Efficient Learning of Halfspaces in the Presence of Bounded Noise

COLT 2015

Pranjal Awasthi, Nina Balcan, Nika Haghtalab, **Ruth Urner**

Main result

There is an algorithm that learns linear separators in polynomial time ($\mathrm{poly}(d, \frac{1}{\epsilon})$) to arbitrarily small excess error for distributions with uniform marginal on the unit ball satisfying the Massart noise condition (bounded by some constant η).

Computational Learning Theory

Efficient algorithms:

- Uniform classification noise

  Blum et al. '98; Servedio '01

- Agnostic model, but $\exp(\frac{1}{\epsilon})$

  Kalai et al.,'08; Ben-David&Simon '00;

  Shalev-Swartz et al. '11

- Adversarial noise, but convergence to $C \cdot \mathrm{opt}$

  Awasthi et al. '14; Daniely '15

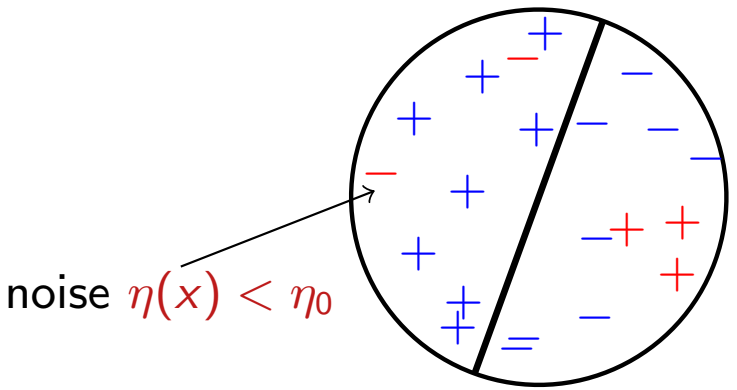
Statistical Learning Theory

Statistical rates:

- Fast rate under bounded noise

  Mammen&Tsybakov '99; Tsybakov '04;

  Massart&Nédélec '06



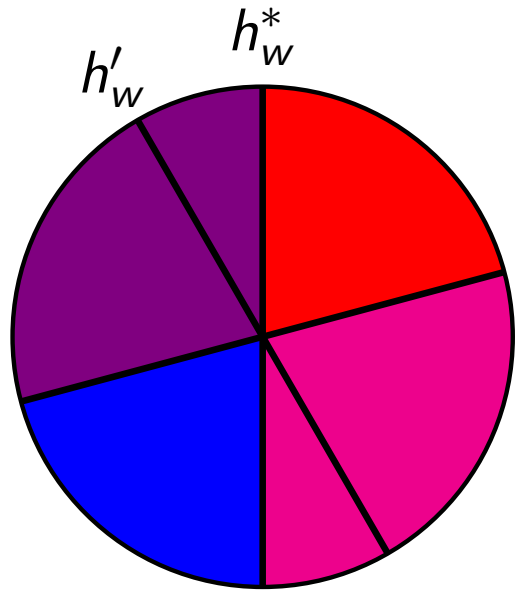
Question: Computational efficient algorithm in this model?

...it's not that easy!

We show that the
Averaging Algorithm and Hinge Loss Minimization
are not consistent in our model.

Hinge loss minimization does not work

For every
bound on Massart noise η,
there is a
benign distribution
on which
hinge loss minimization fails.

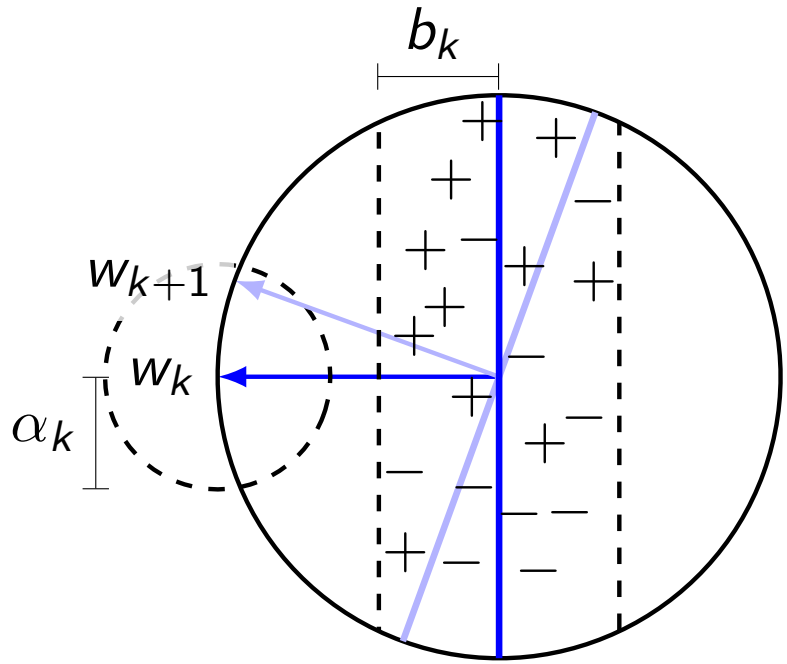


Our algorithm

Margin-based hinge-loss minimization

For rounds $k = 1, \ldots, \log(\frac{1}{\epsilon})$

- Sample m_k points in b_k-band around w_k
- Set w_{k+1} to be the τ_k-hinge minimizer from $B(\alpha_k, w_k)$ in the b_k-band



$\rightarrow$ achieves arbitrarily small excess error in poly-time!

See you at the poster!