Boosting

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

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

2. For that value of $\alpha$, 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 $\alpha$, give a bound on the empirical error of the algorithm that depends only on $\gamma$ 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 $\left\{ \text{sgn}\left( \sum_{t=1}^{T} \alpha_t h_t \right) : \alpha_t \in \mathbb{R} \right\}$ is bounded by $2(s + 1)T \log_2(eT)$). Suppose now that $\gamma$ varies with $m$. Based on the bound derived, what can you say if $\gamma(m) = O\left(\sqrt{\frac{\log m}{m}}\right)$?