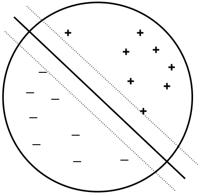
Algorithms and hardness results for parallel large-margin learning

Phil Long (Google) and Rocco Servedio (Columbia)

Setting:



- Target: γ-separated halfspace
- Domain: unit ball in $\mathbf{R}^n$
- Computational model: PRAM (parallel RAM)
- Question: can output an ϵ-accurate hypothesis using
  - $\mathrm{poly}\left(\log n, \log \frac{1}{\gamma}, \log \frac{1}{\epsilon}\right)$ time
  - $\mathrm{poly}\left(n, \frac{1}{\gamma}, \frac{1}{\epsilon}\right)$ processors?

Positive result

Dependence on $1/\epsilon$ already handled by Freund (boost-by-majority). Revised goal:

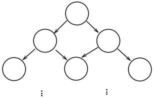
- $\text{poly}\left(\log n, \log \frac{1}{\gamma}\right)$ time
- $\text{poly}\left(n, \frac{1}{\gamma}\right)$ processors.

Algorithm	Number of processors	Running time
Perceptron	$\text{poly}(n, 1/\gamma)$	$\tilde{O}(1/\gamma^2)(\log n)$
SmoothBoost	$\text{poly}(n, 1/\gamma)$	$\tilde{O}(1/\gamma^2)(\log n)$
LP	1	$\text{poly}(n, \log(1/\gamma))$
This paper	$\text{poly}(n, 1/\gamma)$	$\tilde{O}(1/\gamma) + O(\log n)$

Algorithm

- Parallel boost-by-majority to handle ϵ-dependence (Freund)
- Weak learner:
  - Random projection (Johnson/Lindenstrauss, Arriaga/Vempala)
  - Interior point method (Renegar)
    - Compute Hessian using parallel matrix inversion (Reif)
    - Round intermediate solutions (preserve margin)

Negative result



- Some boosters [KM95,MM00,KS02,LS05,LS08] use decision trees and branching programs.
- Calls to weak learners from the same “layer” parallelizable.
- Q: Can save iterations?
- A: No (so $\Omega(1/\gamma^2)$ iterations needed)
- Proof sketch:
  - variables conditionally independent given label
  - in stage *i*, give variable *i* to all weak learners.