15-859(B) Machine Learning Theory

Lecture 7: Boosting I

Avrim Blum
02/06/08

Today: some basic definitional questions in the PAC model

Algorithm PAC-learns a class of functions C if:

- For any given $\varepsilon>0$, $\delta>0$, any target $f \in C$, any dist. D, with prob at least $1-\delta$ the algorithm produces h of $err(h)<\varepsilon$.
- Running time and sample sizes polynomial in relevant parameters: $1/\varepsilon$, $1/\delta$, n, size(f).
- Require h to be poly-time evaluatable (don't require $h \in C$).

Q1: do we need "for all δ"? What if we replace that with "exists $\delta' > 0$ such that alg **succeeds** with prob $\geq \delta'$ "?

Claim: if C is learnable using new def then also learnable with old def

Say A achieves error $\leq \varepsilon/2$ with prob $\geq \delta'$. Uses $|S|=m$.

- Run it $1/\delta'$ times. (m/δ' data points). With prob at least 1-1/e it succeeds at least once.
- Run it $\ln(2/\delta)$ factor more times. With prob at least $1 - \delta/2$ it succeeds at least once.
- Now test hypotheses on new test set S' of size $O((1/\varepsilon)\log(1/(\delta\delta')))$ and pick best. By Chernoff bounds, whp this has error $\leq \varepsilon$. (see hwk)

Q2: do we need to say "for all ε"?

- Def: Say alg A weak-learns class C if there exists $\varepsilon,\delta > 0$ [1/poly(n)] such that for all $f \in C$, all D, A achieves error at most $\frac{1}{2}-\varepsilon$ with probability at least δ.
  - I.e., with some noticeable probability it does noticeably better than guessing.
- If we defined PAC-learning this way, does that change the set of learnable C?
- No. Given alg satisfying this, can "boost" to satisfy original def.

OK, now let's go to the blackboard…