POSTER #124

Differentially Private Learning of Geometric Concepts

Uri Stemmer

Ben-Gurion University

joint work with

Haim Kaplan, Yishay Mansour, and Yossi Matias

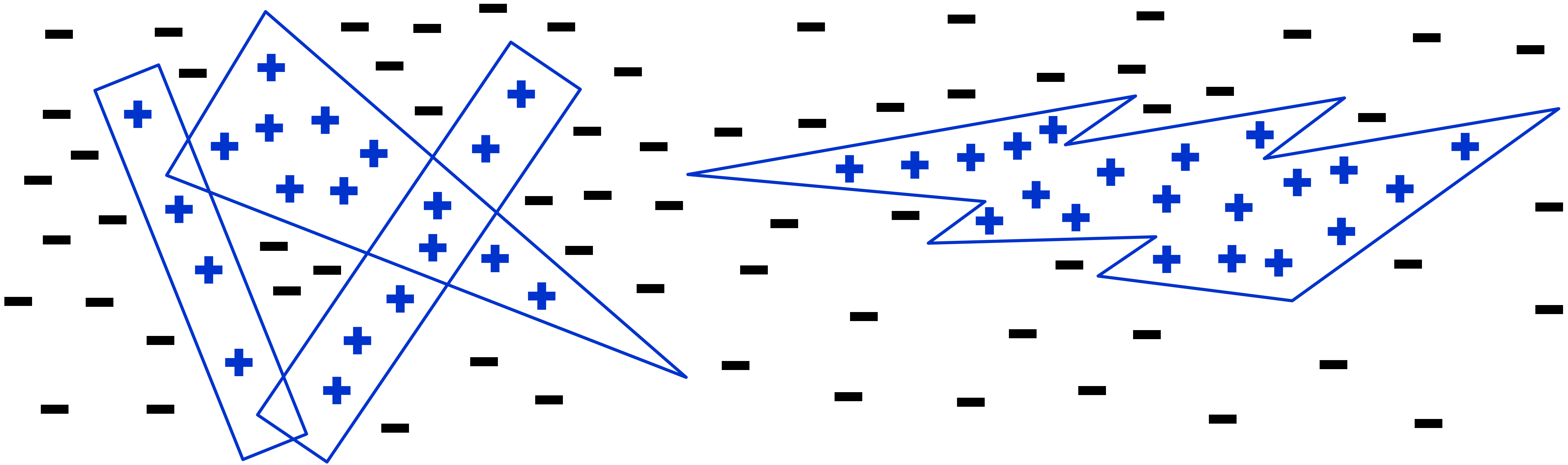
Privately Learning Union of Polygons

POSTER #124

Given: $\boldsymbol{n}$ points in $\mathbb{R}^2$ with binary labels: $\{(\boldsymbol{x_i}, \boldsymbol{y_i})\}_{i=1}^{n}$

Assume: ∃collection of polygons $\{\boldsymbol{P_1}, \dots, \boldsymbol{P_t}\}$ with a total of al most $\boldsymbol{k}$ edges s.t. $\forall \boldsymbol{i} \in [\boldsymbol{n}]$: $\boldsymbol{x_i} \in \bigcup_j \boldsymbol{P_j} \Leftrightarrow \boldsymbol{y_i} = \mathbf{1}$

Find: Hypothesis $\boldsymbol{h}$: $\mathbb{R}^2 \rightarrow \{\mathbf{0}, \mathbf{1}\}$ with small error



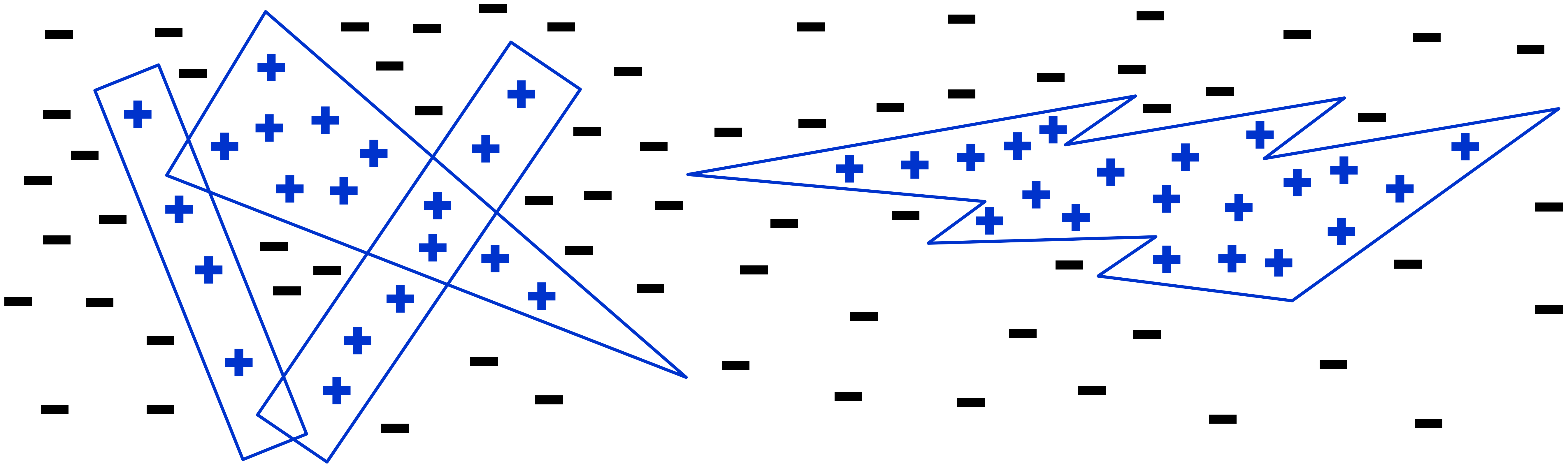
Privately Learning Union of Polygons

POSTER #124

Given: $\boldsymbol{n}$ points in $\mathbb{R}^2$ with binary labels: $\{(\boldsymbol{x_i}, \boldsymbol{y_i})\}_{i=1}^{n}$

Assume: $\exists$collection of polygons $\{\boldsymbol{P_1}, \ldots, \boldsymbol{P_t}\}$ with a total of al most $\boldsymbol{k}$ edges s.t. $\forall \boldsymbol{i} \in [\boldsymbol{n}]$: $\boldsymbol{x_i} \in \bigcup_j \boldsymbol{P_j} \Leftrightarrow \boldsymbol{y_i} = \mathbf{1}$

Find: Hypothesis $\boldsymbol{h}: \mathbb{R}^2 \rightarrow \{\mathbf{0}, \mathbf{1}\}$ with small error, **while providing differential privacy for the training data:**



Privately Learning Union of Polygons

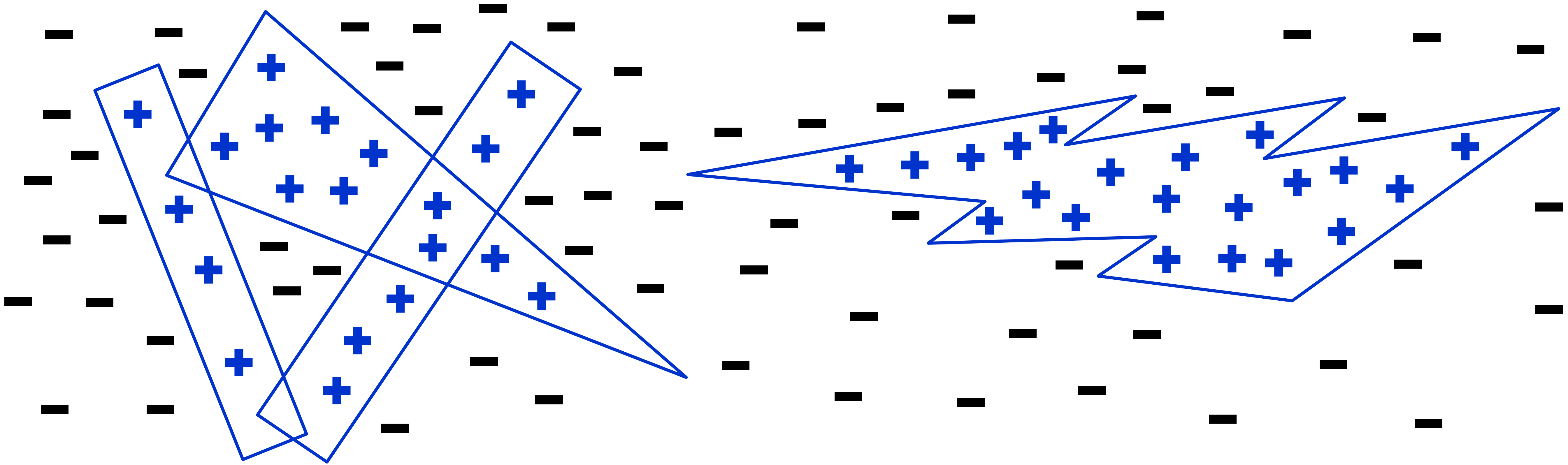
POSTER #124

Given: n points in $\mathbb{R}^2$ with binary labels: $\{(x_i, y_i)\}_{i=1}^{n}$

Assume: $\exists$collection of polygons $\{P_1, \dots, P_t\}$ with a total of al most k edges s.t. $\forall i \in [n]$: $x_i \in \bigcup_j P_j \Leftrightarrow y_i = 1$

Find: Hypothesis $h: \mathbb{R}^2 \to \{0, 1\}$ with small error, **while providing differential privacy for the training data:**

- ✓ Every labeled example represents the (private) information of one individual
- ✓ **Goal:** the output hypothesis does not reveal information that is specific to any single individual



Privately Learning Union of Polygons

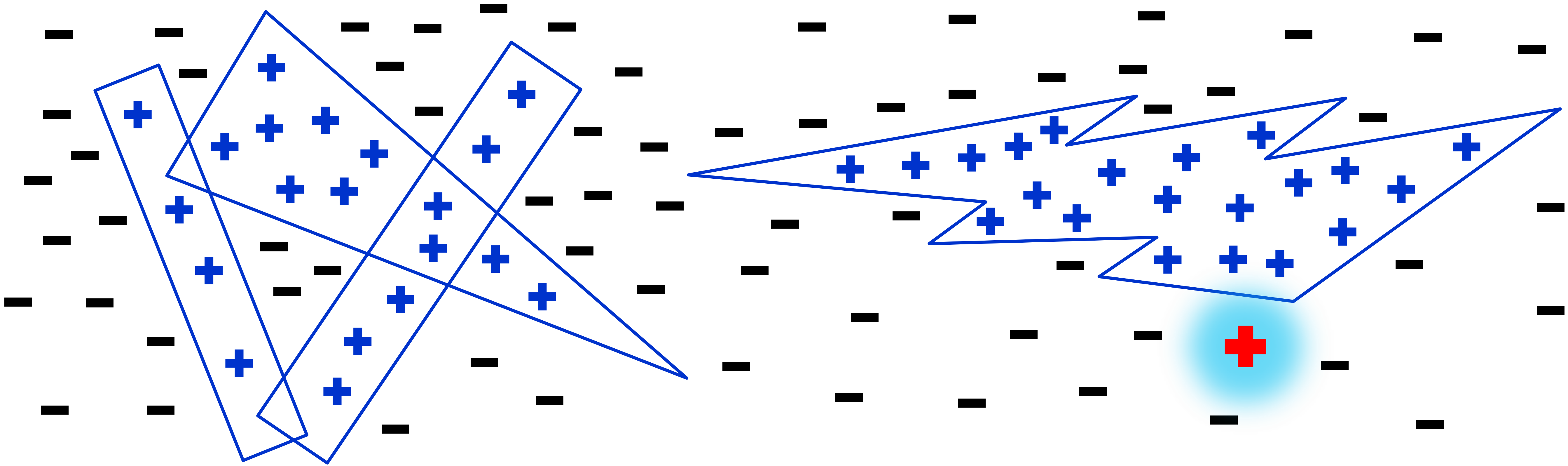
POSTER #124

Given: n points in $\mathbb{R}^2$ with binary labels: $\{(x_i, y_i)\}_{i=1}^{n}$

Assume: $\exists$collection of polygons $\{P_1, \dots, P_t\}$ with a total of al most k edges s.t. $\forall i \in [n]$: $x_i \in \cup_j P_j \Leftrightarrow y_i = 1$

Find: Hypothesis $h: \mathbb{R}^2 \to \{0, 1\}$ with small error, **while providing differential privacy for the training data:**

- ✓ Every labeled example represents the (private) information of one individual
- ✓ **Goal:** the output hypothesis does not reveal information that is specific to any single individual
- ✓ **Requirement:** the output distribution is insensitive to any arbitrarily change of a single input example (an algorithm satisfying this requirement is ***differentially private***)



Privately Learning Union of Polygons

POSTER #124

Given: n points in $\mathbb{R}^2$ with binary labels: $\{(x_i, y_i)\}_{i=1}^{n}$

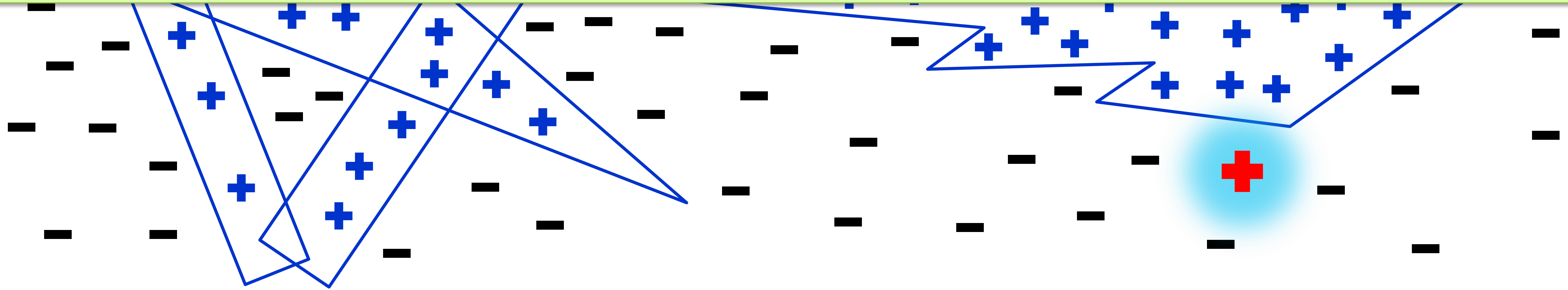
Assume: $\exists$collection of polygons $\{P_1, \dots, P_t\}$ with a total of al most k edges s.t. $\forall i \in [n]$: $x_i \in \cup_j P_j \Leftrightarrow y_i = 1$

Find: Hypothesis $h: \mathbb{R}^2 \to \{0, 1\}$ with small error, **while providing differential privacy for the training data:**

- ✓ Every labeled example represents the (private) information of one individual
- ✓ **Goal:** the output hypothesis does not reveal information that is specific to any single individual
- ✓ **Requirement:** the output distribution is insensitive to any arbitrarily change of a single input example (an algorithm satisfying this requirement is ***differentially private***)

Why is that a good privacy definition?

Even if an observer knows all other data point but mine, and now she sees the outcome of the computation, then she still cannot learn "anything" on my data point



Privately Learning Union of Polygons

POSTER #124

Given: $\boldsymbol{n}$ points in $\mathbb{R}^2$ with binary labels: $\{(\boldsymbol{x_i}, \boldsymbol{y_i})\}_{\boldsymbol{i=1}}^{\boldsymbol{n}}$

Assume: $\exists$collection of polygons $\{\boldsymbol{P_1}, \dots, \boldsymbol{P_t}\}$ with a total of al most $\boldsymbol{k}$ edges s.t. $\forall \boldsymbol{i} \in [\boldsymbol{n}]$: $\boldsymbol{x_i} \in \bigcup_j \boldsymbol{P_j} \Leftrightarrow \boldsymbol{y_i} = \boldsymbol{1}$

Find: Hypothesis $\boldsymbol{h}: \mathbb{R}^2 \rightarrow \{\boldsymbol{0}, \boldsymbol{1}\}$ with small error, **while providing differential privacy for the training data**

Privately Learning Union of Polygons

POSTER #124

Given: $\boldsymbol{n}$ points in $\mathbb{R}^2$ with binary labels: $\{(\boldsymbol{x_i}, \boldsymbol{y_i})\}_{i=1}^{n}$

Assume: $\exists$collection of polygons $\{\boldsymbol{P_1}, \dots, \boldsymbol{P_t}\}$ with a total of al most $\boldsymbol{k}$ edges s.t. $\forall \boldsymbol{i} \in [\boldsymbol{n}]$: $\boldsymbol{x_i} \in \bigcup_j \boldsymbol{P_j} \Leftrightarrow \boldsymbol{y_i} = \mathbf{1}$

Find: Hypothesis $\boldsymbol{h}: \mathbb{R}^2 \to \{\mathbf{0}, \mathbf{1}\}$ with small error, **while providing differential privacy for the training data**

Motivation: Analyzing Users' Location Reports

- ***Analyzing GPS navigation data***
- ***Learning the shape of a flood or a fire based on reports***
- ***Identifying regions with poor cellular reception based on reports***

Privately Learning Union of Polygons

POSTER #124

Given: $\boldsymbol{n}$ points in $\mathbb{R}^2$ with binary labels: $\{(\boldsymbol{x_i}, \boldsymbol{y_i})\}_{i=1}^{n}$

Assume: $\exists$collection of polygons $\{\boldsymbol{P_1}, \dots, \boldsymbol{P_t}\}$ with a total of al most $\boldsymbol{k}$ edges s.t. $\forall \boldsymbol{i} \in [\boldsymbol{n}]$: $\boldsymbol{x_i} \in \bigcup_j \boldsymbol{P_j} \Leftrightarrow \boldsymbol{y_i} = \boldsymbol{1}$

Find: Hypothesis $\boldsymbol{h}: \mathbb{R}^2 \to \{\boldsymbol{0}, \boldsymbol{1}\}$ with small error, **while providing differential privacy for the training data**

Privately Learning Union of Polygons

POSTER #124

Given: $\boldsymbol{n}$ points in $\mathbb{R}^2$ with binary labels: $\{(\boldsymbol{x}_i, \boldsymbol{y}_i)\}_{i=1}^{n}$

Assume: $\exists$collection of polygons $\{\boldsymbol{P}_1, \ldots, \boldsymbol{P}_t\}$ with a total of al most $\boldsymbol{k}$ edges s.t. $\forall \boldsymbol{i} \in [\boldsymbol{n}]$: $\boldsymbol{x}_i \in \bigcup_j \boldsymbol{P}_j \Leftrightarrow \boldsymbol{y}_i = \mathbf{1}$

Find: Hypothesis $\boldsymbol{h}: \mathbb{R}^2 \to \{\mathbf{0}, \mathbf{1}\}$ with small error, **while providing differential privacy for the training data**

Differential Privacy and Discretization

- Impossibility results for differential privacy show that this problem (and even much simpler problems) cannot be solved over infinite domains

Privately Learning Union of Polygons

POSTER #124

Given: $\boldsymbol{n}$ points in $\mathbb{R}^2$ with binary labels: $\{(\boldsymbol{x_i}, \boldsymbol{y_i})\}_{i=1}^{n}$

Assume: ∃collection of polygons $\{\boldsymbol{P_1}, \dots, \boldsymbol{P_t}\}$ with a total of al most $\boldsymbol{k}$ edges s.t. $\forall \boldsymbol{i} \in [\boldsymbol{n}]: \boldsymbol{x_i} \in \bigcup_j \boldsymbol{P_j} \Leftrightarrow \boldsymbol{y_i} = \mathbf{1}$

Find: Hypothesis $\boldsymbol{h}: \mathbb{R}^2 \to \{\mathbf{0}, \mathbf{1}\}$ with small error, **while providing differential privacy for the training data**

Differential Privacy and Discretization

- Impossibility results for differential privacy show that this problem (and even much simpler problems) cannot be solved over infinite domains
- We assume that input points come from $[\boldsymbol{d}]^2 = \{\mathbf{1}, \mathbf{2}, \dots, \boldsymbol{d}\} \times \{\mathbf{1}, \mathbf{2}, \dots, \boldsymbol{d}\}$ for a discretization parameter $\boldsymbol{d}$

Privately Learning Union of Polygons

POSTER #124

Given: $\boldsymbol{n}$ points in $\mathbb{R}^2$ with binary labels: $\{(\boldsymbol{x_i}, \boldsymbol{y_i})\}_{i=1}^{n}$

Assume: $\exists$collection of polygons $\{\boldsymbol{P_1}, \ldots, \boldsymbol{P_t}\}$ with a total of al most $\boldsymbol{k}$ edges s.t. $\forall \boldsymbol{i} \in [\boldsymbol{n}]$: $\boldsymbol{x_i} \in \cup_j \boldsymbol{P_j} \Leftrightarrow \boldsymbol{y_i} = \mathbf{1}$

Find: Hypothesis $\boldsymbol{h}$: $\mathbb{R}^2 \rightarrow \{\mathbf{0}, \mathbf{1}\}$ with small error, **while providing differential privacy for the training data**

Differential Privacy and Discretization

- Impossibility results for differential privacy show that this problem (and even much simpler problems) cannot be solved over infinite domains
- We assume that input points come from $[\boldsymbol{d}]^2 = \{\mathbf{1}, \mathbf{2}, \ldots, \boldsymbol{d}\} \times \{\mathbf{1}, \mathbf{2}, \ldots, \boldsymbol{d}\}$ for a discretization parameter $\boldsymbol{d}$
- Furthermore, the sample complexity must grow with the size of the discretization

Privately Learning Union of Polygons

POSTER #124

Given: $\boldsymbol{n}$ points in $[d]^2$ with binary labels: $\{(\boldsymbol{x}_i, \boldsymbol{y}_i)\}_{i=1}^{n}$

Assume: $\exists$collection of polygons $\{\boldsymbol{P}_1, \dots, \boldsymbol{P}_t\}$ with a total of al most $\boldsymbol{k}$ edges s.t. $\forall \boldsymbol{i} \in [\boldsymbol{n}]$: $\boldsymbol{x}_i \in \bigcup_j \boldsymbol{P}_j \Leftrightarrow \boldsymbol{y}_i = \mathbf{1}$

Find: Hypothesis $\boldsymbol{h}: \mathbb{R}^2 \rightarrow \{\mathbf{0}, \mathbf{1}\}$ with small error, **while providing differential privacy for the training data**

Privately Learning Union of Polygons

POSTER #124

Given: $\boldsymbol{n}$ points in $[\boldsymbol{d}]^2$ with binary labels: $\{(\boldsymbol{x_i}, \boldsymbol{y_i})\}_{\boldsymbol{i=1}}^{\boldsymbol{n}}$

Assume: $\exists$collection of polygons $\{\boldsymbol{P_1}, \dots, \boldsymbol{P_t}\}$ with a total of al most $\boldsymbol{k}$ edges s.t. $\forall \boldsymbol{i} \in [\boldsymbol{n}]$: $\boldsymbol{x_i} \in \bigcup_j \boldsymbol{P_j} \Leftrightarrow \boldsymbol{y_i} = \mathbf{1}$

Find: Hypothesis $\boldsymbol{h}: \mathbb{R}^2 \to \{\mathbf{0}, \mathbf{1}\}$ with small error, **while providing differential privacy for the training data**

Previous Result

Private learner with sample complexity

$\mathbf{O}(\boldsymbol{k} \cdot \mathbf{log}\, \boldsymbol{d})$ and runtime $\approx \boldsymbol{d}^{\boldsymbol{k}}$

(using a generic tool of MT'07)

Privately Learning Union of Polygons

POSTER #124

Given: n points in $[d]^2$ with binary labels: $\{(x_i, y_i)\}_{i=1}^{n}$

Assume: $\exists$collection of polygons $\{P_1, \ldots, P_t\}$ with a total of al most k edges s.t. $\forall i \in [n]$: $x_i \in \cup_j P_j \Leftrightarrow y_i = 1$

Find: Hypothesis $h: \mathbb{R}^2 \to \{0, 1\}$ with small error, **while providing differential privacy for the training data**

Previous Result

Private learner with sample complexity

$\mathrm{O}(k \cdot \log d)$ and runtime $\approx d^k$

(using a generic tool of MT'07)

New Result

Private learner with sample complexity

$\tilde{\mathrm{O}}(k \cdot \log d)$ and runtime $\mathrm{poly}(k, \log d)$

Privately Learning Union of Polygons

POSTER #124

Given: n points in $[d]^2$ with binary labels: $\{(x_i, y_i)\}_{i=1}^n$

Assume: $\exists$collection of polygons $\{P_1, \ldots, P_t\}$ with a total of al most k edges s.t. $\forall i \in [n]$: $x_i \in \cup_j P_j \Leftrightarrow y_i = 1$

Find: Hypothesis $h: \mathbb{R}^2 \to \{0, 1\}$ with small error, **while providing differential privacy for the training data**

Previous Result

Private learner with sample complexity

$O(k \cdot \log d)$ and runtime $\approx d^k$

(using a generic tool of MT'07)

New Result

Private learner with sample complexity

$\tilde{O}(k \cdot \log d)$ and runtime $\mathrm{poly}(k, \log d)$

Summary

- ✓ New algorithm for privately learning union of polygons
- ✓ Efficient runtime and sample complexity
- ✓ Applications to privately analyzing users' location data