

VC dimension

Sunday, October 17, 2021 8:56 PM

jahn

We have seen PAC and Mistake bound algorithms for many concept classes.

In the case of halfspaces there was a $\frac{1}{\gamma^2}$ dependence on the margin γ .

In fact, one can make this $\log \frac{1}{\gamma}$.

Suppose we predict majority of all surviving w .

i.e. suppose after examples $x^1, x^2, \dots x^l$,

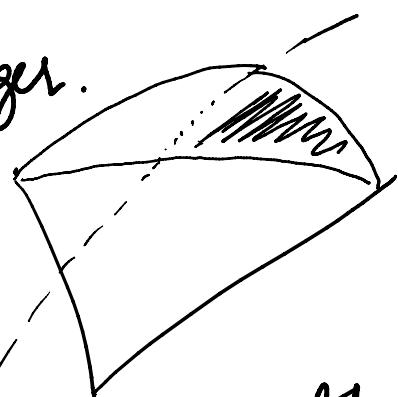
we have $W = \{w : w^T x^i \geq 0, \|w\| \leq 1\}$

as candidates and we consider which of

$W \cap \{w : w^T x^{l+1} \geq 0\}$, $W \cap \{w : w^T x^{l+1} < 0\}$

is larger.

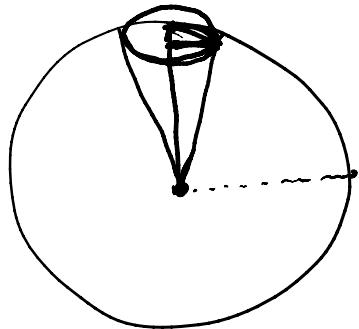
Predict according to that.



Then in each step we eliminate $\frac{1}{2}$ the volume.
 $\text{vol}(W)$ starts at $\text{vol}(B)$.
... (with base?)

$\text{vol}(W)$ starts at $\text{vol}(D)$.

At the end it is at least $\text{Vol}(\gamma\text{-cone})$



$$\text{Vol}(B) = \int_0^1 (\sqrt{1-t^2})^{n-1} \text{vol}(B^n) dt$$

$$\text{Vol}(\gamma\text{-cap}) = \int_0^1 (\sqrt{1-t^2})^{n-1} \text{vol}(B^n) dt$$

$$\frac{\text{Vol}(\gamma\text{-cap})}{\text{Vol}(B)} = \frac{\int_0^1 (1-t^2)^{\frac{n-1}{2}} dt}{\int_0^1 (1-t^2)^{\frac{n-1}{2}} dt} \geq c \cdot \gamma^n.$$

$$\therefore \# \text{mistakes} = O(n \log \frac{1}{\delta}).$$

How to estimate volume fraction?

We can sample the current W and take the majority vote of the sample. Even 1 sample suffices!

Alternatively we can use Linear Programming to find a feasible w for all constraints so far.

To bound the number of examples needed we can use a more general theory.

a more general theory.

VC-dimension. m points.

Concept class H .

How many distinct subsets of m points are defined by $h \in H$? $H[m] \leq m^{\text{VC-dim.}}$

More precisely: $\text{VC-dim} = \max_{(H)} m \text{ s.t. } \exists m \text{ points}$
that can be shattered, i.e.
split in all possible ways by H .

e.g. intervals on a line $\text{VC-dim} = 2$

rectangles in 2-d $\text{VC-dim} = 4$

Halfspaces in \mathbb{R}^d $\text{VC-dim} = d+1$

Thm 1. For a concept class H of VC-dim d , # distinct ways to split m points using $h \in H$ is $\leq m^d$.

Thm 2. # examples needed to (ϵ, δ) -PAC learn H is $\leq 2(\log(H[2m]) + \log \frac{1}{\delta})$.

$$\begin{aligned} n &\leq \frac{2}{\varepsilon} (\log_2 H[2m]) + \log \frac{1}{8} \\ &= O\left(\frac{1}{\varepsilon} (d \log \frac{1}{\varepsilon} + \log \frac{1}{8})\right). \end{aligned}$$

PF(Thm 1). We will show $H[m] \leq \sum_{i=0}^d \binom{m}{i} = \binom{m}{\leq d}$.

Let S be a set of m points.

Induction on m . True for $m \leq d$.

Let $x \in S$. Consider $S \setminus \{x\}$.

By induction $H(S \setminus \{x\}) \leq \binom{m-1}{\leq d}$.

Also note that

$$\binom{m}{\leq d} = \binom{m-1}{\leq d} + \binom{m-1}{\leq d-1}$$

So it suffices to show that

$$H(S) - H(S \setminus \{x\}) \leq \binom{m-1}{\leq d-1}$$

How can $H(S)$ be larger? There must be labelings h and h' s.t. they agree on all

labelings h and h' S.t. may not
both except x .

$$\text{Let } T = \{h \in H(S) : h(x) = 1, h' \in H(S)\}.$$

Then we are interested in bounding $|T|$.

Let $\text{VC-dim}(T) = d'$. So $2^{d'}$ points can
be shattered by T . But then $d'+1$ points can
be shattered by H . So $d'+1 \leq d$.

$$\text{i.e. } d' \leq d-1.$$

$$\text{Hence } H(T) \leq \binom{m-1}{\leq d-1}.$$

Pf (Th 2). We find a hypothesis h_s that
correctly classifies m points. We want to show
that with prob $\geq 1-\delta$

$$\Pr_D(h_s(x) \neq h(x)) \leq \epsilon. \quad \begin{cases} \text{Let } A \text{ be the} \\ \text{complement of} \\ \text{this event} \end{cases}$$

Consider a different setting where we pick
2 subsets of size m , say
 \dots, a, b, \dots

↙ answer of "you", "

S and S' . let B be the event that a hypothesis h has error on S and $\text{err} > \frac{\varepsilon}{2}$ on S' .

Claim. $\Pr(B) \geq \frac{1}{2} \Pr(A).$

$$\Pr(B) = \Pr(A) \Pr(B/A)$$

$\Pr(B/A)$: $\Pr(h \text{ has error } \geq \frac{\varepsilon}{2} \text{ on } m \text{ points}$
given that it has error $\geq \varepsilon$ on D)

This is a simple Chernoff bound.

$$\Pr\left(\sum X_i - \mathbb{E}(\sum X_i) < \delta \mathbb{E}(\sum X_i)\right) \leq e^{-\frac{\delta^2 \mathbb{E}(X)}{2}}.$$

$$\Pr\left(\sum X_i < \frac{\varepsilon m}{2}\right) \leq e^{-\frac{\varepsilon m}{8}}.$$

$$m \geq \frac{8}{\varepsilon} \Rightarrow \Pr(B/A) \geq \frac{1}{2}$$

So we want to show $\Pr(B) \leq \frac{8}{2}$.

For this we pick $2m$ points, "partition them randomly into two subsets S, S' of m points".

randomly into two subsets S, S' of m points.
 Then we want to bound $P_S(\text{err}_h(S) = 0, \text{err}_h(S') > \frac{\epsilon}{2})$.

Pair up the $2m$ points $(a_1, b_1), \dots, (a_m, b_m)$.

Fix hypothesis h .

If h makes error on both a_i and b_i ,
 then $P_S = 0$. (since no errors allowed on S).
 Also at least $\frac{\epsilon m}{2}$ indices i must make an error.

$$\text{So } P_S(\text{all } \frac{\epsilon m}{2} \text{ errors fall in } S') \leq \frac{1}{2^{\frac{\epsilon m}{2}}}$$

$$\#\text{possible } h \leq H(2m)$$

$$\therefore \text{suffices to have } 2^{-\frac{\epsilon m}{2}} H[2m] \leq \frac{\delta}{2}$$

$$\text{i.e. } m \geq \frac{2}{\epsilon} \left(\log 2H[2m] + \log \frac{1}{\delta} \right).$$
