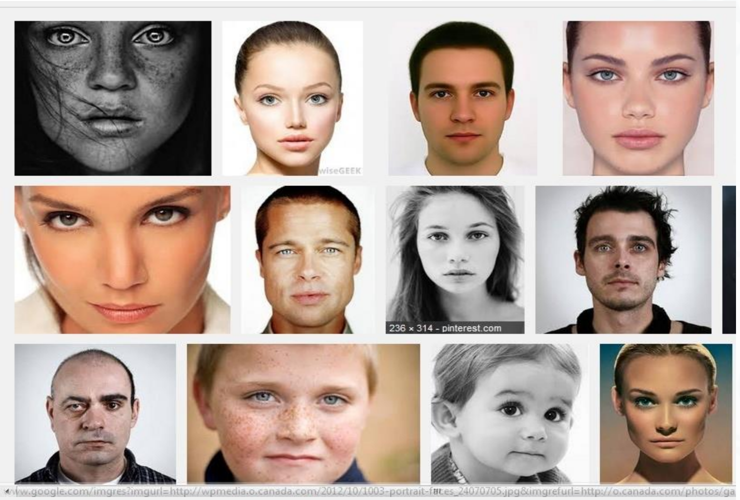
Multiple Instance Learning

Dr. Fayyaz ul Amir Afsar Minhas
PIEAS Biomedical Informatics Research Lab
Department of Computer and Information Sciences
Pakistan Institute of Engineering & Applied Sciences
PO Nilore, Islamabad, Pakistan
http://faculty.pieas.edu.pk/fayyaz/

Binary Classification



Faces (+1)

Non Faces (-1)

+1

-1

-1

+1

Labeling Ambiguities

- What if the labels are **ambiguous**?
  - Part-Whole Ambiguity
  - Polymorphism Ambiguity

Part-Whole Ambiguity

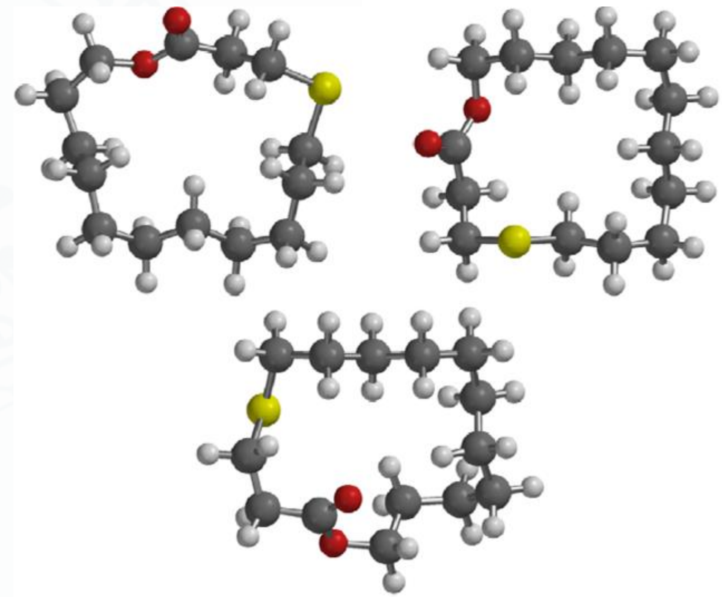
- Label of the object depends on one or more **PARTS** instead of **WHOLE** of the object
- Face Detection Problem

+1

Polymorphism Ambiguity

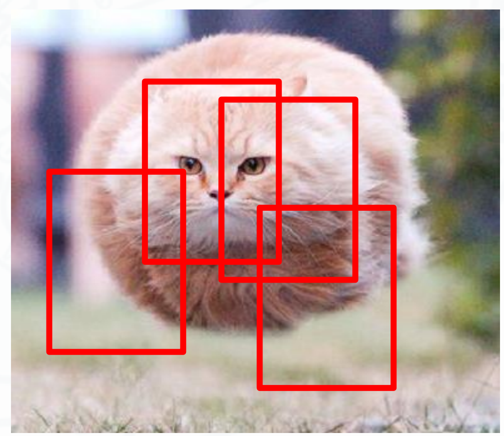
- Label of the object depends on one or more **PARTICULAR FORMS** of all

- Drug Activity Prediction



Multiple Instance Learning

- Allows to learn a classifier when such ambiguities are present
- Use group (bag) level labels instead of instances



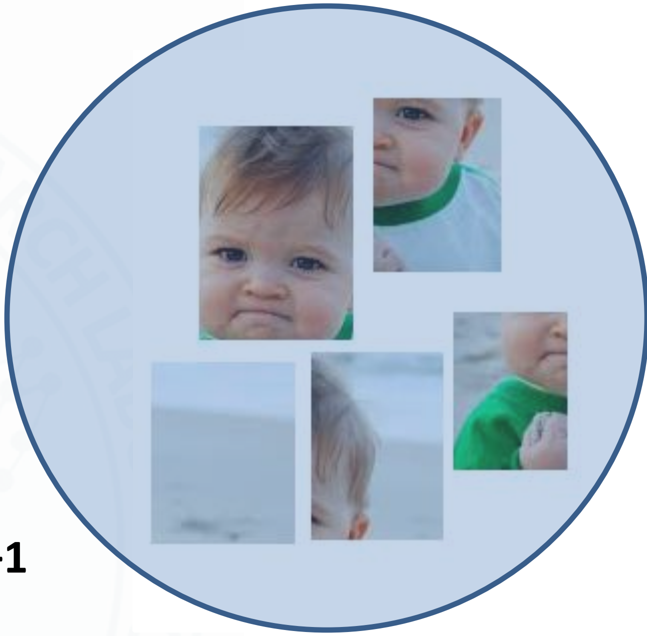
-1

Problem Formulation of MIL- Bags

- A bag is a set of input patterns
- With each bag B_I is associated a label Y_I
  - If $Y_I = -1$, then $y_i = -1$ for all $i \in I$
  - If $Y_I = +1$, then at least one pattern $x_i \in B_I$ is a positive example

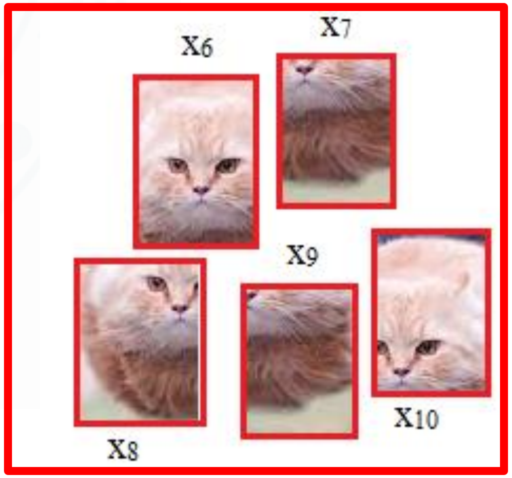
B_1

Y_1=+1



B_2

Y_1=-1



Problem Formulation of MIL

Find a discriminant function f *parameterized by* w *which can generate output values of any given training bag* B_I

$$Y_I = f(\boldsymbol{B_I}; \boldsymbol{w}) \quad \forall I$$

such that for classification the following constraints hold,

$$\sum_{i \in I} \frac{y_i + 1}{2} \geq 1, \qquad \forall\, I \; s.t \;\; Y_I = 1, \qquad and \tag{1}$$

$$y_i = -1 \qquad \forall\, I \; s.t. \; Y_I = -1 \tag{2}$$

where, y_i *is the label for instance* $\boldsymbol{x}_i$ *and* $\boldsymbol{x}_i \in \boldsymbol{B_I}$.

Multiple Instance Learning – The Goal

- To classify the unseen
  - bags
  - instances

Applications

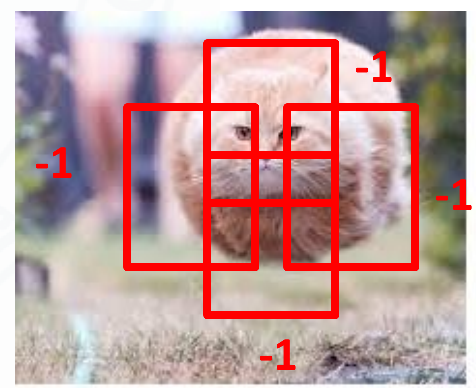
- Drug Activity Prediction
- Protein Interactions
- Computer Aided Diagnostics
- Visual Tracking
- Content-based image retrieval and classification
- Text categorization
- Natural Scene Classification

Naïve Solution to MIL Problem

- Assign the label of the "whole" to the "parts"
- Assign all the appearances the same label
- Use any classification technique
- Classification Accuracy Compromised!

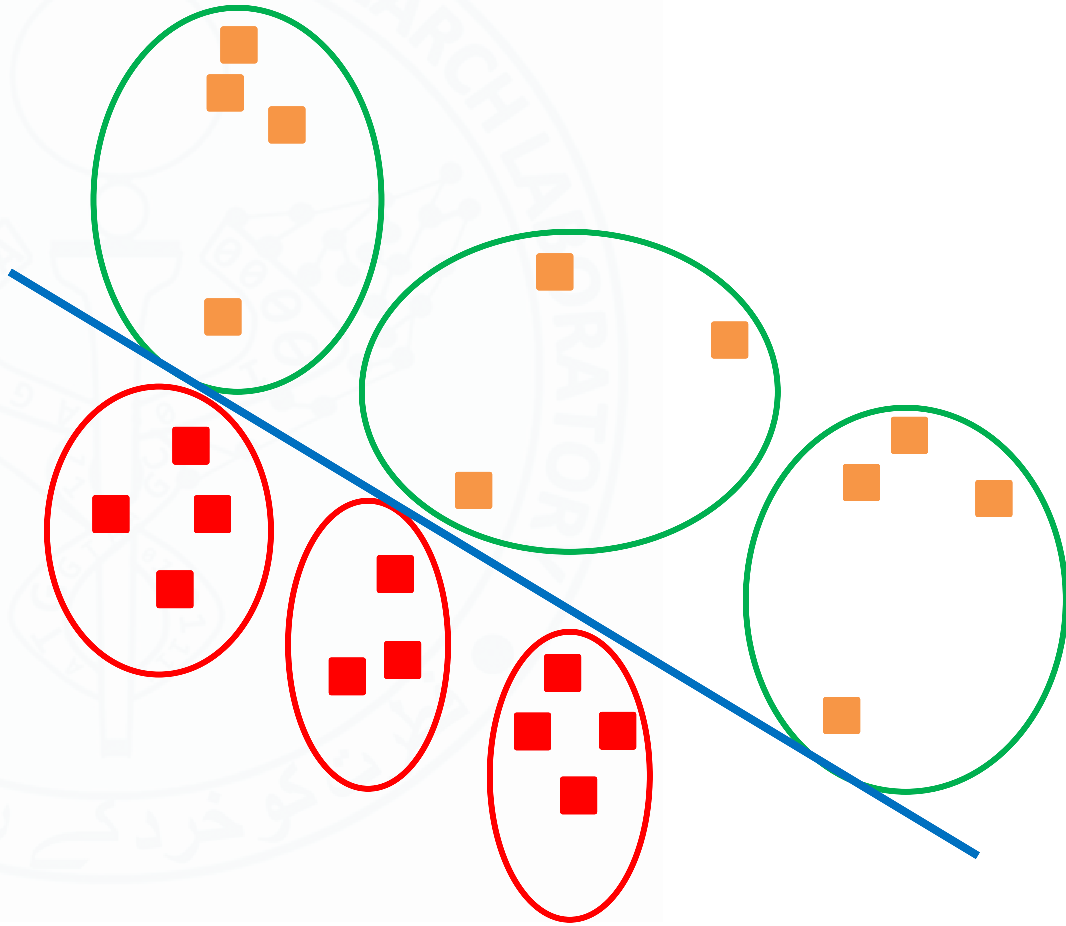
+1

-1



Properties of the Naïve Solution

- Labeling noise introduced in the data
- low margin and poor generalization
- Does not model the relationship between examples in a bag
  - Assumes that examples are independently labeled



Classifiers for MIL

- Learning Axis-Parallel Concepts
- Logistic Regression
- Diverse Density (DD) and its EM version
- Boosting
- Citation kNN
- Support Vector Machines

SVM for MIL

- Proposed by Andrews et al. (2002)

- Two Formulations
  - Maximum Pattern Margin Formulation (mi-SVM)
    - Instance level Margin Maximization
  - Maximum Bag Margin Formulation (MI-SVM)
    - Bag level Margin Maximization

mi-SVM

- Find the optimal labels of the examples in the positive training bags

$$\min_{\{y_i\}} \min_{\boldsymbol{w},\xi} \frac{1}{2} \|\boldsymbol{w}\|^2 + C \sum_i \xi_i$$

$s.t.\quad \forall i:$

$$y_i(\langle \boldsymbol{w}, \boldsymbol{x_i} \rangle + b) \geq 1 - \xi_i,$$

$$\xi_i \geq 0,$$

$$y_i \in \{-1, 1\} \quad and \quad (1,2)\ hold$$

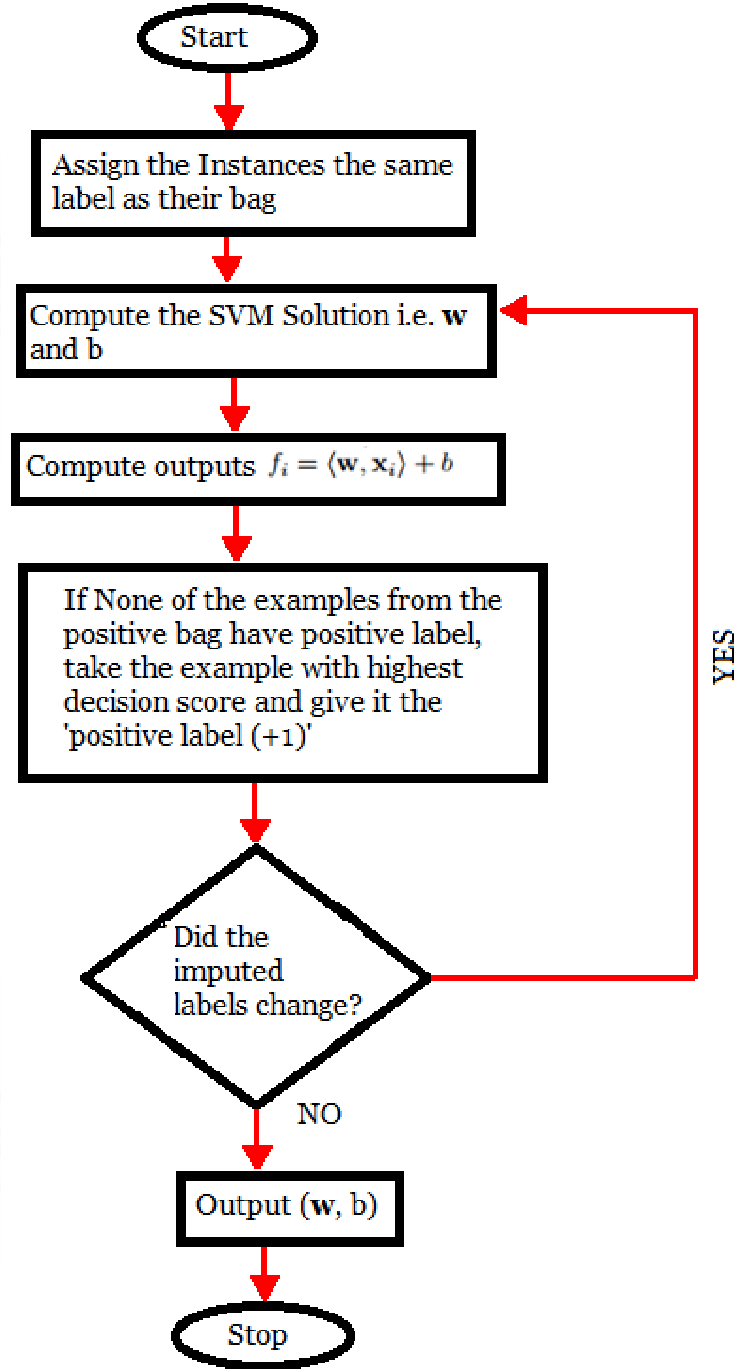
$$\sum_{i \in I} \frac{y_i + 1}{2} \geq 1, \qquad \forall\, I\ s.t\ \ Y_I = 1, \qquad and \tag{1}$$

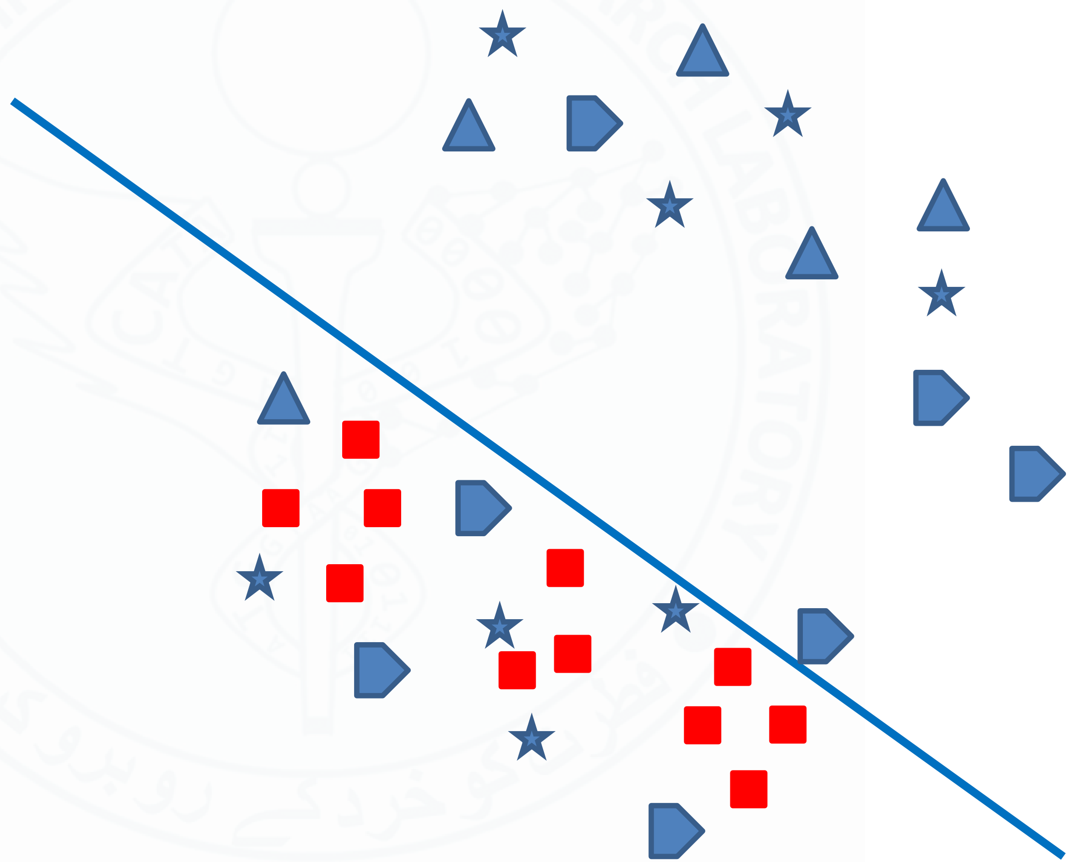
$$y_i = -1 \qquad \forall\, I\ s.t.\ Y_I = -1 \tag{2}$$

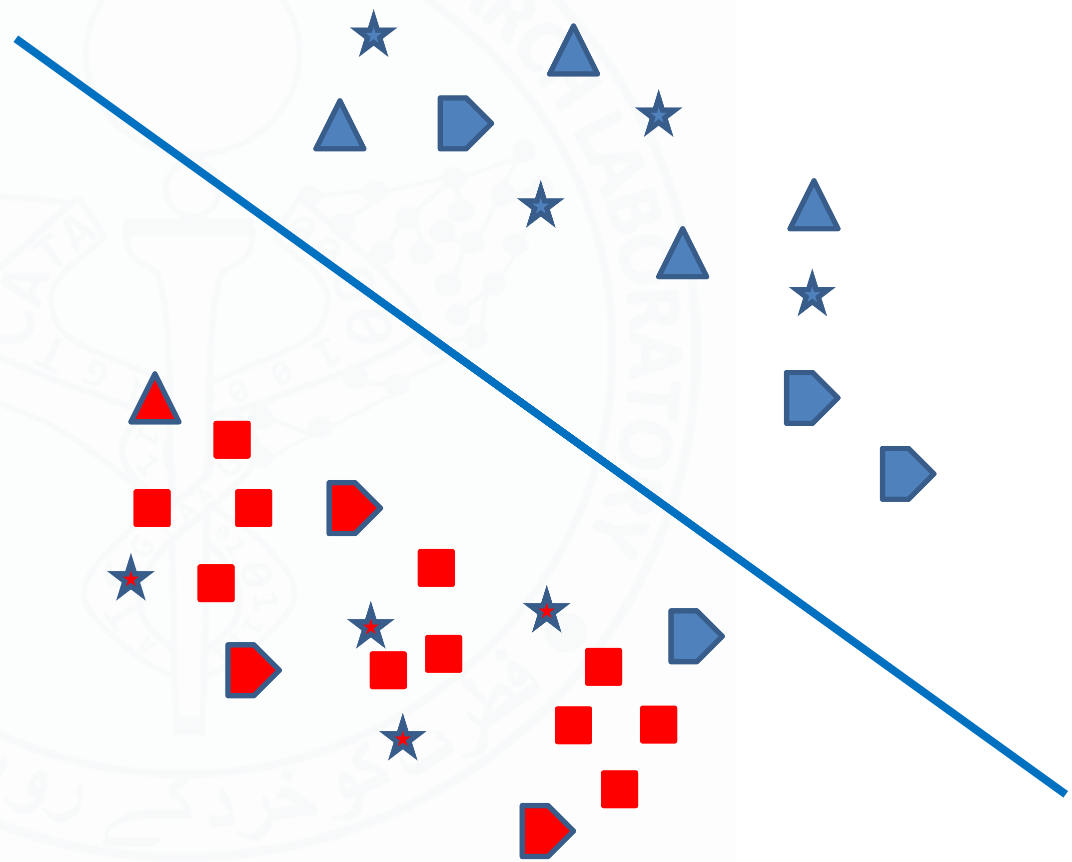
where, y_i *is the label for instance* $\boldsymbol{x}_i$ *and* $\boldsymbol{x}_i \in \boldsymbol{B_I}$.

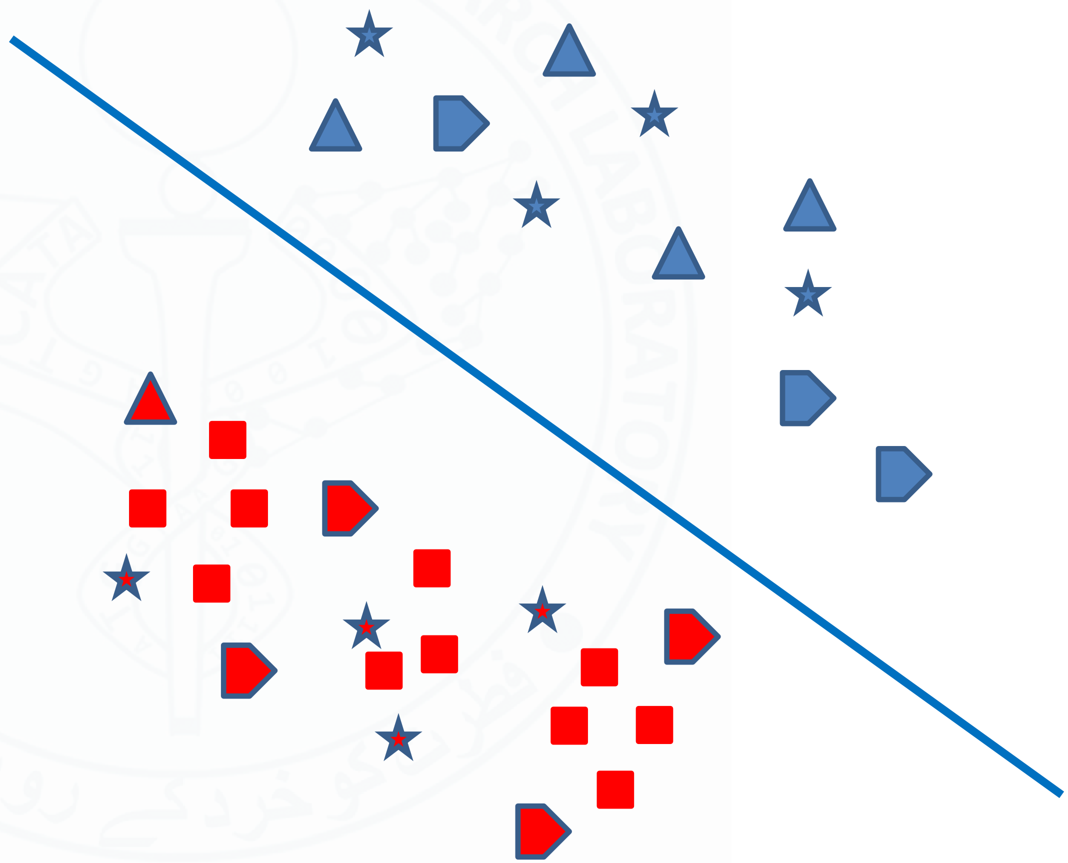
mi-SVM

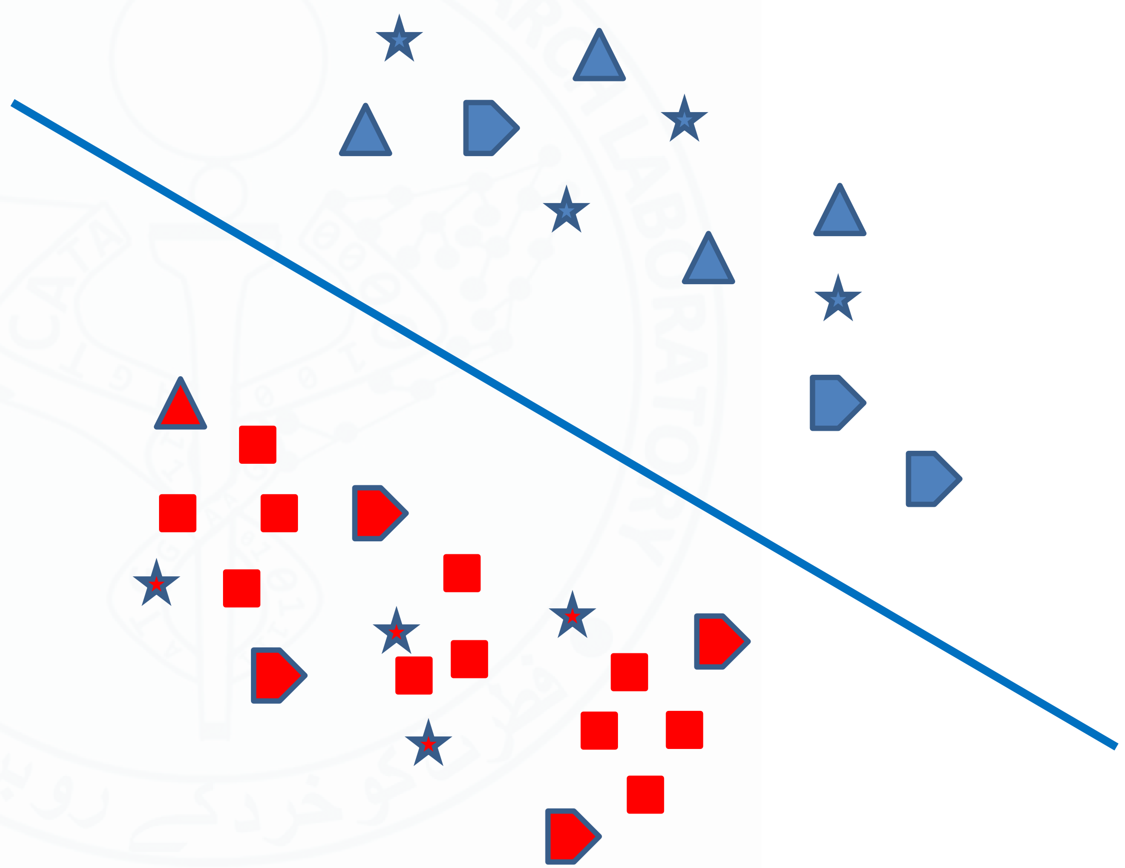
initialize $y_i = Y_I$ for $i \in I$
REPEAT
    compute SVM solution $\mathbf{w}, b$ for data set with imputed labels
    compute outputs $f_i = \langle \mathbf{w}, \mathbf{x}_i \rangle + b$ for all $\mathbf{x}_i$ in positive bags
    set $y_i = \mathrm{sgn}(f_i)$ for every $i \in I$, $Y_I = 1$
    FOR (every positive bag B_I)
        IF $(\sum_{i \in I} (1 + y_i)/2 == 0)$
            compute $i^* = \arg\max_{i \in I} f_i$
            set $y_{i^*} = 1$
        END
    END
WHILE (imputed labels have changed)
OUTPUT $(\mathbf{w}, b)$









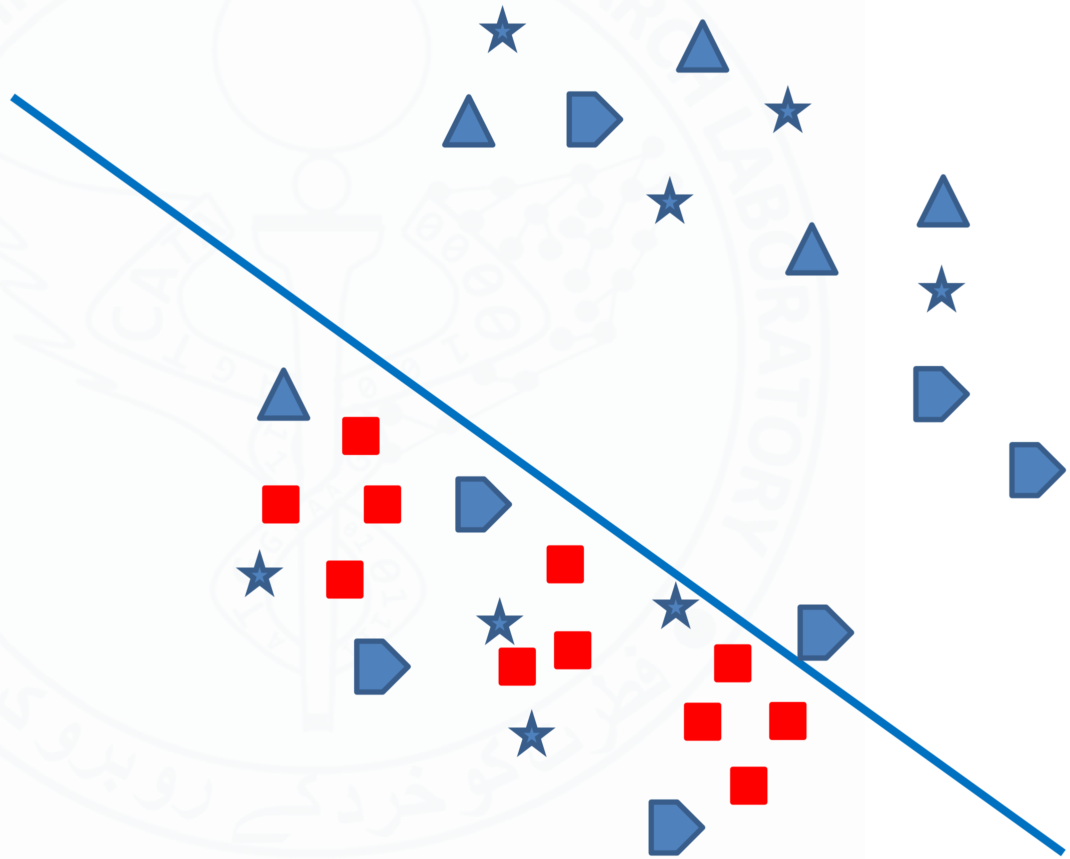


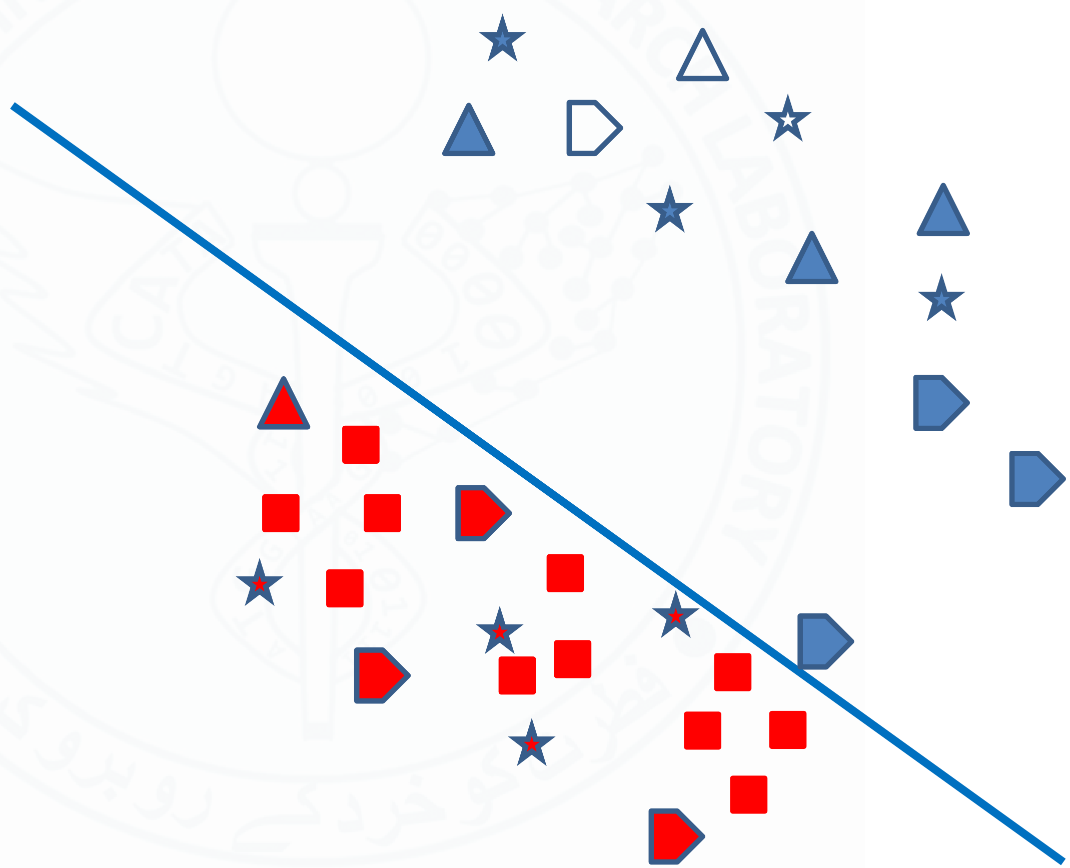
MI-SVM

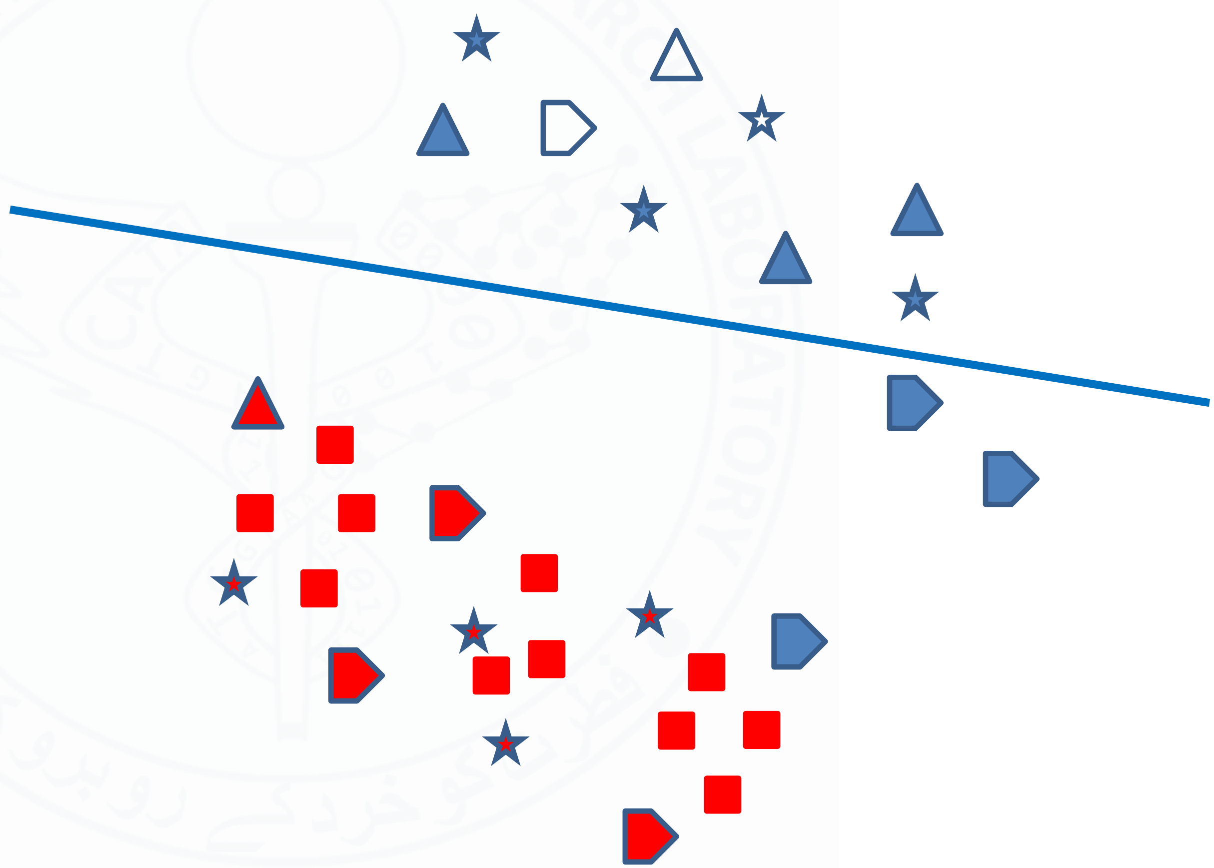
$$\text{MI-SVM} \qquad \min_{\mathbf{w},b,\xi} \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_I \xi_I$$

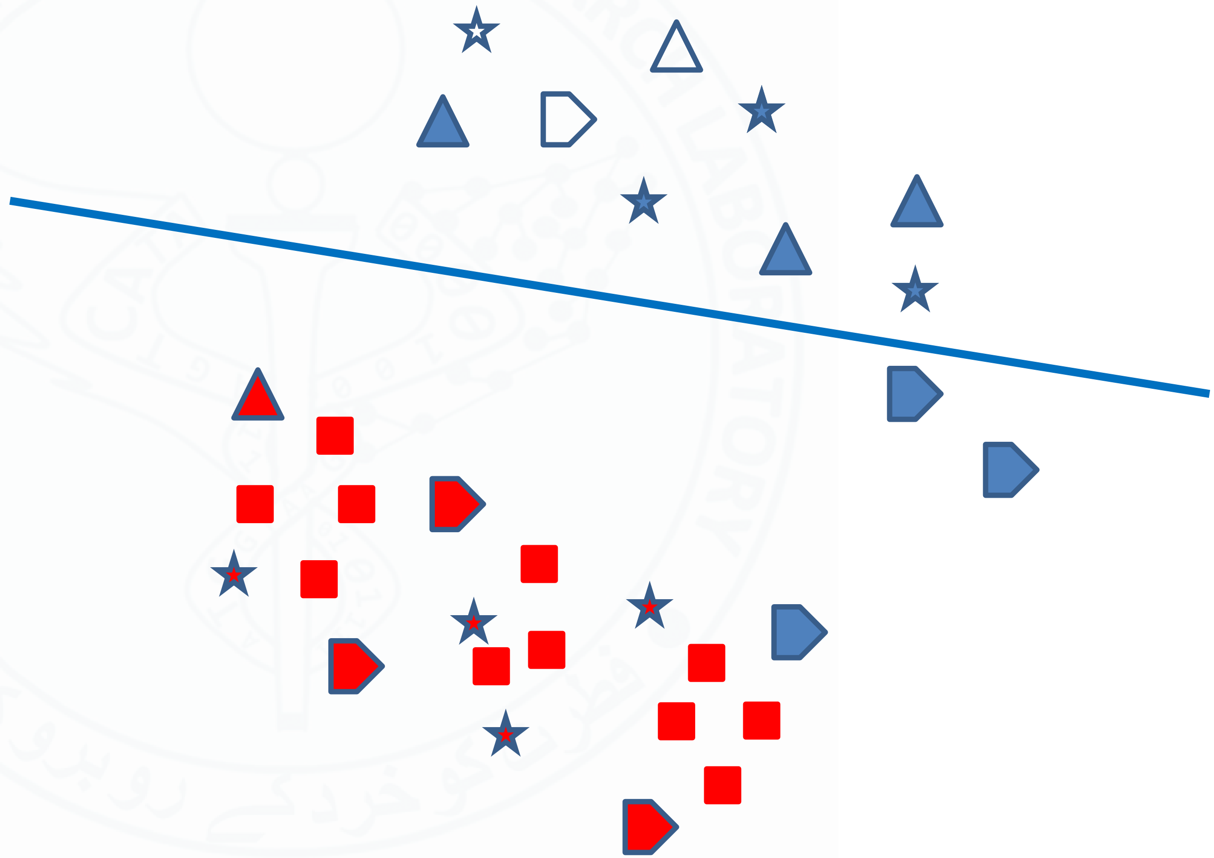
$$\text{s.t.} \quad \forall I: \; Y_I \max_{i \in I}(\langle \mathbf{w}, \mathbf{x}_i \rangle + b) \geq 1 - \xi_I, \; \xi_I \geq 0.$$

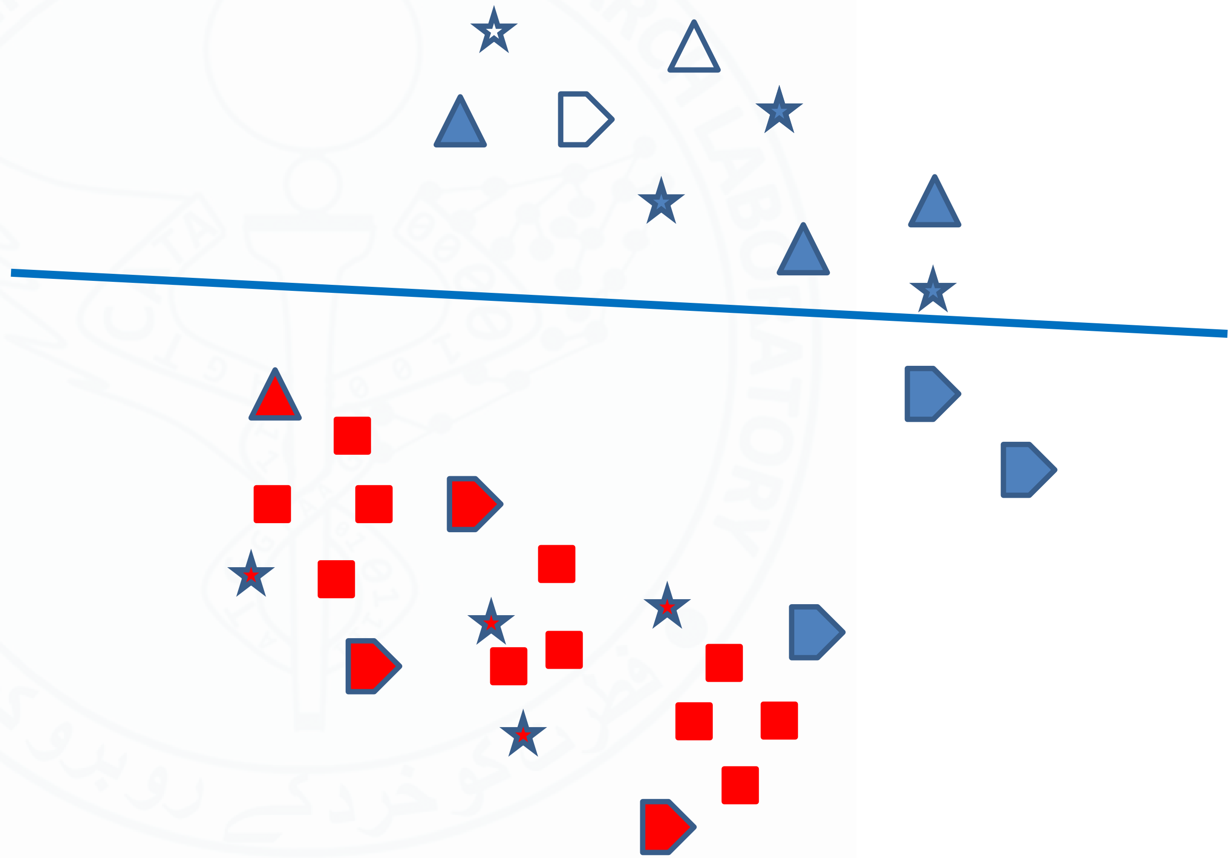
initialize $\mathbf{x}_I = \sum_{i \in I} \mathbf{x}_i / |I|$ for every positive bag B_I
REPEAT
  compute QP solution $\mathbf{w}, b$ for data set with
    positive examples $\{\mathbf{x}_I : Y_I = 1\}$
  compute outputs $f_i = \langle \mathbf{w}, \mathbf{x}_i \rangle + b$ for all $\mathbf{x}_i$ in positive bags
  set $\mathbf{x}_I = \mathbf{x}_{s(I)}$, $s(I) = \arg\max_{i \in I} f_i$ for every I, $Y_I = 1$
WHILE (selector variables $s(I)$ have changed)
OUTPUT $(\mathbf{w}, b)$







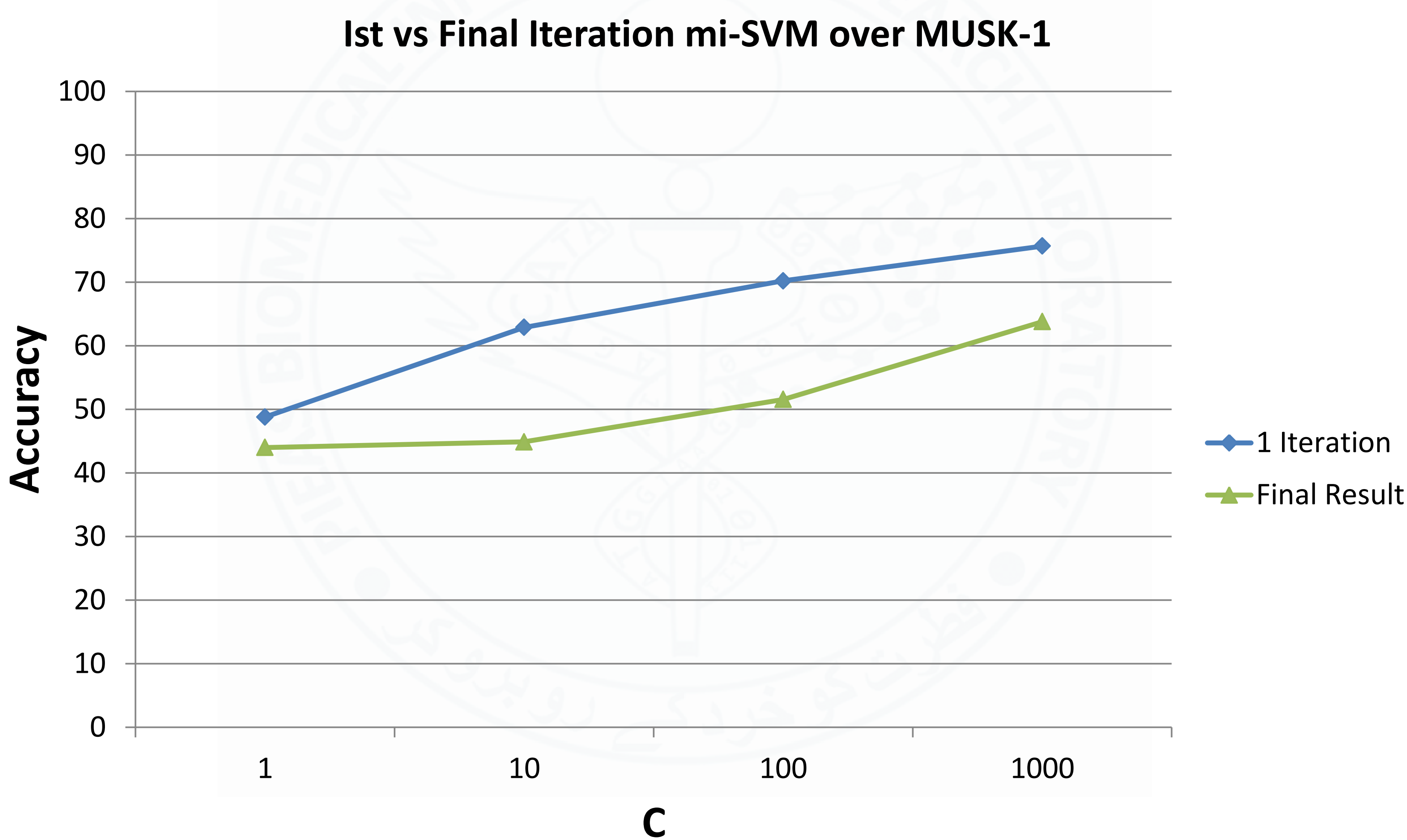




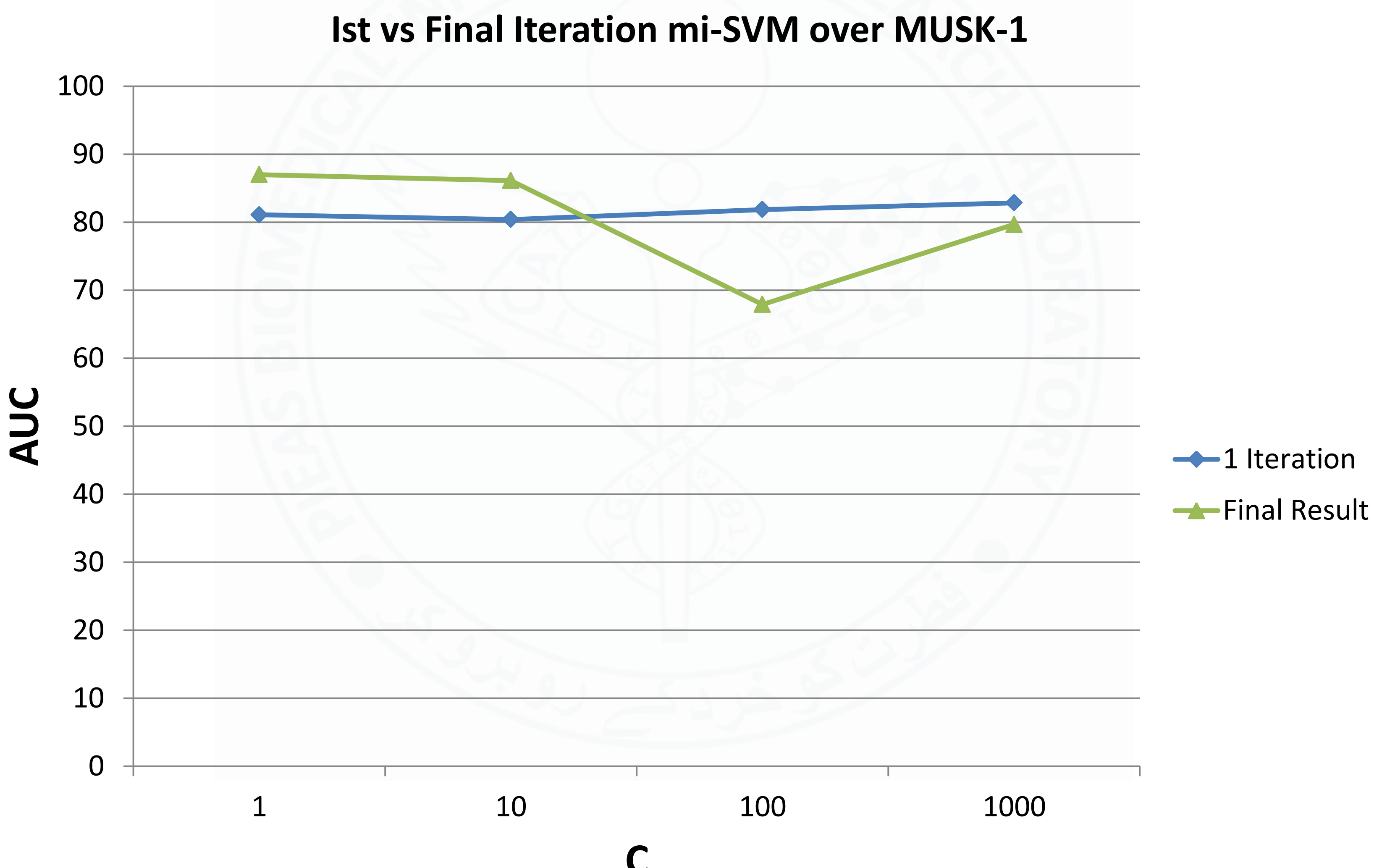
Problems

- mi-SVM and MI-SVM both use local optimization heuristics
  - Can produce sub-optimal solutions
- Require iterative training of the classifier
  - Computationally expensive
- Cannot work for big data

No Guaranteed Optimal Solution



No Guaranteed Optimal Solution



Multiple Instance Ranking

Find a ranking function f parameterized by w which can generate output values for any given training bag B_I

$$Y_I = f(\boldsymbol{B_I}; \boldsymbol{w}) \quad \forall I$$

such that,

$$f(\boldsymbol{B_I}; \boldsymbol{w}) > f\left(\boldsymbol{B_J}; \boldsymbol{w}\right) \quad if \;\; Y_I > Y_J \tag{3}$$

Multiple Instance Regression

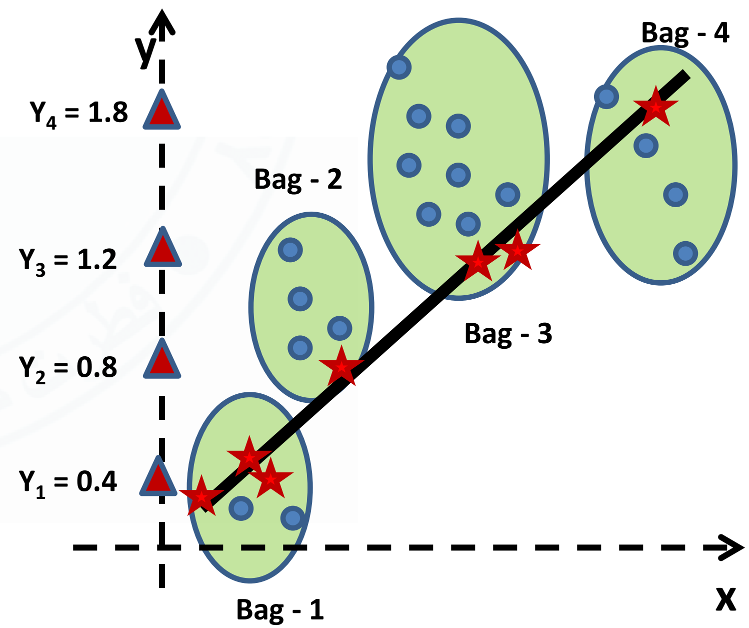
Find a function f parameterized by w which can generate output values of any given training bag B_I

$$Y_I = f(\boldsymbol{B_I}; \boldsymbol{w}) \quad \forall I$$

such that,

$$|Y_I - f(\boldsymbol{B_I}; \boldsymbol{w})| \leq \epsilon \qquad (4)$$

Where, $\epsilon \geq 0$ is the maximum error allowed.



pyLEMMINGS

- **PY**thon based Larg**E** **M**argin **M**ultiple **I**nstance learni**NG** **S**ystem
  - Object Oriented Implementation for
    - Linear and Locally Linear SSGO Multiple Instance Learning
  - Built-in module for
    - K-fold Cross Validation
    - Leave One Out Cross Validation
  - Support for sparse data format
  - Support for Parallelization for Cross Validation module
  - User-friendly implementation
  - Help for users and maintenance

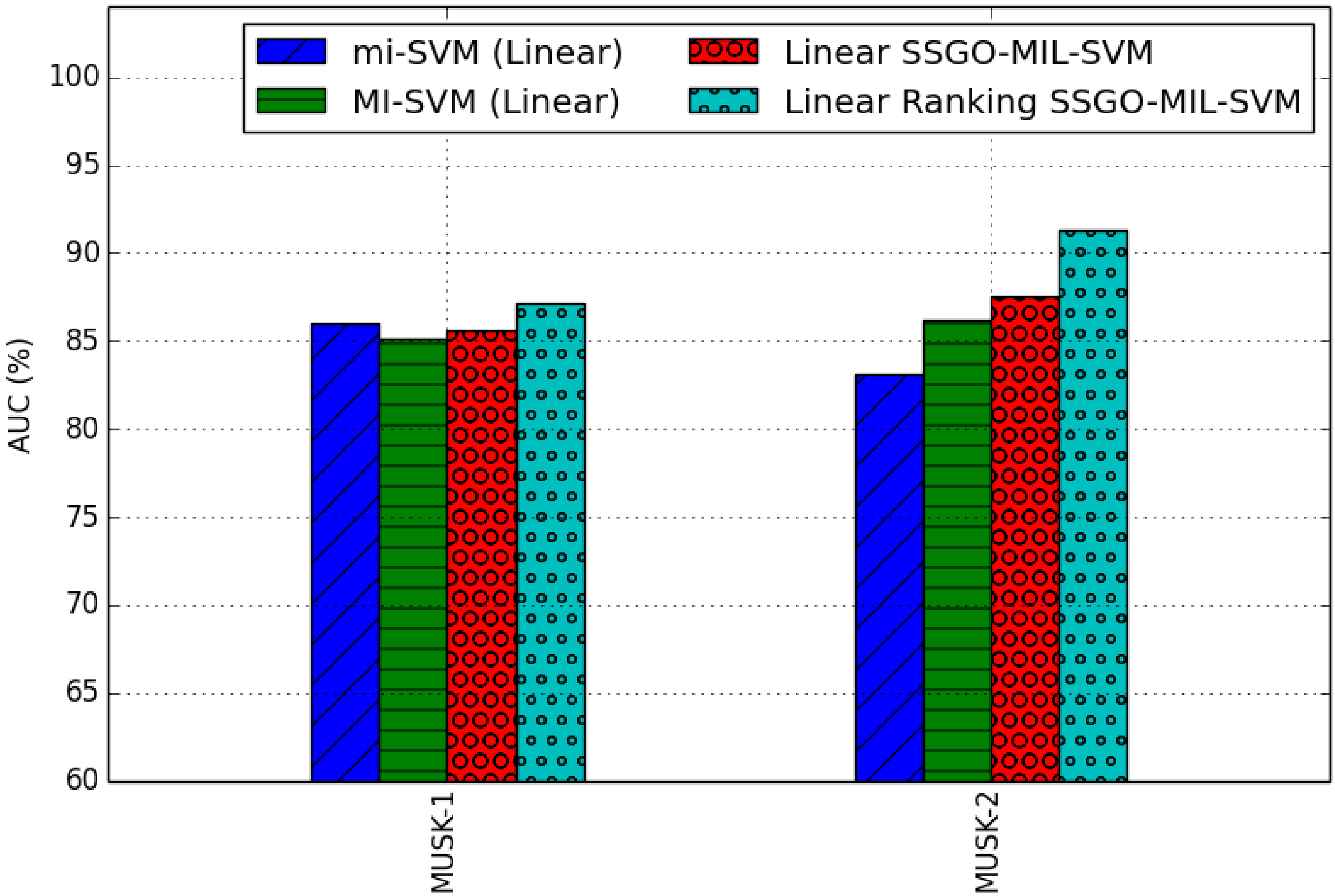
Benchmark Results

- Datasets

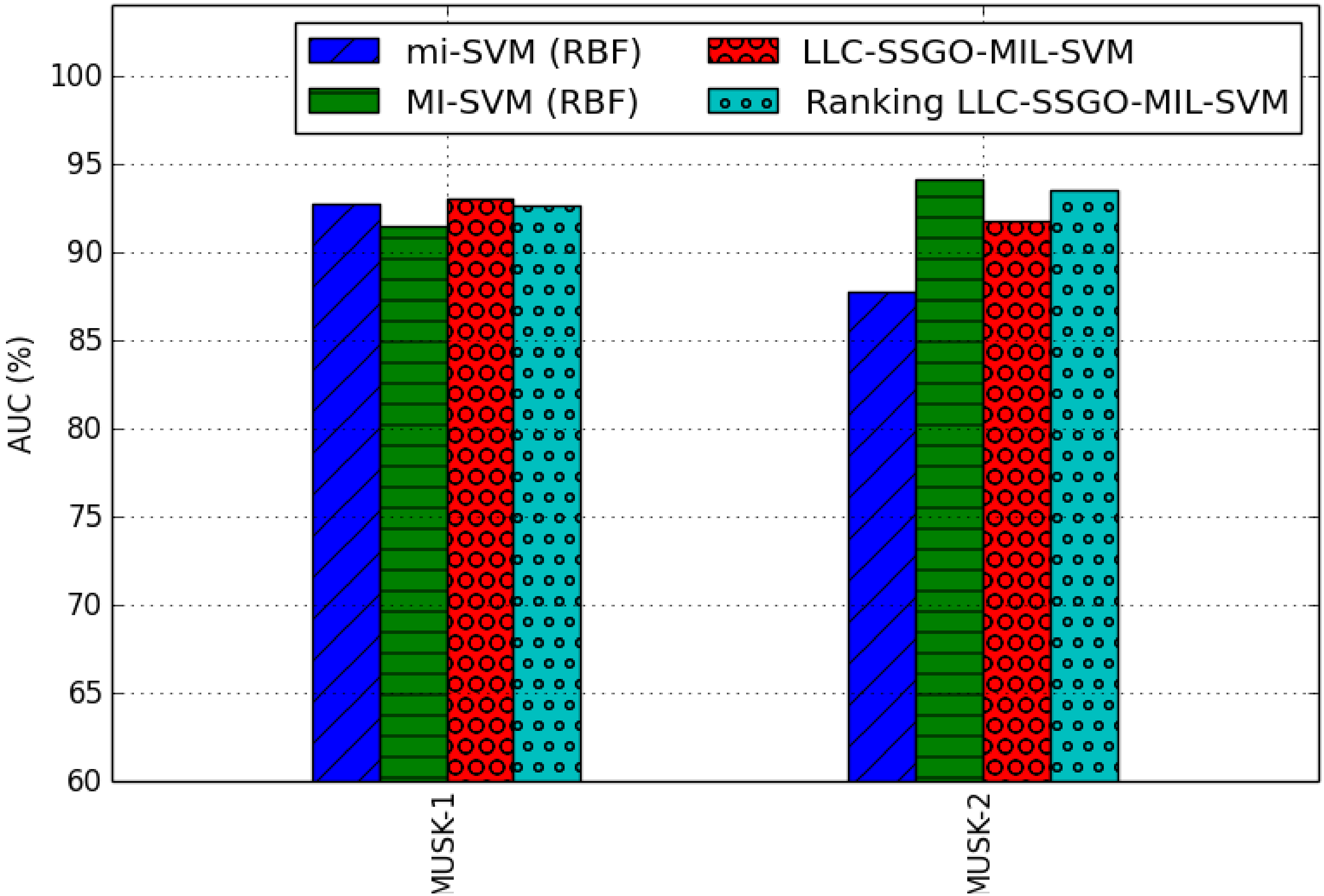
Dataset	Instances	Bags	Positive Bags	Negative Bags	Dimensions
MUSK-1	476	92	47	45	166
MUSK-2	6598	102	39	63	166

- Evaluation Model
  - Ten Fold Cross Validation
- Performance Metrics
  - AUC
  - Running Time

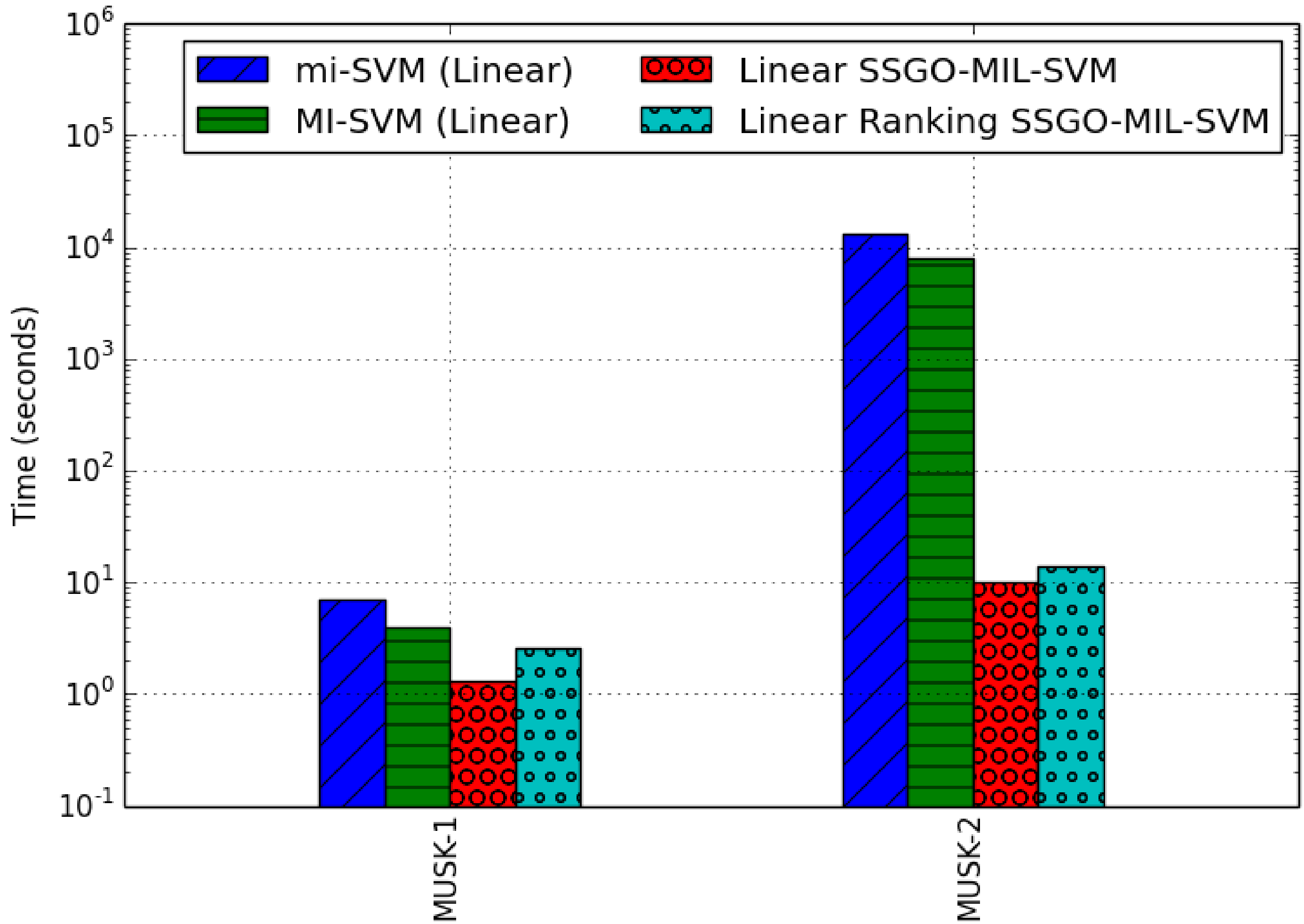
Benchmark Results- AUC



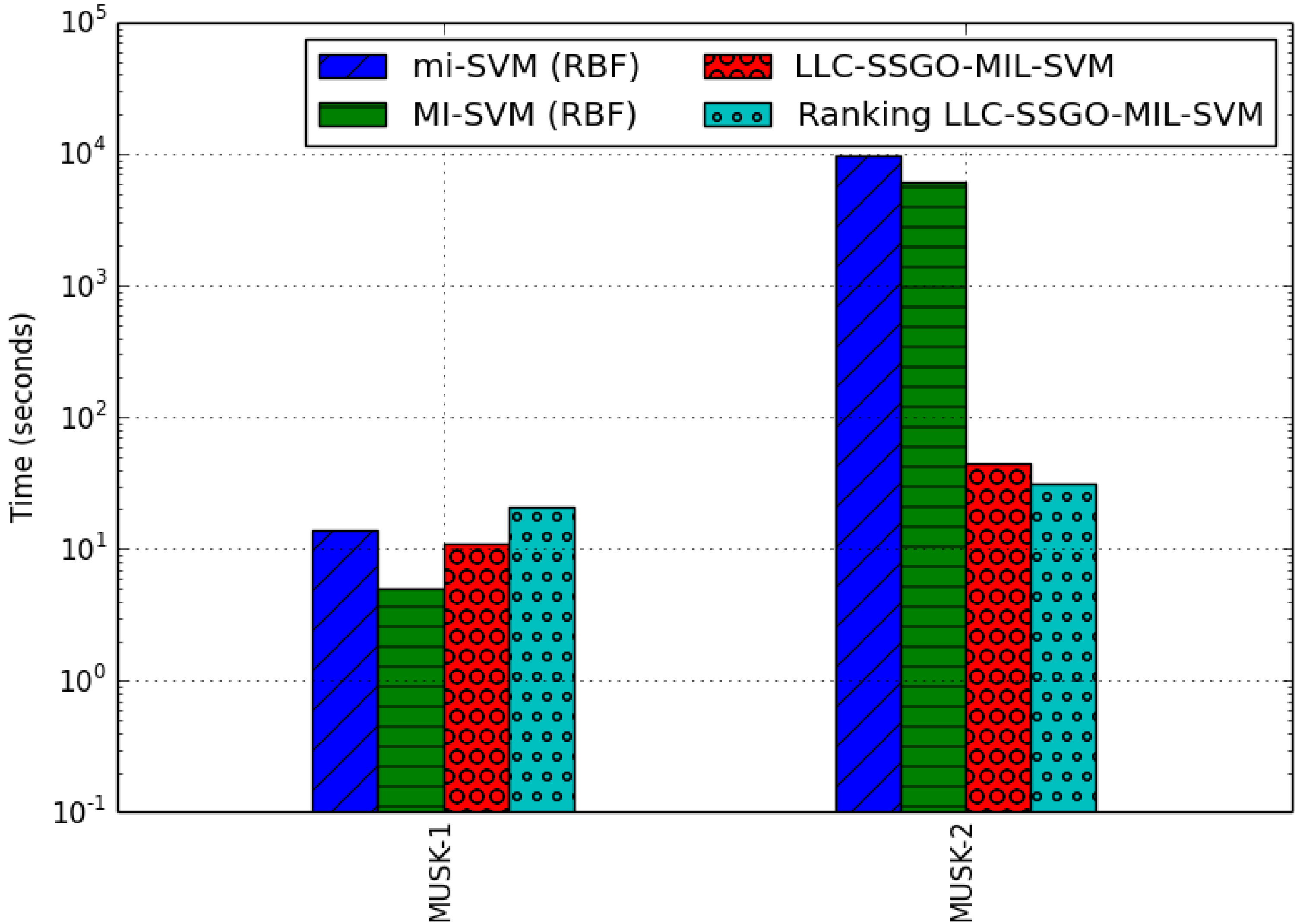
Benchmark Results- AUC



Benchmark Results- Running Time



Benchmark Results- Running Time



Applications

Method	Features	AUC	$AUC_{0.1}$	TH %	FH %
Vanilla SVM	1-Spec	95.5	53.9	**66**	2.6
	PD-1	95.6	54.5	64	2.5
	Comb.	95.9	55.1	65	2.1
	Max. Std.	0.16	0.59	2.2	0.15
mi-SVM	1-Spec	95.5	**54.4**	64	2.6
	PD-1	96.0	55.8	69	2.1
	Comb.	96.2	55.6	68	1.9
MI-1 SVM	1-Spec	**96.0**	54.3	62	**2.1**
	PD-1	**96.8**	**58.5**	**72**	**1.3**
	Comb.	**96.9**	**59.0**	**75**	**1.2**
	Max Std.	0.14	0.80	3.4	0.11
	Gappy	96.5	58.5	68	1.6

- Andrews, Stuart, Ioannis Tsochantaridis, and Thomas Hofmann. 2003. “Support Vector Machines for Multiple-Instance Learning.” In *Advances in Neural Information Processing Systems 15*, 561–68. MIT Press.
- Leistner, Christian, Amir Saffari, and Horst Bischof. 2010. “MIForests: Multiple-Instance Learning with Randomized Trees.” In *Computer Vision – ECCV 2010*, edited by Kostas Daniilidis, Petros Maragos, and Nikos Paragios, 29–42. Lecture Notes in Computer Science 6316. Springer Berlin Heidelberg. http://link.springer.com/chapter/10.1007/978-3-642-15567-3_3.
- https://en.wikipedia.org/wiki/Multiple-instance_learning

We want to make a machine that will be proud of us.

- Danny Hillis