Boosting

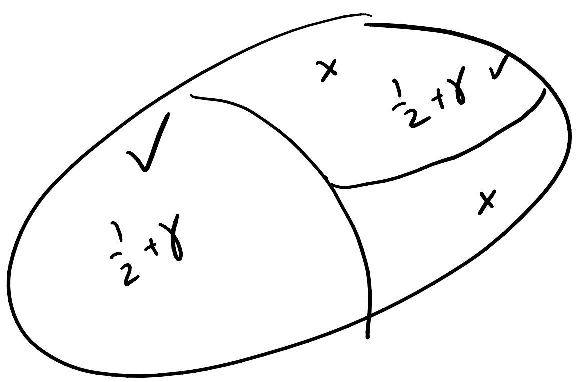
Monday, October 25, 2021 6:59 AM

yeh

Suppose we know how to learn $h \in \mathcal{H}$ s.t. it has accuracy $\frac{1}{2}+\gamma$ i.e.

$Pr_D(h(x) = \ell(x)) \geq \frac{1}{2}+\gamma$ for any D.

Can we improve this <u>generically</u>?



How to combine?

<u>Boosting</u>:

Each example has a weight.

Initially all $w_i = 1$.

When example is misclassified weight goes up by a factor of $\frac{\frac{1}{2}+\gamma}{1-\gamma}$.

weight goes up by a factor of $\frac{\quad}{\frac{1}{2}-\gamma}$.

Run for T iterations using weak learner each time. At the end, label x by the <u>majority</u> vote of the T classifiers.

Suppose # mistakes is $\hat{m}$.
Then each of them has weight $\geq \left(\frac{\frac{1}{2}+\gamma}{\frac{1}{2}-\gamma}\right)^{T/2}$.

In each round, max increase in total weight is if $\frac{1}{2}-\gamma$ examples increase their weight.

$$W(t+1) \leq \left(\left(\frac{\frac{1}{2}+\gamma}{\frac{1}{2}-\gamma}\right)\cdot\left(\frac{1}{2}-\gamma\right) + \frac{1}{2}+\gamma\right)\cdot W(t)$$

$$= (1+2\gamma)\,W(t)$$

$W(0) = n$.

$$\leq (1+2\gamma)^{T}\, n$$

So $\ldots)^{T/2} \ldots)^{T}$

So

$$\hat{m}\left(\frac{(1+2\gamma)}{(1-2\gamma)}\right)^{T/2} \leq (1+2\gamma)^T n$$

$$\hat{m} \leq (1-4\gamma^2)^{T/2} \cdot n$$

Set $T = \frac{\ln n}{2\gamma^2}$ then $\hat{m} < 1$, i.e. $\hat{m} = 0$.

But how large n do we need to ensure the hypothesis generalizes?

What is the VC-dimension of majority of T classifiers?

<u>Lemma.</u> VC-dim(k-maj of $h \in \mathcal{H}$)

$$\leq 2 k d \log(kd)$$

where $d = \text{VC-dim}(\mathcal{H})$.

So $n = O\left(\frac{1}{\varepsilon}\left(kd \log(kd) \log\frac{1}{\varepsilon} + \log\frac{1}{\delta}\right)\right)$

examples suffice, with $T = \frac{\ln n}{\gamma^2}$.