Efficient Semi-supervised and Active Learning of Disjunctions

Maria-Florina Balcan, Christopher Berlind, Steven Ehrlich, and Yingyu Liang

Georgia Institute of Technology

ICML, Jun 19th, 2013

Modern Challenge for Learning Paradigm

Passive Supervised Learning

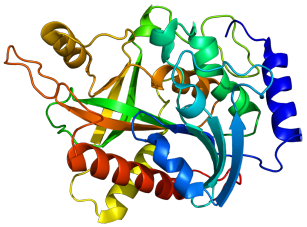
- Given labeled examples, find function that correctly labels future examples

face

car

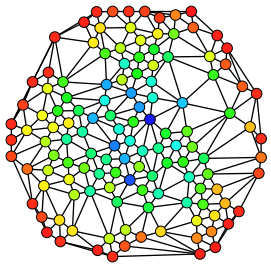
Classic paradigm insufficient nowadays

- Massive amounts of unlabeled data
- Only small fraction can be labeled



protein sequences

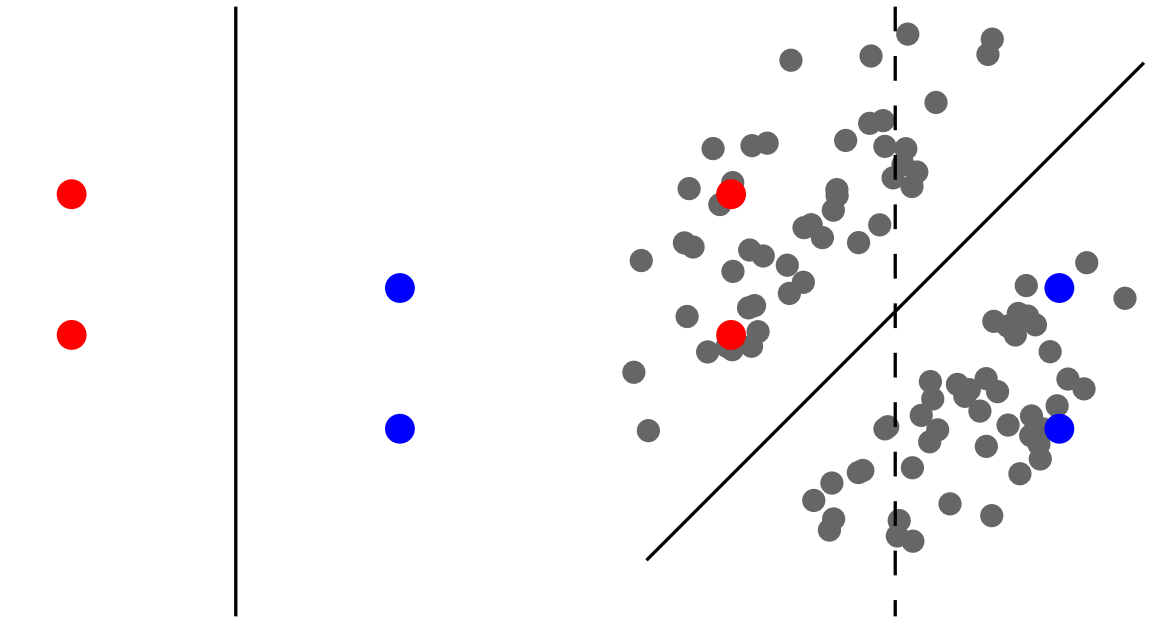
astronomical data



social networks

How Unlabeled Data Helps

Common assumption: large margin



Labeled data only

Labeled and unlabeled

Two-Sided Disjunctions

n boolean features: positive, negative, and non-indicators

x_1 x_2 x_3 x_4 x_5 x_6 x_7 x_8 x_9 x_{10}

Two-Sided Disjunctions

n boolean features: positive, negative, and non-indicators

Training examples labeled according to contained indicators

- Every example has an indicator
- No example has conflicting indicators

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	label
1	1	0	0	0	0	0	0	0	0	+
0	0	1	1	0	0	0	0	1	0	+
0	0	0	0	0	0	1	1	1	1	−
0	0	0	0	1	1	1	0	0	1	−

Two-sided Disjunctions
Example

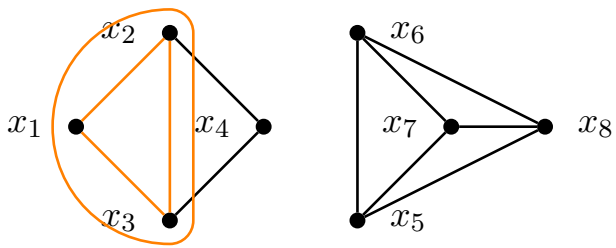
Imagine to distinguish two languages we don't speak...

Words are features, documents are examples

Easy Case: No Non-indicators

Features set to 1 in an example are indicators of the same type

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	label
1	1	1	0	0	0	0	0	?
0	1	0	1	0	0	0	0	?
0	0	1	1	0	0	0	0	?
0	0	0	0	0	0	1	1	?
0	0	0	0	1	1	1	0	?
0	0	0	0	0	1	1	1	?
0	0	0	0	1	0	0	1	?



General case: With Non-indicators

Our results for this open problem:

- efficient active algorithm
- efficient semi-supervised algorithms

Connection to margin assumptions:

- $L_\infty L_1$ margin $\geq O(\frac{1}{\#\text{non-indicators}})$
- Different from $L_2 L_2$ (Perceptron) or $L_1 L_\infty$ (Winnow) margin