PAC and the VC dimension

PAC learnability is a property of the concept class C. It is determined by a combinatorial measure of C that is called *the VC dimension*.

Definition: Given a nonempty concept class C, the Vapnik-Chervonenkis (VC) dimension of C is the cardinality of the largest finite set of examples X, such that for any subset $X' \subset X$ there is a concept $c \in C$ that classifies all examples in X' as positive, and all examples in X but not in X' as negative.

A relationship between consistent algorithms, PAC learnability, and VC dimension is given by the following theorems.

Theorem: Let d be the VC dimension of a concept class C. Any algorithm consistent with m random examples achieves ϵ error with confidence parameter δ (in the PAC learning sense) where:

$$m \geq \frac{1}{\epsilon}(4\log_2(2/\delta) + 8d\log_2(13/\epsilon))$$

If an algorithm is consistent with a significantly smaller number of traning examples then it may not have sufficient accuracy:

Theorem: If $d \geq 2$, $0 < \epsilon < \frac{1}{8}$, and $0 < \delta < \frac{1}{100}$ then there is a probability distribution D and a concept class C such that a learning algorithm applied to learn a concept from C using m examples may have an error of more than ϵ where:

$$m < \frac{1}{\epsilon}\max\left(\ln\frac{1}{\delta}, \frac{d-1}{32}\right)$$

Computing VC dimension is usually nontrivial. Special cases that are relevant to us are that the VC dimension of a finite concept class with v concepts is at most $\log_2 v$, and that the VC dimension of the concept class of all half spaces in n dimensions is $n+1$.