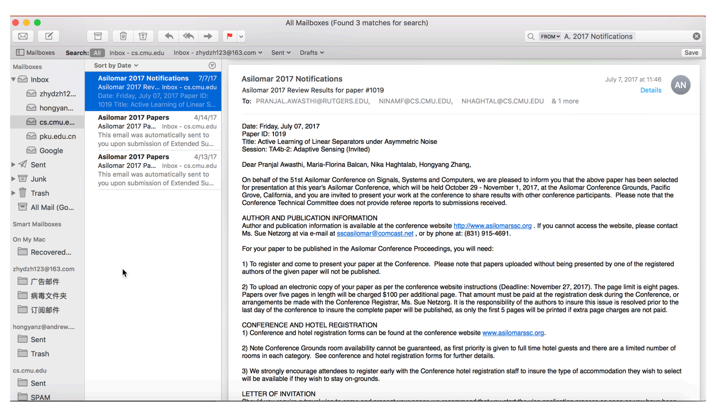
Sample and Computationally Efficient Learning Algorithms under S-Concave Distributions

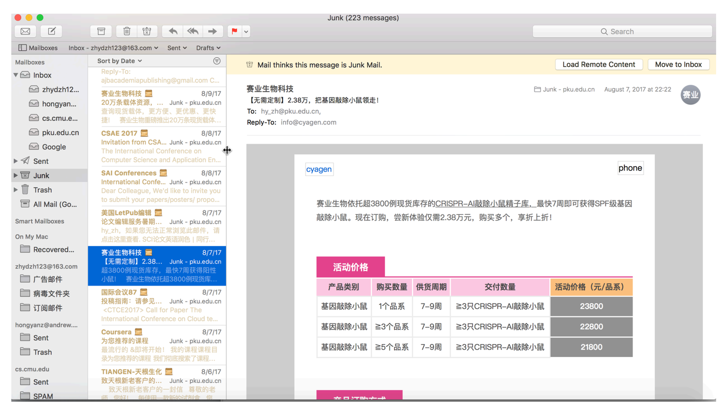
Maria-Florina Balcan Hongyang Zhang
Machine Learning Department, Carnegie Mellon University

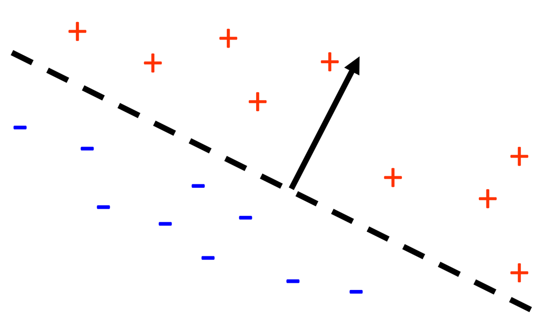
Hardness of Learning Algorithms

Learning algorithms are ubiquitous:

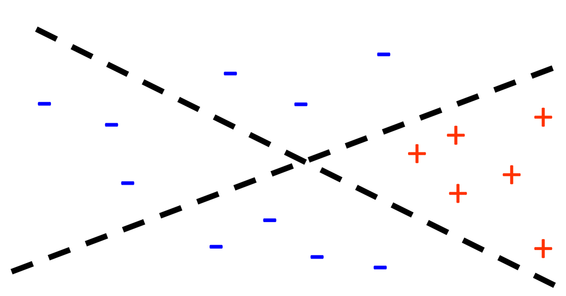
Spam Not Spam







NP-hard when there is noise and no assumption is made on the underlying distribution



NP-hard when there is no assumption on the underlying distribution

Assumption on the underlying dist.:

Uniform Dist. ⇒ Log-concave Dist. ⇒ General Dist.?

(The logarithm of density is concave)

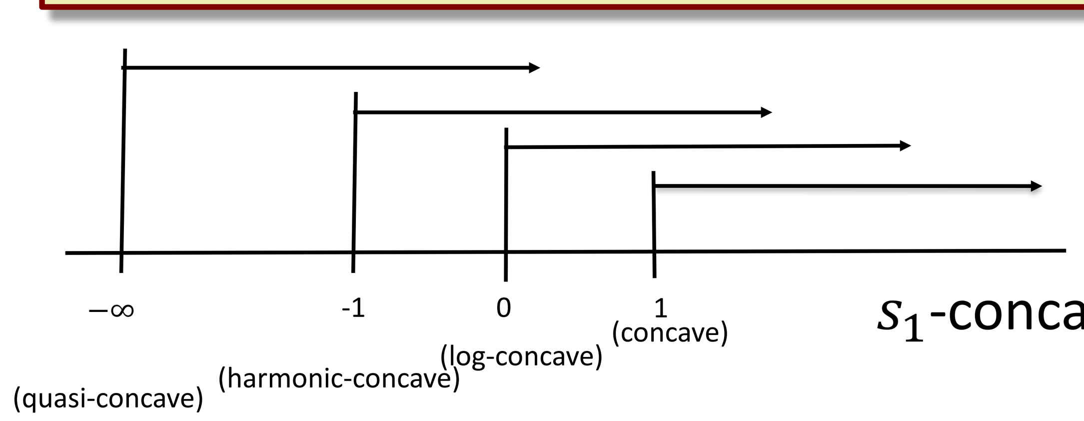
Can we learn halfspace and intersection of halfspaces in polynomial time, when the class of underlying distribution goes beyond the log-concave distributions?

The answer is affirmative!

S-Concave Distributions

S-Concave Distributions:

A function $f(x)$ is s-concave, if $f(x)^s$ is a concave function. A distribution D is s-concave, if its density is s-concave.



$s_1 < s_2$

s_1-concave $\subset$ s_2-concave

Examples of s-concave distributions:

- Cauchy distribution
- Pareto distribution
- *t*-distribution ...

(All are heavy-tailed here)

Structural Results

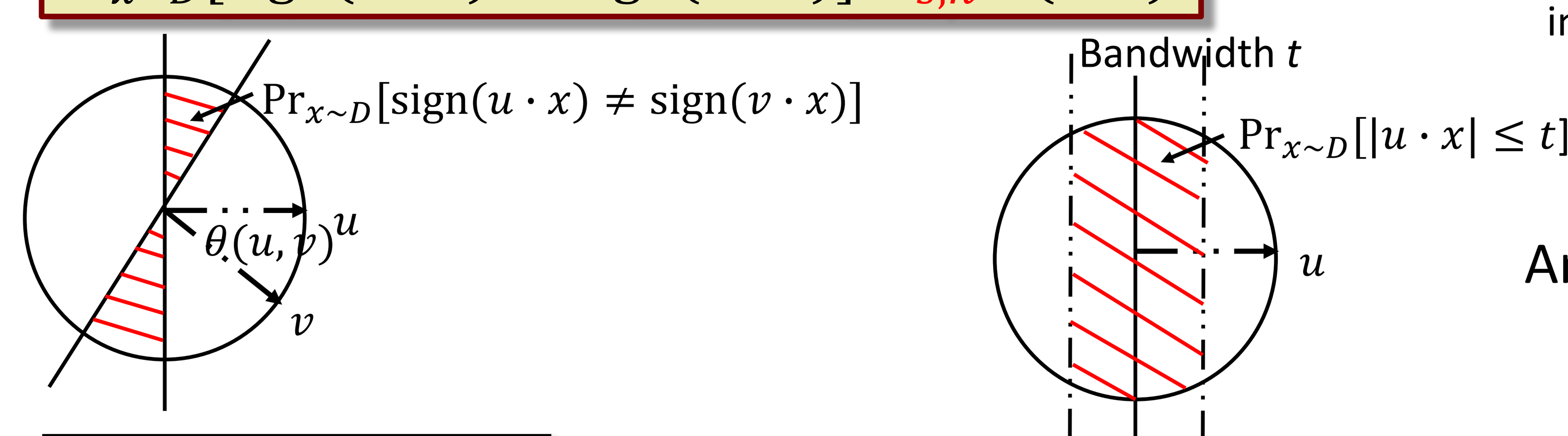
Weakly Closed under Marginal:

The marginal of isotropic *s*-concave distribution over *m* arguments is isotropic $\frac{s}{1+ms}$ - concave.

- Special case: The marginal of log-concave remains log-concave

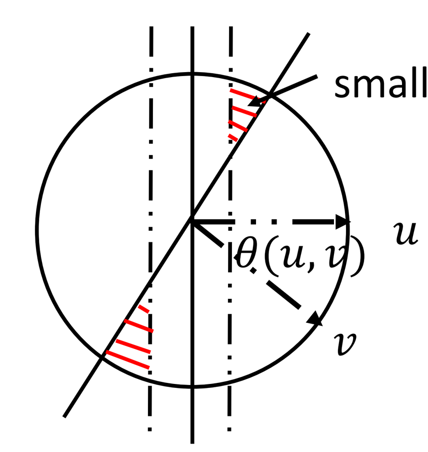
Hyperplane Disagreement:

For any unit vectors $u, v \in \mathbb{R}^n$,
$\Pr_{x\sim D}[\text{sign}(u\cdot x) \neq \text{sign}(v\cdot x)] \propto_{s,n} \theta(u,v)$.



Probability of Band:

For any unit vectors $w \in \mathbb{R}^n$,
$\Pr_{x\sim D}[|u\cdot x| \le t] \propto_{s,n} t$.



Disagreement outside Margin:

For any unit vectors $u, v \in \mathbb{R}^n$,
$\Pr_{x\sim D}[\text{sign}(u\cdot x) \neq \text{sign}(v\cdot x),\ |v\cdot x| \gtrsim_{s,n} \theta(u,v)] \lesssim_{s,n} \theta(u,v)$.

Tail Probability:

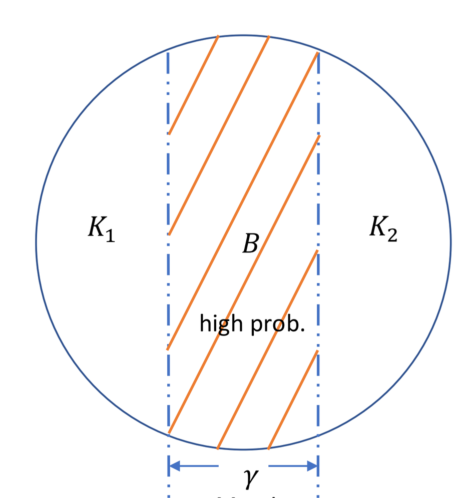
We have $\Pr_{x\sim D}[\|x\| > \sqrt{n}t] \le \left[1 - \frac{cst}{1+ns}\right]^{(1+ns)/s}$

- Heavy-tailed distribution
- Special case: When $s \to 0$, $\Pr_{x\sim D}[||x|| > \sqrt{n}t] \le \exp(-ct)$. (light-tailed dist.)

Intuition and Proof Outline

Isoperimetry (s-concave):

If two sets K_1 and K_2 are well-separated, then the area B between them has large measure *relative to the measure of the two sets.*

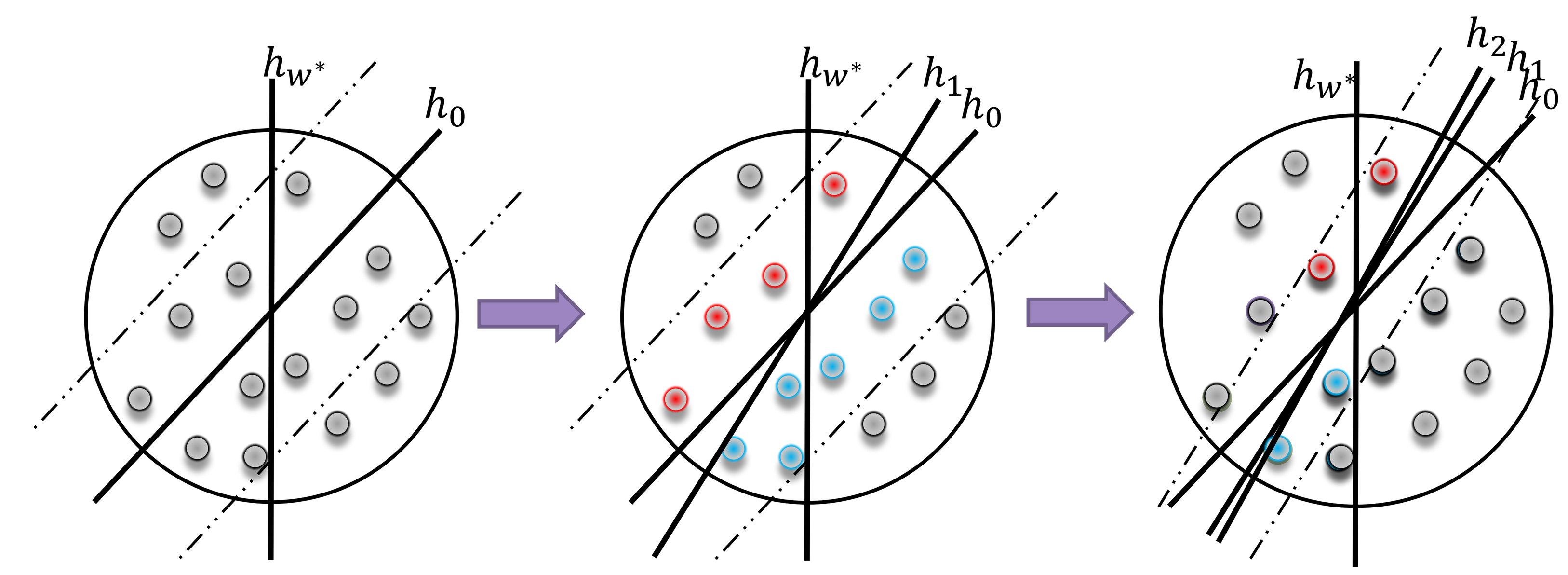


n-D *s*-concave → 1-D γ-concave → 1-D $h(t) = \alpha(1+\beta t)^{1/\gamma}$

Extension of Prekopa-Leindler

Baseline Function

Applications in Active Learning



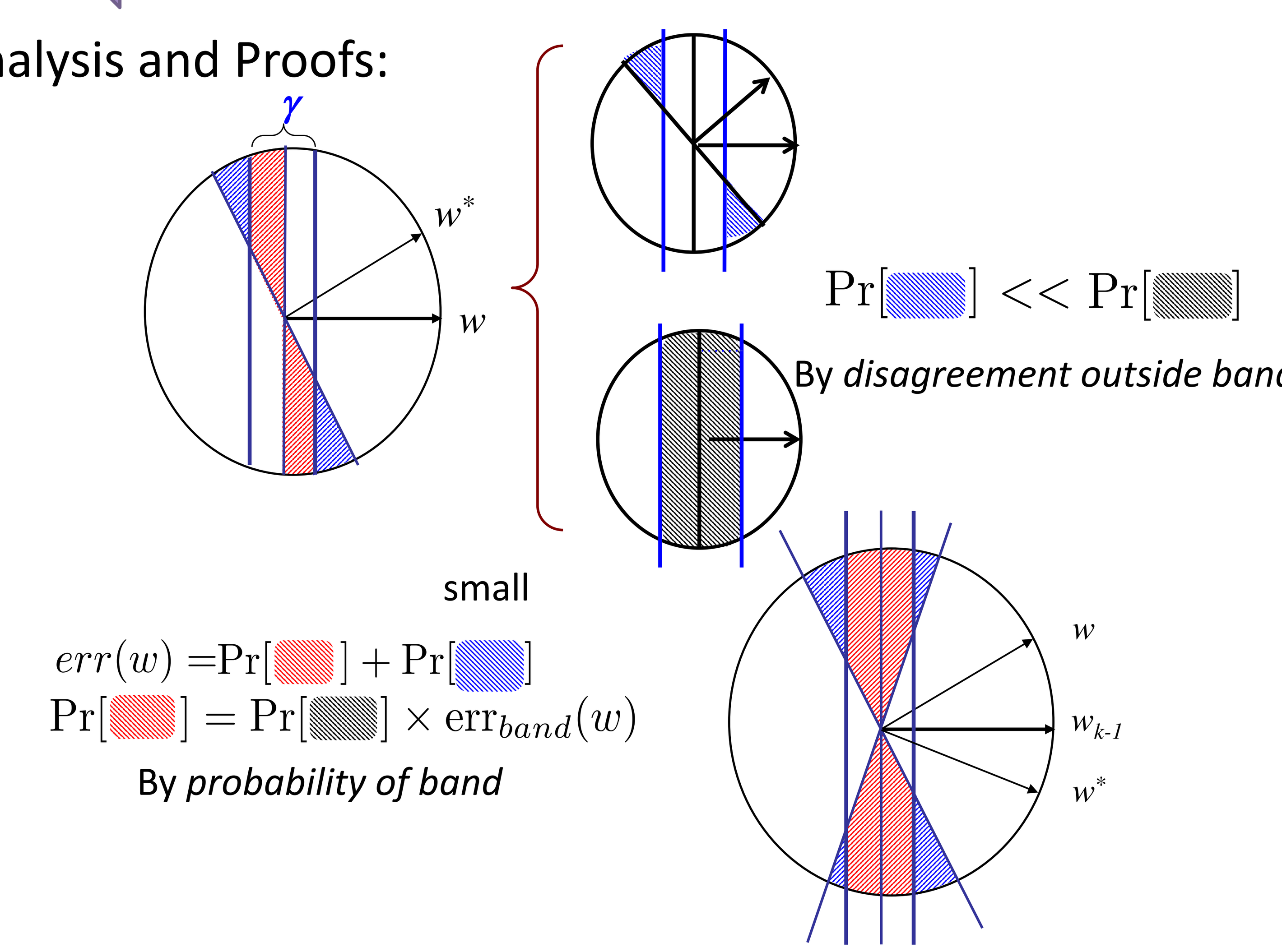
Sample unlabeled data and have an initial guess

Request labels in the band, do hinge loss minimization to constant error

Halve the bandwidth around h_1, request labels in the band, do hinge loss minimization to constant error

Repeat $\log(1/\varepsilon)$ times

Analysis and Proofs:



Theorem 1 (Margin-Based Active Learning):

Label Complexity: poly$_s$(*n*, log(1/*δ*), log(1/*ε*))
Guarantee: $OPT + \varepsilon$

Applications in Baum's Algorithm

Baum's Algorithm under s-concave distributions:

Baum's algorithm succeeds under s-concave dist. if the samples are large (depend on *s*).

- Key observation:
  $\mu_D(E) \approx_s \mu_D(-E)$