0.06	2.32$\pm$0.06	2.35$\pm$0.05
Ours (FixMatch, CTA)	5.05$\pm$0.12	-	4.68$\pm$0.14	-	4.35$\pm$0.06	2.63$\pm$0.23	-	2.34$\pm$0.15	-	2.15$\pm$0.03

Our method generalized to various SSL algorithms



Visit our poster / project page for more details!

Thanks for watching!



<https://jason718.github.io/project/semi/main>

