Mehryar Mohri
Foundations of Machine Learning
Courant Institute of Mathematical Sciences
Homework assignment 4
Due: April 27, 2007

This is a shorter assignment to leave you more time to work on your project.

Boosting

Suppose we simplify AdaBoost by setting the parameter α_t to a fix value $\alpha_t = \alpha > 0$, independent of the boosting round t.

1. Let γ be such that $(\frac{1}{2} - \epsilon_t) \geq \gamma > 0$ where ϵ_t is defined as in class. Find the best value of α as a function of γ by analyzing the empirical error.

2. For that value of α, does the algorithm assign the same probability mass to correctly classified and misclassified examples at each round? Which set is assigned a higher probability mass?

3. Using the previous value of α, give a bound on the empirical error of the algorithm that depends only on γ and the number of rounds of boosting T.

4. Using the previous bound, show that for $T > \frac{\log m}{2\gamma^2}$, the resulting hypothesis is consistent with the sample of size m.

5. Let s be the VC dimension of the base learners used. Give a bound on the generalization error of the consistent hypothesis obtained after $T = \left\lfloor \frac{\log m}{2\gamma^2} \right\rfloor + 1$ rounds of boosting (*hint*: you can use the fact that the VC dimension of the family of functions $\left\{ \text{sgn}(\sum_{t=1}^{T} \alpha_t h_t) : \alpha_t \in \mathbb{R} \right\}$ is bounded by $2(s+1)T \log_2(eT)$). Suppose now that γ varies with m. Based on the bound derived, what can you say if $\gamma(m) = O(\sqrt{\frac{\log m}{m}})$?