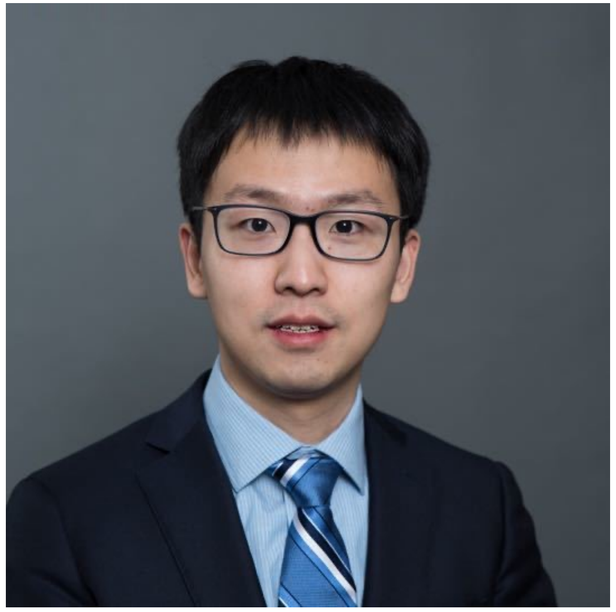
Non-Convex Robust Hypothesis Testing Using Sinkhorn Uncertainty Sets

Jie Wang
Georgia Institute of Technology

ISIT 2024
Session TU1.R4: “Hypothesis Testing 1”

This work is supported by NSF and Coca-Cola Foundation



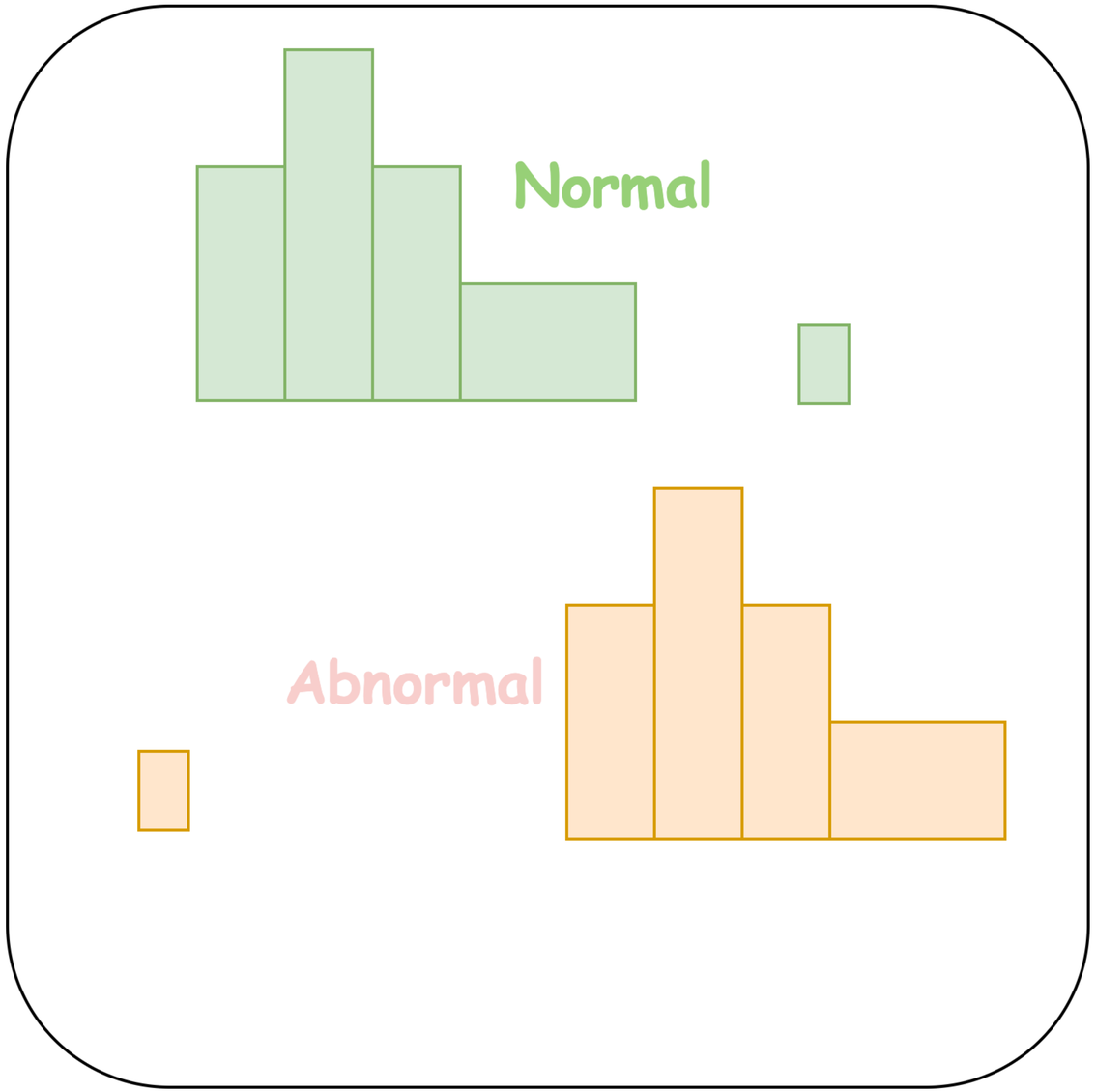
Rui Gao

The University of Texas at Austin

Yao Xie

Georgia Institute of Technology

Introduction



Non-parametric hypothesis test with **unbalanced and limited** data

- Empirical distributions may have non-overlapping support;
- Intractable to leverage (optimal) likelihood ratio method

Problem Formulation

- Composite Hypothesis Test:

$$H_1: \ \omega \sim \mathbb{P}_1, \quad \mathbb{P}_1 \in \mathscr{P}_1 \qquad\qquad H_2: \ \omega \sim \mathbb{P}_2, \quad \mathbb{P}_2 \in \mathscr{P}_2$$

- **Information**: Training samples $\{x_i^k\}_{i\in[n]}$ generated from $\mathbb{P}_k, k = 1,2$

- **Goal**: Find detector $T: \ \Omega \to \mathbb{R}$ to accept H_1 or H_2.

- Risk of Detector:

$$\mathscr{R}(T; \mathbb{P}_1, \mathbb{P}_2) := \max\left(\underbrace{\Pr_{\omega\sim\mathbb{P}_1}\{T(\omega) \le 0\}}_{\textbf{Type-I Error}}, \ \underbrace{\Pr_{\omega\sim\mathbb{P}_2}\{T(\omega) > 0\}}_{\textbf{Type-II Error}}\right)$$

- Robust Testing: minimize worst-case risk

$$\inf_{T:\ \Omega\to\mathbb{R}} \left\{ \sup_{\mathbb{P}_k\in\mathscr{P}_k, k=1,2} \mathscr{R}(T; \mathbb{P}_1, \mathbb{P}_2) \right\}$$

Data-Driven Distributional Uncertainty Set

- Empirical distributions $\widehat{\mathbb{P}}_k, k = 1,2$ based on training samples:

$$\widehat{\mathbb{P}}_k = \frac{1}{n} \sum_{i \in [n]} \delta_{x_i^k}, \qquad \text{where } \{x_i^k\}_i \sim \mathbb{P}_k$$

- Uncertainty sets:

$$\mathscr{P}_k = \left\{ \mathbb{P} : \ \mathscr{D}(\widehat{\mathbb{P}}_k, \mathbb{P}) \leq \rho \right\}, \quad k = 1,2$$

Distributional Uncertainty Set

References	Uncertainty sets	Remarks
[A. Magesh, Z. Sun, V. V. Veeravalli, and S. Zou, 2023, 2024]	Moment	Descriptive Statistical Approach
[Levy 2009]	KL-Divergence	All distributions supported only on training samples
[L. Xie, R. Gao, and Y. Xie, 2018, 2021]	Wasserstein	Include discrete and continuous distributions; LFD supported only on training samples
[Z. Sun and S. Zou]	Kernel MMD	Efficient Computation
[J. Wang, R. Gao, and Y. Xie, 2021, 2024]	**Sinkhorn Discrepancy (Entropic Regularized Wasserstein)**	**All distributions are absolutely continuous!**

Notable Features

- Model:

$$\inf_{T:\,\Omega\to\mathbb{R}} \left\{ \max_{k=1,2} \sup_{\mathbb{P}_k\in\mathscr{P}_k} \Pr_{\omega\sim\mathbb{P}_k}\{(-1)^{k+1}\cdot T(\omega)\le 0\} \right\}$$

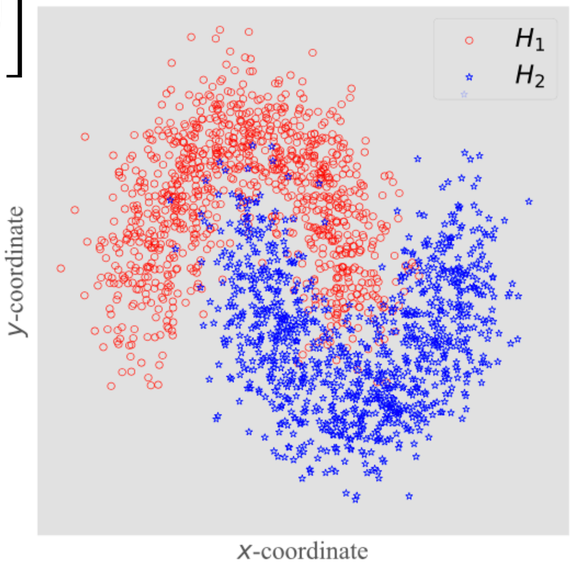
$$\mathscr{P}_k: \quad \mathbf{W}_\epsilon(\widehat{\mathbb{P}}_k, \mathbb{P}_k) \le \rho$$

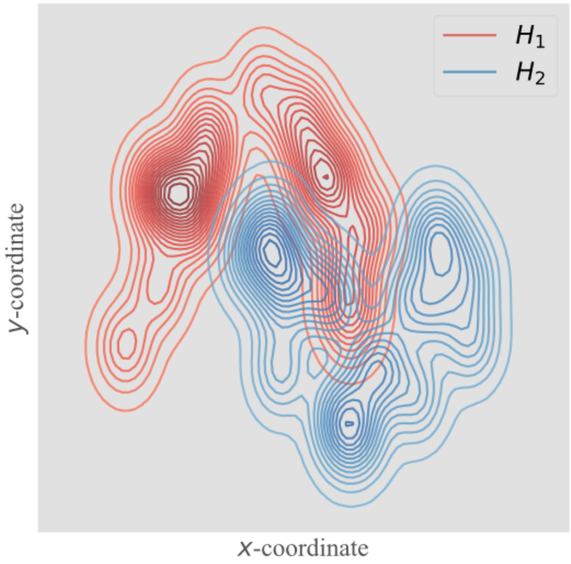
- Least Favorable Distribution (LFD):

$$\frac{\mathrm{d}\mathbb{P}_k^*(z)}{\mathrm{d}z} = \mathbb{E}_{x\sim\widehat{\mathbb{P}}_k}\left[\alpha_x \cdot \exp\left(\frac{\mathbf{1}\{(-1)^{k+1}\cdot T(z)\le 0\} - \lambda_k^*\|x-z\|^2}{\lambda_k^*\epsilon}\right)\right]$$

Normalizing Constant

Density contributed by sample x





Outlines

- Model:

$$\inf_{T:\,\Omega\to\mathbb{R}} \left\{ \max_{k=1,2} \sup_{\mathbb{P}_k\in\mathscr{P}_k} \Pr_{\omega\sim\mathbb{P}_k}\{(-1)^{k+1}\cdot T(\omega)\le 0\} \right\}$$

$$\mathscr{P}_k: \quad \mathbf{W}_\epsilon(\widehat{\mathbb{P}}_k, \mathbb{P}_k) \le \rho$$

  - How to solve this optimization?
    - **Exact and Approximation Algorithms**
  - What is the benefit of Sinkhorn discrepancy-based robust testing?
    - **Regularization effects**

Finite-Dimensional Reformulation

$$\min_{s\geq 0,\ T:\ \Omega\to\mathbb{R}} \quad s$$

$$\text{s.t.} \quad \sup_{\mathbb{P}_1\in\mathscr{P}_1} \mathbb{P}_1\{T(\omega)\leq 0\}\leq s, \qquad \sup_{\mathbb{P}_2\in\mathscr{P}_2} \mathbb{P}_2\{(-1)\cdot T(\omega)\leq 0\}\leq s$$

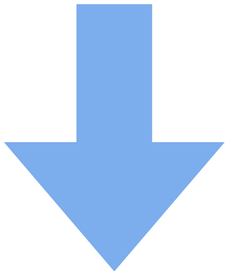
- Random Feature Model: $\mathscr{F}_D = \left\{ T:\ \ x\mapsto \langle \theta, \Phi(x)\rangle, \exists \theta\in\mathbb{R}^D \right\}$
- True detector belongs to RKHS with kernel $K(x,y) = \int \phi(x;\omega)\phi(y;\omega)\mathrm{d}\pi(\omega)$
- Random feature $\Phi(x) = \left(\phi(x;\omega_d)\right)_{d\in[D]}, \quad \omega_d\sim\pi$

$D = O\left(\epsilon^{-2}\right)$ is enough to control approximation error within ϵ

Finite-Dimensional Reformulation

$$\min_{s\geq 0,\ T:\ \Omega\to\mathbb{R}} \quad s$$

$$\text{s.t.} \quad \sup_{\mathbb{P}_1\in\mathscr{P}_1} \mathbb{P}_1\{T(\omega)\leq 0\}\leq s, \qquad \sup_{\mathbb{P}_2\in\mathscr{P}_2} \mathbb{P}_2\{(-1)\cdot T(\omega)\leq 0\}\leq s$$



Random Feature Model

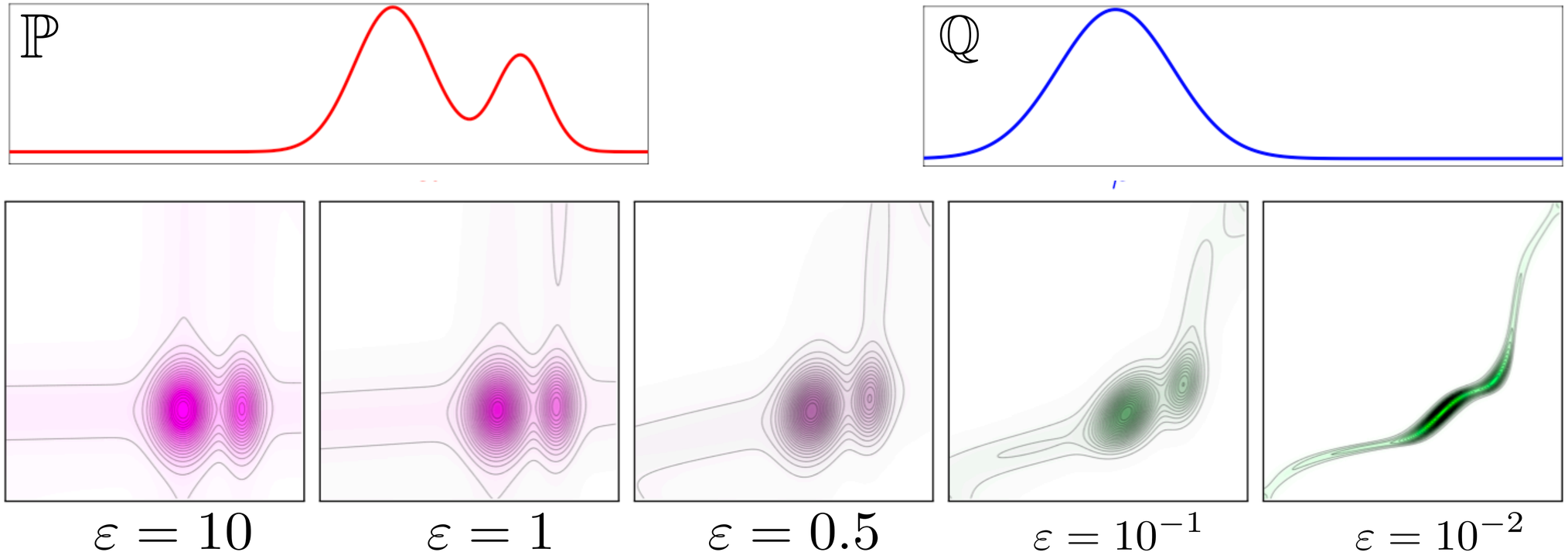
$$\min_{s\geq 0,\ \theta\in\mathbb{R}^D} \quad s$$

$$\text{s.t.} \quad \sup_{\mathbb{P}_1\in\mathscr{P}_1} \mathbb{P}_1\{\langle\theta,\Phi(\omega)\rangle\leq 0\}\leq s, \qquad \sup_{\mathbb{P}_2\in\mathscr{P}_2} \mathbb{P}_2\{(-1)\cdot\langle\theta,\Phi(\omega)\rangle\leq 0\}\leq s$$

Sinkhorn DRO Model [J. Wang, R. Gao, Y. Xie, 2021]

Sinkhorn discrepancy:

$$\mathbf{W}_\epsilon(\mathbb{P},\mathbb{Q}) = \inf_{\gamma\in\Gamma(\mathbb{P},\mathbb{Q})} \left\{ \mathbb{E}_{(x,y)\sim\gamma}[c(x,y)] + \epsilon\mathbb{E}_{(x,y)\sim\gamma}\left[\log\left(\frac{\mathrm{d}\gamma(x,y)}{\mathrm{d}\gamma(x)\mathrm{d}y}\right)\right]\right\}$$



Sinkhorn DRO Model [J. Wang, R. Gao, Y. Xie, 2021]

Sinkhorn discrepancy:

$$\mathbf{W}_\epsilon(\mathbb{P},\mathbb{Q}) = \inf_{\gamma\in\Gamma(\mathbb{P},\mathbb{Q})} \left\{ \mathbb{E}_{(x,y)\sim\gamma}[c(x,y)] + \epsilon \mathbb{E}_{(x,y)\sim\gamma}\left[\log\left(\frac{\mathrm{d}\gamma(x,y)}{\mathrm{d}\gamma(x)\mathrm{d}y}\right)\right]\right\}$$

Under mild conditions, $V_{\mathbf{P}} = V_{\mathbf{D}}$

$$V_{\mathbf{P}} = \sup_{\mathbb{P}} \left\{ \mathbb{E}_{z\sim\mathbb{P}}[f(z)] : \quad \mathbf{W}_\epsilon(\widehat{\mathbb{P}}, \mathbb{P}) \le \rho \right\}$$

$$V_{\mathbf{D}} = \inf_{\lambda\ge 0} \left\{ \lambda\bar{\rho} + \mathbb{E}_{z\sim\widehat{\mathbb{P}}}\left[\lambda\epsilon \log \mathbb{E}_{z'\sim\mathbf{N}(z,\epsilon\mathbf{I})}[e^{f(z)/(\lambda\epsilon)}]\right]\right\}$$

Finite-Dimensional Reformulation

$$\min_{s\geq 0,\ T:\ \Omega\to\mathbb{R}} \quad s$$

$$\text{s.t.} \quad \sup_{\mathbb{P}_1\in\mathscr{P}_1} \mathbb{P}_1\{T(\omega)\leq 0\}\leq s, \qquad \sup_{\mathbb{P}_2\in\mathscr{P}_2} \mathbb{P}_2\{(-1)\cdot T(\omega)\leq 0\}\leq s$$

Random Feature Model

$$\min_{s\geq 0,\ \theta\in\mathbb{R}^D} \quad s$$

$$\text{s.t.} \quad \sup_{\mathbb{P}_1\in\mathscr{P}_1} \mathbb{P}_1\{\langle\theta,\Phi(\omega)\rangle\leq 0\}\leq s, \qquad \sup_{\mathbb{P}_2\in\mathscr{P}_2} \mathbb{P}_2\{(-1)\cdot\langle\theta,\Phi(\omega)\rangle\leq 0\}\leq s$$

$\mathbf{W}_\epsilon(\widehat{\mathbb{P}}_1,\mathbb{P}_1)\leq\rho$

$\mathbf{W}_\epsilon(\widehat{\mathbb{P}}_2,\mathbb{P}_2)\leq\rho$

Finite-Dimensional Reformulation

$$\min_{s\geq 0,\ \theta\in\mathbb{R}^D} \quad s$$

$$\text{s.t.} \quad \sup_{\mathbb{P}_1\in\mathscr{P}_1} \mathbb{P}_1\{\langle\theta,\Phi(\omega)\rangle\leq 0\}\leq s, \qquad \sup_{\mathbb{P}_2\in\mathscr{P}_2} \mathbb{P}_2\{(-1)\cdot\langle\theta,\Phi(\omega)\rangle\leq 0\}\leq s$$

$$\mathbf{W}_\epsilon(\widehat{\mathbb{P}}_1,\mathbb{P}_1)\leq\rho \qquad \mathbf{W}_\epsilon(\widehat{\mathbb{P}}_2,\mathbb{P}_2)\leq\rho$$

Strong Duality Theory from Sinkhorn DRO

$$\min_{s\geq 0,\ \theta\in\mathbb{R}^D,\lambda_1,\lambda_2\geq 0} \quad s$$

$$\text{s.t.} \quad \lambda_k\bar{\rho}_k+\mathbb{E}_{x\sim\widehat{\mathbb{P}}_k}\left[\lambda_k\epsilon_k\log\mathbb{E}_{y\sim\mathbf{N}(x,\epsilon\mathbf{I})}\exp\left(\frac{\mathbf{1}\{(-1)^{k+1}\langle\theta,\Phi(y)\rangle<0\}}{\lambda_k\epsilon_k}\right)\right]\leq 0, \quad k=1,2$$

Finite-Dimensional Reformulation

$$\min_{s\geq 0,\ \theta\in\mathbb{R}^D,\lambda_1,\lambda_2\geq 0} \quad s$$

$$\text{s.t.} \quad \lambda_k\bar{\rho}_k + \mathbb{E}_{x\sim\widehat{\mathbb{P}}_k}\left[\lambda_k\epsilon_k \log \mathbb{E}_{y\sim\mathbf{N}(x,\epsilon\mathbf{I})} \exp\left(\frac{\mathbf{1}\{(-1)^{k+1}\langle\theta,\Phi(y)\rangle < 0\}}{\lambda_k\epsilon_k}\right)\right] \leq 0, \quad k=1,2$$

Cons:

1. Inner expectation $\mathbb{E}_{y\sim\mathbf{N}(x,\epsilon\mathbf{I})}$ is **not finite-dimensional representable**.
2. Indicator function is highly **non-convex and non-smooth.**

Sample Average Approximation

$$\min_{s\geq 0,\ \theta\in\mathbb{R}^D,\lambda_1,\lambda_2\geq 0} \quad s$$

$$\text{s.t.} \quad \lambda_k\bar{\rho}_k + \mathbb{E}_{x\sim\widehat{\mathbb{P}}_k}\left[\lambda_k\epsilon_k \log \mathbb{E}_{y\sim\mathbf{N}(x,\epsilon\mathbf{I})} \exp\left(\frac{\mathbf{1}\{(-1)^{k+1}\langle\theta,\Phi(y)\rangle < 0\}}{\lambda_k\epsilon_k}\right)\right] \leq 0$$

Cons:

1. Inner expectation $\mathbb{E}_{y\sim\mathbf{N}(x,\epsilon\mathbf{I})}$ is **not finite-dimensional representable**.

A. Take $\widehat{\mathbb{P}}_k = \frac{1}{n}\sum_{i\in[n]}\delta_{x_i^k}$. For each x_i^k, sample m i.i.d. samples $y_{i,j}^k \sim \mathbf{N}(x_i^k, \epsilon\mathbf{I})$.

Consistency Holds

B. Solve

$$\min_{s\geq 0,\ \theta\in\mathbb{R}^D,\lambda_1,\lambda_2\geq 0} \quad s$$

$$\text{s.t.} \quad \lambda_k\bar{\rho}_k + \frac{1}{n}\sum_{i\in[n]}\left[\lambda_k\epsilon_k \log \left[\frac{1}{m}\sum_{j\in[m]} \exp\left(\frac{\mathbf{1}\{(-1)^{k+1}\langle\theta,\Phi(y_{i,j}^k)\rangle < 0\}}{\lambda_k\epsilon_k}\right)\right]\right] \leq 0$$

Mixed-Integer Conic Reformulation

$$\min_{s\geq 0,\ \theta\in\mathbb{R}^D,\lambda_1,\lambda_2\geq 0} \quad s$$

$$\text{s.t.} \quad \lambda_k\bar{\rho}_k + \frac{1}{n}\sum_{i\in[n]}\left[\lambda_k\epsilon_k \log\left[\frac{1}{m}\sum_{j\in[m]}\exp\left(\frac{\mathbf{1}\{(-1)^{k+1}\langle\theta,\Phi(y_{i,j}^k)\rangle < 0\}}{\lambda_k\epsilon_k}\right)\right]\right] \leq 0$$

Cons:

1. Inner expectation $\mathbb{E}_{y\sim\mathbf{N}(x,\epsilon\mathbf{I})}$ is **not finite-dimensional representable**.
2. Indicator function is highly **non-convex and non-smooth.**
   A. **Big-M reformulation** of indicator $\mathbf{1}\left\{(-1)^{k+1}\langle\theta,\Phi(y_{i,j}^k)\rangle < 0\right\}$:

$$(-1)^{k+1}\langle\theta,\Phi(y_{i,j}^k)\rangle \leq M_{i,j}^k(1-z_{i,j}^k), \quad z_{i,j}^k\in\{0,1\}$$

   B. Log-sum-exp structure is **exponential conic representable**.

Mixed-Integer Conic Reformulation

Minimize s

$$
\text{s.t.} \quad \begin{cases} \|\theta\|_2 \le 1 \\ (-1)^{k+1}\langle \theta, \Phi(y_{i,j}^k)\rangle \le M_{i,j}^k (1 - z_{i,j}^k) \end{cases}
$$

$$
\begin{cases} \lambda_k \bar{\rho}_k + \dfrac{1}{n}\displaystyle\sum_{i\in[n]} t_i^k \le s \\ \lambda_k \varepsilon_k \ge \dfrac{1}{m}\displaystyle\sum_{j\in[m]} a_{i,j}^k \\ (\lambda_k \varepsilon_k, a_{i,j}^k, z_{i,j}^k - t_i^k) \in \mathcal{K}_{\exp}, \\ i \in [n], j \in [m], k \in \{1,2\} \end{cases}
$$

subject to the following decision variables

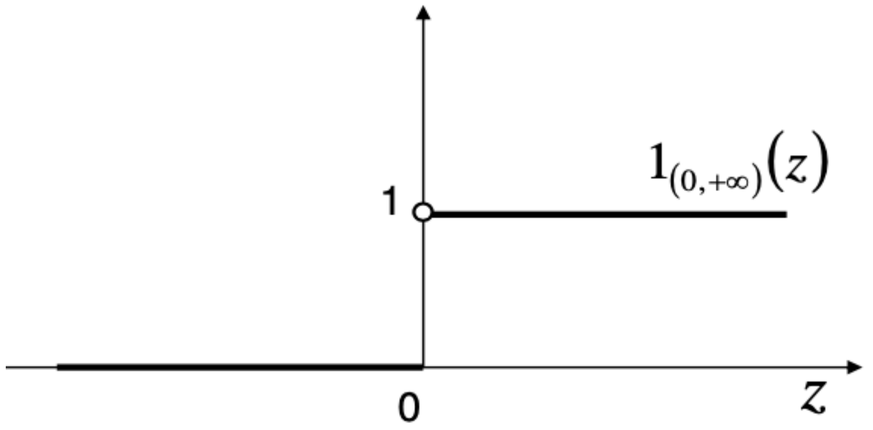
$$
s \in [0,1], \theta \in \mathbb{R}^D, \lambda_1, \lambda_2 \in \mathbb{R}_+, \{t_i^k\}_{i,k} \in \mathbb{R}^{n\times 2},
$$

$$
\{z_{i,j}^k\}_{i,j,k} \in \{0,1\}^{n\times m\times 2}, \{a_{i,j}^k\}_{i,j,k} \in \mathbb{R}^{n\times m\times 2}.
$$

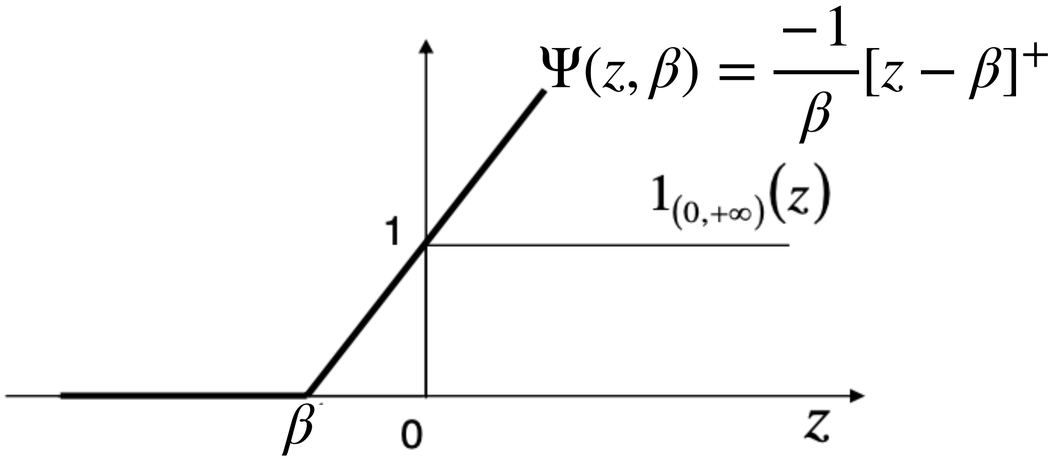
Global Optimization Technique

1. Directly handle it by off-the-shelf solver **Mosek**
2. Develop customized **branch and cut** algorithm.

Conditional Value-at-Risk (CVaR) Approximation



Indicator $\mathbf{1}\{z \geq 0\}$



CVaR Approximation $\Psi(z, \beta)$

[A. Nemirovski and A. Shapiro, 2007]

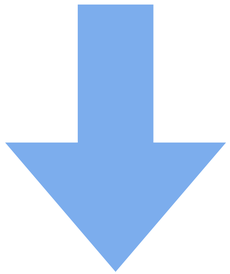
$$\inf_{\beta \leq 0} \left\{ \epsilon\beta + \mathbb{E}\left[Z-\beta\right]_+ \right\} \leq 0 \quad \Rightarrow \quad \mathbb{P}\{Z \geq 0\} \leq \epsilon$$

$$= \mathbb{E}\left[\mathbf{1}\{Z \geq 0\}\right]$$

Convex Approximation (I)

$$\min_{s\geq 0,\ \theta\in\mathbb{R}^D} \quad s$$

$$\text{s.t.} \quad \sup_{\mathbb{P}_k\in\mathscr{P}_k} \mathbb{P}_k\{(-1)^{k+1}\cdot\langle\theta, T(\omega)\rangle \leq 0\} \leq s, \quad k=1,2$$



CVaR Approximation

$$\min_{s\geq 0,\ \theta\in\mathbb{R}^D} \quad s$$

$$\text{s.t.} \quad \sup_{\mathbb{P}_k\in\mathscr{P}_k} \inf_{\beta_k\leq 0} \left\{ s\beta_k + \mathbb{E}_{\mathbb{P}_k}\left[(-1)^k\langle\theta,\Phi(\omega)\rangle - \beta_k\right]_+ \right\} \leq 0, \quad k=1,2$$

Convex Approximation (II)

$$\min_{s\geq 0,\ \theta\in\mathbb{R}^D} \quad s$$

$$\text{s.t.} \quad \sup_{\mathbb{P}_k\in\mathscr{P}_k} \inf_{\beta_k\leq 0} \left\{ s\beta_k + \mathbb{E}_{\mathbb{P}_k}\left[(-1)^k\langle\theta,\Phi(\omega)\rangle - \beta_k\right]_+ \right\} \leq 0, \quad k=1,2$$

MiniMax Theorem

$$\min_{s\geq 0,\ \theta\in\mathbb{R}^D} \quad s$$

$$\text{s.t.} \quad \inf_{\beta_k\leq 0} \sup_{\mathbb{P}_k\in\mathscr{P}_k} \left\{ s\beta_k + \mathbb{E}_{\mathbb{P}_k}\left[(-1)^k\langle\theta,\Phi(\omega)\rangle - \beta_k\right]_+ \right\} \leq 0, \quad k=1,2$$

Convex Approximation (III)

$$\min_{s\geq 0,\ \theta\in\mathbb{R}^D} \quad s$$

$$\text{s.t.} \quad \inf_{\beta_k\leq 0} \sup_{\mathbb{P}_k\in\mathscr{P}_k} \left\{ s\beta_k + \mathbb{E}_{\mathbb{P}_k}\left[(-1)^k\langle\theta,\Phi(\omega)\rangle - \beta_k\right]_+ \right\} \leq 0, \quad k=1,2$$

Strong Duality Theory from Sinkhorn DRO

$$\min_{s\geq 0,\ \beta_k\leq 0,\lambda_k\geq 0,\ k=1,2} \left\{ s:\ \ G_k(s,\beta_k,\lambda_k)\leq 0,\ \ k=1,2 \right\}$$

$$\text{where } G_k(s,\beta_k,\lambda_k) = s\beta_k + \left\{ \lambda_k\bar{\rho} + \mathbb{E}_{x\sim\widehat{\mathbb{P}}_k}\left[\lambda_k\epsilon\log\mathbb{E}_{y\sim\mathbf{N}(x,\epsilon\mathbf{I})}\left[e^{[(-1)^k\langle\theta,\Phi(y)\rangle-\beta_k]_+/(\lambda_k\epsilon)}\right]\right] \right\}$$

Bisection Search for Convex Approximation

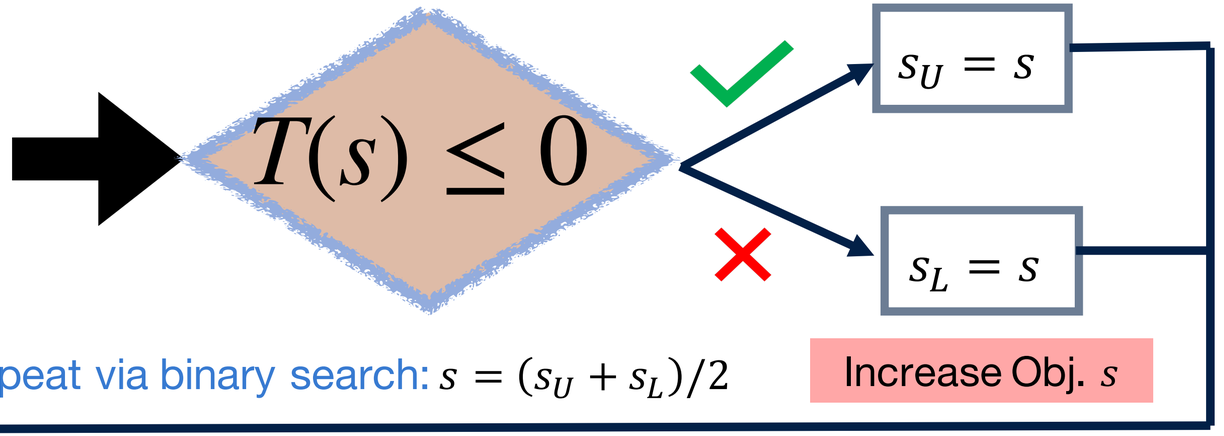
$$\min_{s\geq 0,\ \beta_k\leq 0,\lambda_k\geq 0,\ k=1,2} \left\{ s:\ \ G_k(s,\beta_k,\lambda_k)\leq 0,\ \ \ k=1,2 \right\}$$

$$\text{where } G_k(s,\beta_k,\lambda_k) = s\beta_k + \left\{ \lambda_k\bar{\rho} + \mathbb{E}_{x\sim\widehat{\mathbb{P}}_k}\left[\lambda_k\epsilon \log \mathbb{E}_{y\sim\mathbf{N}(x,\epsilon\mathbf{I})}\left[e^{[(-1)^k\langle\theta,\Phi(y)\rangle-\beta_k]_+/(\lambda_k\epsilon)} \right] \right] \right\}$$

Solve CVaR-Type Loss Function

$$T(s) = \min_{\substack{\theta\in\mathbb{R}^D,\beta_k\leq 0,\\ \lambda_k\geq 0, k=1,2}} \left\{ \max_{k=1,2}\ G_k(s,\beta_k,\lambda_k) \right\}$$

Check Feasibility



Regularization Effects

- Risk of Detector:

$$\mathscr{R}(\theta; \mathbb{P}_1, \mathbb{P}_2) := \max_{k=1,2} \underbrace{\Pr_{\omega\sim\mathbb{P}_k}\{(-1)^{k+1} \cdot \langle \theta, \Phi(\omega)\rangle \le 0\}}_{\textbf{Type-I/Type-II Error}}$$

- Robust hypothesis testing:

$$\inf_{\theta\in\mathbb{R}^D} \left\{ \sup_{\mathbb{P}_k\in\mathscr{P}_k, k=1,2} \mathscr{R}(\theta; \mathbb{P}_1, \mathbb{P}_2) \right\}$$

$$\mathscr{P}_k: \quad \mathbf{W}_\epsilon(\widehat{\mathbb{P}}_k, \mathbb{P}_k) \le \rho$$

- Regularization Effects:

$$\text{(Robust Testing)} \approx \inf_{\theta\in\mathbb{R}^D} \left\{ \mathscr{R}(\theta; \widehat{\mathbb{P}}_1, \widehat{\mathbb{P}}_2) + \text{Regularization} \right\}$$

Regularization Effects

- Robust hypothesis testing:

$$\inf_{\theta\in\mathbb{R}^D}\left\{\sup_{\mathbb{P}_k\in\mathscr{P}_k,k=1,2}\mathscr{R}(\theta;\mathbb{P}_1,\mathbb{P}_2)\right\}$$

$$\mathscr{P}_k:\quad \mathbf{W}_\epsilon(\widehat{\mathbb{P}}_k,\mathbb{P}_k)\le\rho$$

- **Case 1:** $\bar{\rho}/\epsilon\to 0$

$$\text{(Robust Testing)}\approx\inf_{\theta\in\mathbb{R}^D}\left\{\max_{k=1,2}\left(\Pr_{\omega\sim\widehat{\mathbb{P}}_k}\left\{(-1)^{k+1}\langle\theta,\Phi(\omega)\rangle\le 0\right\}\right.\right.$$
$$\left.\left.+\mathbb{E}_{x\sim\widehat{\mathbb{P}}_k}\left[\mathbb{V}\mathrm{ar}_{z\sim\mathbf{N}(x,\epsilon\mathbf{I})}\left[1\{(-1)^{k+1}\langle\theta,\Phi(\omega)\rangle\le 0\}\right]\right]\right)\right\}$$

- Hypothesis test on **perturbed data** should be **deterministic**!

Regularization Effects

- Robust hypothesis testing:

$$\inf_{\theta\in\mathbb{R}^D}\left\{\sup_{\mathbb{P}_k\in\mathscr{P}_k,k=1,2}\mathscr{R}(\theta;\mathbb{P}_1,\mathbb{P}_2)\right\}$$

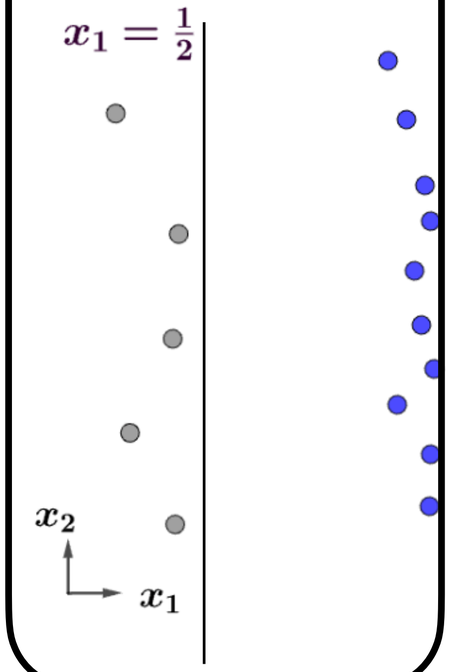
$$\mathscr{P}_k:\quad \mathbf{W}_\epsilon(\widehat{\mathbb{P}}_k,\mathbb{P}_k)\le\rho$$

- **Case 2:** $\bar{\rho}/\epsilon\to\infty$

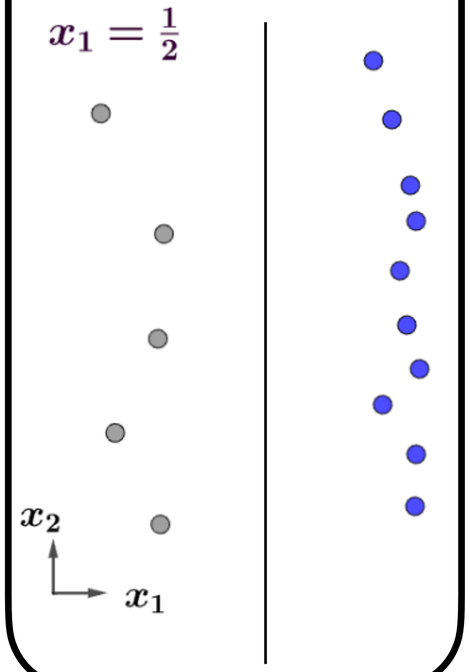
$$\text{(Robust Testing)}\approx\inf_{\theta\in\mathbb{R}^D}\left\{\max_{k=1,2}\left(\Pr_{\omega\sim\widehat{\mathbb{P}}_k}\left\{(-1)^{k+1}\langle\theta,\Phi(\omega)\rangle\le 0\right\}+\mathfrak{g}_k(\theta)\right)\right\}$$

- Here $\mathfrak{g}_k(\theta)$ quantifies the **density of $\mathbb{P}_k$ around the decision boundary**

Large $\mathfrak{g}_1(\theta)$

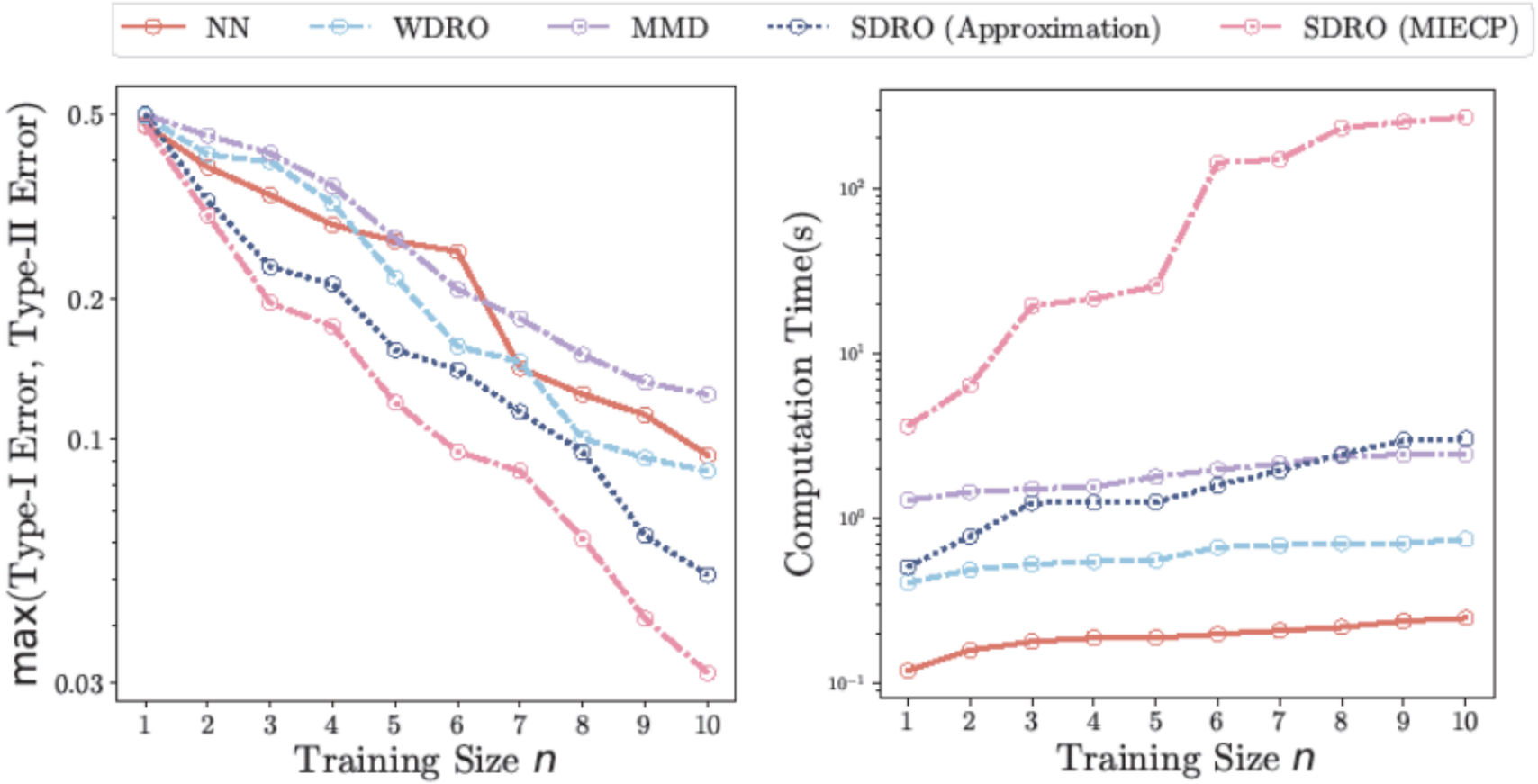


Small $\mathfrak{g}_1(\theta)$



Experiment with Gaussian Mixture Model

- Baselines: **NN** (Neural Network Detector), **WDRO** (Wasserstein robust testing), **MMD** (MMD robust testing), **SDRO** (**Exact** and **Approximation** Formulations)



Experiment with Real Datasets

Parameters	MNIST (Label 1 and 2)	CIFAR-10 (Label 1 and 2)	Lung Cancer (Label 1 and 2)	Sepsis
Training Size	50	50	12	20000
Testing Size	2115	2000	10	3662
Data Dimension	784	1024	56	39

Method	MNIST (Label 1 and 2)	CIFAR-10 (Label 1 and 2)	Lung Cancer (Label 1 and 2)	Sepsis
NN	0.310	0.456	0.400	0.385
WDRO	0.129	0.232	**0.300**	0.256
MMD	0.356	0.445	0.500	0.297
SDRO (Approximation)	**0.0912**	**0.133**	**0.300**	**0.223**

Take Aways

- **Robust hypothesis testing with Sinkhorn discrepancy and 0-1 Loss**
  - Exact and Approximation Algorithms
  - Regularization effects for different scalings of ρ/ϵ