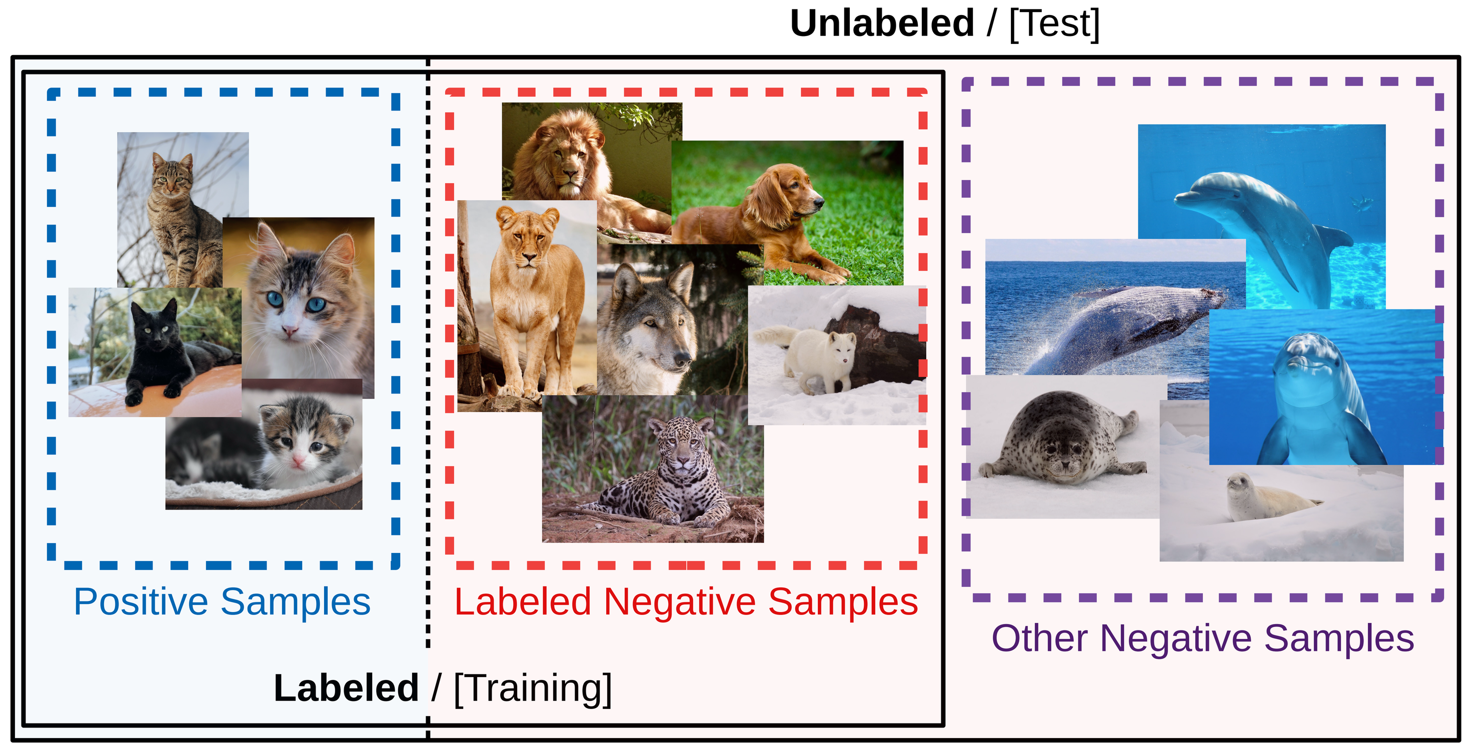
Classification from Positive, Unlabeled and Biased Negative Data

Yu-Guan Hsieh[1] Gang Niu[2] Masashi Sugiyama[2,3]

[1] ENS Paris, France [2] RIKEN, Japan [3] The University of Tokyo, Japan

Motivation

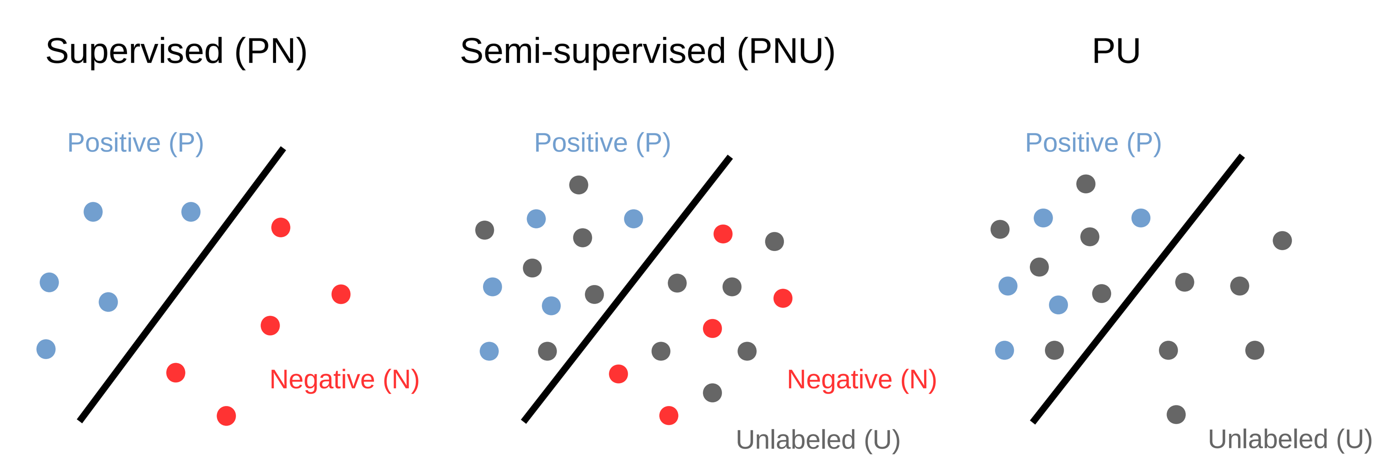


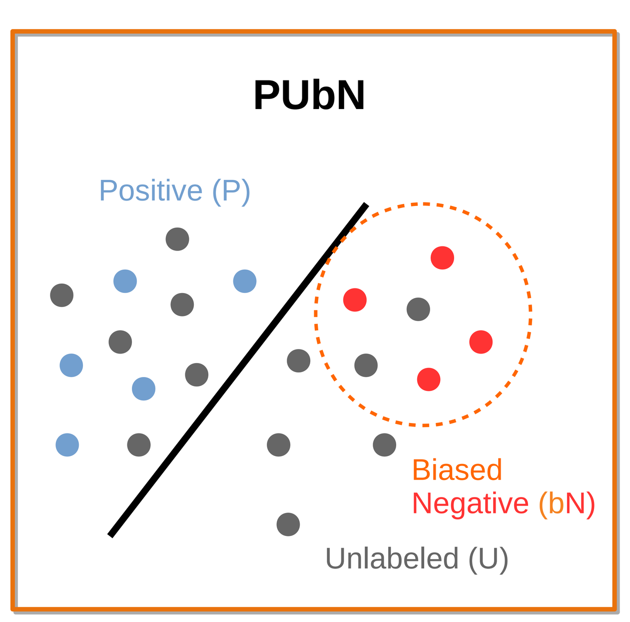
- Information retrieval, text classification
- Medical diagnosis: healthy population that goes through physical exams is biased

Related Work

- **Semi-supervised learning:**
  N data in general implicitly assumed unbiased ; U data used for regularization
- **Dataset shift:** a variant that has been rarely studied
  Ex. Covariate shift ; Source component shift
- **PU learning:** add bN data
- Pseudo-labeling / Importance-weighting

Problem Setting





x : feature
y ∈ {+1, −1} : label
s ∈ {+1, −1} : latent variable causing the bias

N data with selection bias
p(s = +1 | x, y = +1) = 1

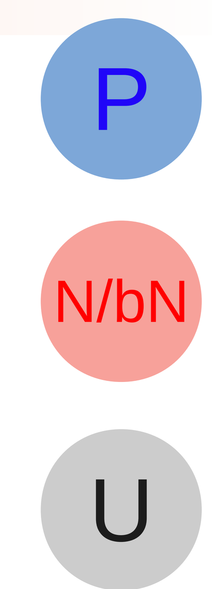
- $\sim p_{\mathrm{P}}(x) := p(x \mid y=+1)$
- $\sim p(x)$
- $\sim p_{\mathrm{bN}}(x) := p(x \mid y=-1, s=+1)$

Method

Empirical Risk Minimization

$$\underbrace{\min_{g\in\mathcal{G}} \mathbb{E}_{(x,y)\sim p(x,y)}[\ell(yg(x))]}_{R(g)} \longleftarrow \boxed{\underbrace{\min_{g\in\mathcal{G}} \frac{1}{n}\sum_{i=1}^{n} \ell(y_i g(x_i))}_{\hat{R}(g)}}$$

Risk Minimization — Unbiased Estimator — #P data + #N data — Empirical Risk Minimization



Q: The N data are biased

$$R(g) = \underbrace{\boxed{\pi R_{\mathrm{P}}^{+}(g)}}_{\text{\#P data}} + \underbrace{\boxed{\rho R_{\mathrm{bN}}^{-}(g)}}_{\text{\#bN data}} + \boxed{\underbrace{(1-\pi-\rho)R_{s=-1}^{-}(g)}_{\bar{R}_{s=-1}^{-}(g)}} \ \textbf{?}$$

Idea: σ(x) probability of x being labeled
η>0 determining how much attention paid to U data

$$\bar{R}_{s=-1}^{-}(g) = \boxed{\mathbb{E}_{x\sim p(x)}[1_{\sigma(x)\le\eta}\, \ell(-g(x))(1-\sigma(x))]}$$

#U data, Partial risk for samples with low probability of being labeled

$$+ \underbrace{\boxed{\pi\mathbb{E}_{x\sim p_{\mathrm{P}}(x)}\left[1_{\sigma(x)>\eta}\,\ell(-g(x))\frac{1-\sigma(x)}{\sigma(x)}\right]}}_{\text{\#P data}} + \underbrace{\boxed{\rho\mathbb{E}_{x\sim p_{\mathrm{bN}}}\left[1_{\sigma(x)>\eta}\,\ell(-g(x))\frac{1-\sigma(x)}{\sigma(x)}\right]}}_{\text{\#bN data}}$$

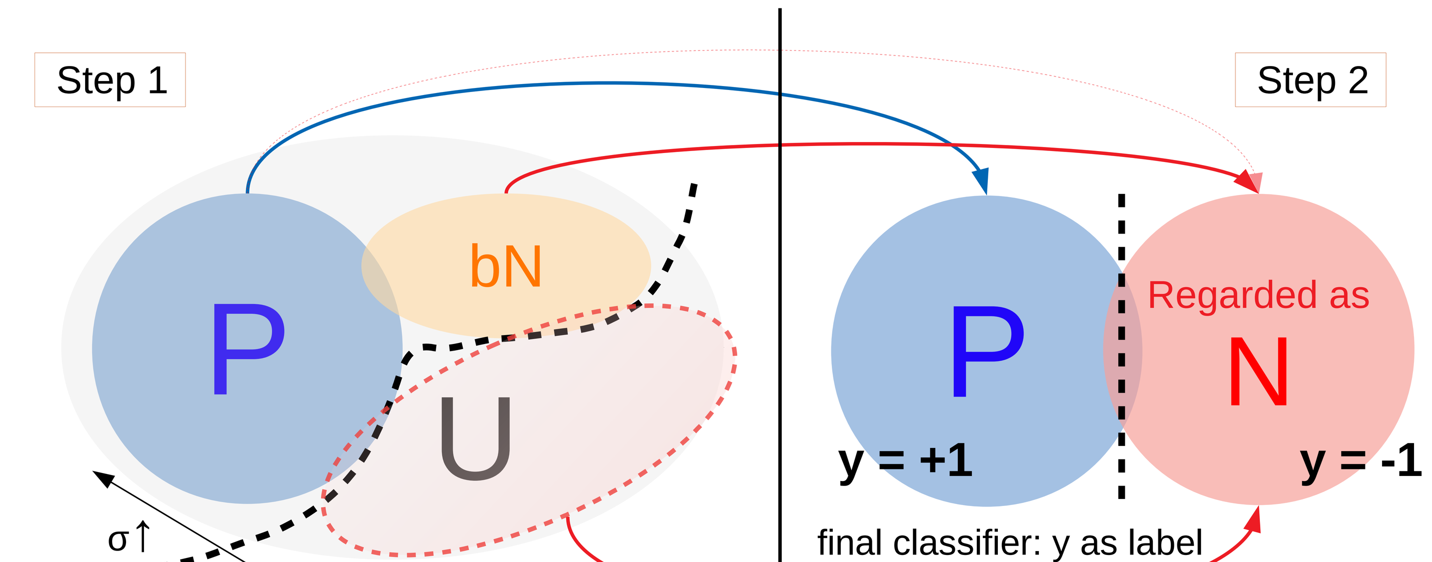
Partial risk for samples with high probability of being labeled

$R_{\mathrm{P}}^{+}(g) := \mathbb{E}_{x\sim p_{\mathrm{P}}(x)}[\ell(g(x))]$ $\pi := p(y=+1)$ ℓ : loss function

$R_{\mathrm{bN}}^{-}(g) := \mathbb{E}_{x\sim p_{\mathrm{bN}}(x)}[\ell(-g(x))]$ $\rho := p(y=-1, s=+1)$

$R_{s=-1}^{-}(g) := \mathbb{E}_{x\sim p(x|s=-1)}[\ell(-g(x))]$ $\sigma(x) := p(s=+1 \mid x)$

Algorithm Outline



estimate σ = p(s=+1|.): s as label

final classifier: y as label

ERM:
pseudo labeling
+ weight adjustment

nnPU classifier (Kiryo+ NeurIPS 2017)

PU risk estimator

$$R(g) = \underbrace{\boxed{\mathbb{E}_{x\sim p(x)}\ell(-g(x))}}_{\text{\#U data}} + \underbrace{\boxed{\pi\mathbb{E}_{x\sim p_{\mathrm{P}}(x)}[\ell(g(x)) - \ell(-g(x))]}}_{\text{\#P data}}$$

Q: Severe overfitting

A: Avoid regarding all U as N

Non-negative correction

N partial risk ≥ 0

$$\tilde{R}_{\mathrm{pu}}(g) = \underbrace{\boxed{\frac{\pi}{n_{\mathrm{P}}}\sum_{x\in\mathcal{X}_{\mathrm{P}}}[\ell(g(x))]}}_{\text{\#P data}} + \max\left\{0, \underbrace{\boxed{\frac{1}{n_{\mathrm{U}}}\sum_{x\in\mathcal{X}_{\mathrm{U}}}\ell(-g(x))}}_{\text{\#U data}} - \underbrace{\boxed{\frac{\pi}{n_{\mathrm{P}}}\sum_{x\in\mathcal{X}_{\mathrm{P}}}[\ell(-g(x))]}}_{\text{\#P data}}\right\}$$

Estimation Error Bound

With probability at least 1-δ

$$R(\hat{g}) - R(g^*) \le \underbrace{\boxed{\frac{4\pi L_\ell}{\eta}\mathfrak{R}_{n_{\mathrm{P}},p_{\mathrm{P}}}(\mathcal{G}) + \frac{2\pi C_\ell}{\eta}\sqrt{\frac{\ln(6/\delta)}{2n_{\mathrm{P}}}}}}_{\text{\#P data}} + \underbrace{\boxed{\frac{4\rho L_\ell}{\eta}\mathfrak{R}_{n_{\mathrm{bN}},p_{\mathrm{bN}}}(\mathcal{G}) + \frac{2\rho C_\ell}{\eta}\sqrt{\frac{\ln(6/\delta)}{2n_{\mathrm{bN}}}}}}_{\text{\#bN data}}$$

$$+ \underbrace{\boxed{4L_\ell\mathfrak{R}_{n_{\mathrm{U}},p}(\mathcal{G}) + 2C_\ell\sqrt{\frac{\ln(6/\delta)}{2n_{\mathrm{U}}}}}}_{\text{\#U data}} + \underbrace{\boxed{2C_\ell\sqrt{\zeta\epsilon} + \frac{2C_\ell}{\eta}\sqrt{(1-\zeta)\epsilon}}}_{\text{Bias due to inexact approximation of }\sigma}$$

$\mathfrak{R}_{n,q}(\mathcal{G})$: Rademacher Complexity

$\hat{\sigma}$: estimate of σ $\zeta := p(\hat{\sigma}(x) \le \eta)$ $\epsilon := \mathbb{E}_{x\sim p(x)}[|\hat{\sigma}(x) - \sigma(x)|^2]$

Assumption ℓ is L_ℓ-Lipschitz $\sup_{g\in\mathcal{G}} \|g\|_\infty \le C_g$ $\sup_{|z|\le C_g} \ell(z) \le C_\ell$

1) If RC terms vanish asymptotically, it holds a.s.

$$\limsup R(\hat{g}) - R(g^*) \le \underbrace{\boxed{2C_\ell\sqrt{\zeta\epsilon} + 2(C_\ell/\eta)\sqrt{(1-\zeta)\epsilon}}}_{\text{Bias due to inexact approximation of }\sigma}$$

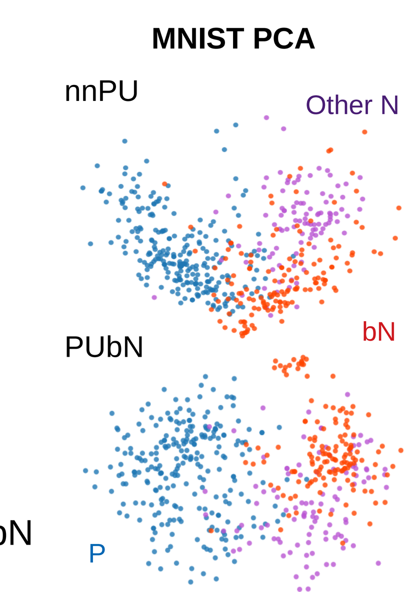
2) Classical convergence rate + bias

3) To control ε: approximation error + estimation error

Experiments

Setting

- Models: ConvNet / ResNet / FCN + Training: Amsgrad
- **VALIDATION!** equally composed of P+U+bN
- 1/10 #U ~ #P = #bN ; Same model for the two steps



Baselines

- nnPNU (Sakai+ 2017 ICML):
  linear combination of PU and PN risk
- PU→PN: one classifier for s and one to separate P from bN

Results

bN data helps

Dataset	P	π	bN	ρ	nnPU/nnPNU	PUbN(\N)	PU→PN
MNIST	2, 4, 6, 8, 10	0.49	Not given	NA	5.76 ± 1.04	**4.64 ± 0.62**	NA
			1, 3, 5	0.3	5.33 ± 0.97	4.05 ± 0.27	**4.00 ± 0.30**
			9 > 5 > others	0.2	4.60 ± 0.65	3.91 ± 0.66	**3.77 ± 0.31**
CIFAR-10	Airplane, automobile, ship, truck	0.4	Not given	NA	12.02 ± 0.65	**10.70 ± 0.57**	NA
			Cat, dog, horse	0.3	10.25 ± 0.38	**9.71 ± 0.51**	10.37 ± 0.65
			Horse > deer = frog > others	0.25	9.98 ± 0.53	**9.92 ± 0.42**	10.17 ± 0.35
CIFAR-10	Cat, deer, dog, horse	0.4	Not given	NA	23.78 ± 1.04	**21.13 ± 0.90**	NA
			Bird, frog	0.2	22.00 ± 0.53	**18.83 ± 0.71**	19.88 ± 0.62
			Car, truck	0.2	22.00 ± 0.74	**20.19 ± 1.06**	21.83 ± 1.36
20 Newsgroups	alt., comp., misc., rec.	0.56	Not given	NA	14.67 ± 0.87	**13.30 ± 0.53**	NA
			sci.	0.21	14.69 ± 0.46	**13.10 ± 0.90**	13.58 ± 0.97
			talk.	0.17	14.38 ± 0.74	**12.61 ± 0.75**	13.76 ± 0.66
			soc. > talk. > sci.	0.1	14.41 ± 0.76	**12.18 ± 0.59**	12.92 ± 0.51