Computational Learning Theory

What general laws constrain inductive learning?

Seek theory to relate:

- Probability of successful learning
- Number of training examples
- Complexity of hypothesis space
- Accuracy to which target concept is approximated
- Manner in which training examples presented

Some Specific Questions/Frameworks

- *Learning in the limit:* Guaranteed to converge to correct hypothesis in the limit as training data increases?

- *Sample complexity:* How many training examples required to succeed$\hat{*}$?

- *Computational complexity:* How much computational effort required to succeed$\hat{*}$?

- *Mistake bound:* How many mistakes must the learner make before succeeding$\hat{*}$?

* What do we mean by *succeed*?

The PAC Learning Setting

Given:

- set of possible instances X
- hypotheses H
- target concepts C
- training instances generated by a fixed, unknown probability distribution $\mathcal{D}$ over X

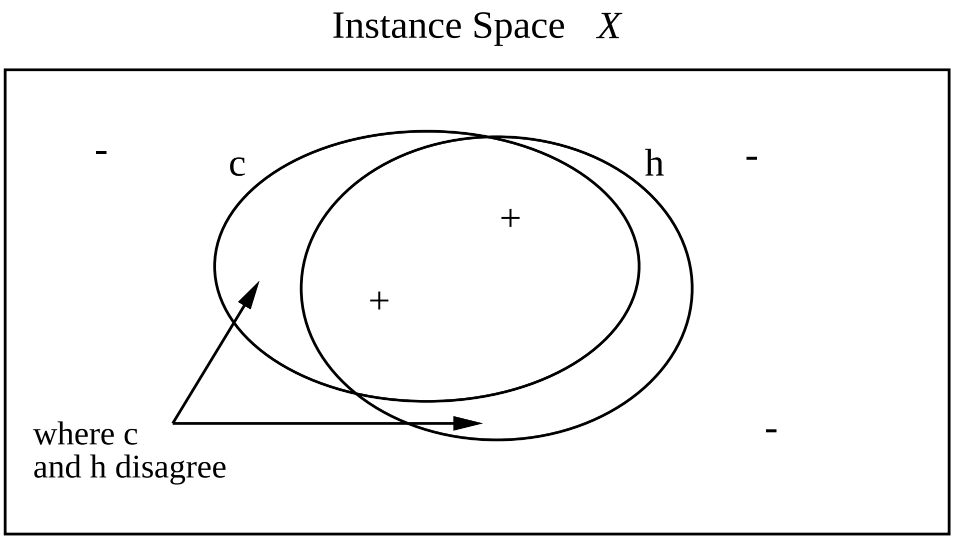
Learner observes a seqence D of training examples of form $\langle x, c(x) \rangle$, for some target concept $c \in C$

- instances x are drawn from distribution $\mathcal{D}$
- teacher provides target value $c(x)$ for each

Learner must output a hypothesis h estimating c

- h is evaluated by its performance on subsequent instances drawn according to $\mathcal{D}$

Note: probabilistic instances, noise-free classifications



Definition: Given some target concept c, hypothesis h, and instance distribution $\mathcal{D}$, the **error of h with respect to c and** $\mathcal{D}$ is the probability that h and c disagree on the classification of an instance drawn at random according to $\mathcal{D}$.

$$error(h) = Pr_{x \in \mathcal{D}}[c(x) \neq h(x)]$$

Training and true error

Training error of h:

- $r(h) =$ percent of training examples for which $h(x) \neq c(x)$

(True) error of h:

- $error(h) = Pr_{x \in \mathcal{D}}[c(x) \neq h(x)]$

Our concern:

- How probable is it that $r(h)$ gives misleading estimate of $error(h)$
- How does this depend on the number of training examples, H, etc.

PAC Learnability

Don't require perfect learning, but only

- output hypothesis has $error(h) \leq \epsilon$
- learner fails to find such a hypothesis with probability $\leq \delta$

I.e., require that learner **probably** $(1 - \delta)$ output a hypothesis that is **approximately correct** (ϵ)

PAC Learnability

What classes C of target concepts can be PAC learned using a "reasonable" ammount of training data and computational effort?

- Reasonable = polynomial in $1/\epsilon$, $1/\delta$ and problem size

Definition: Consider a concept class C defined over a set of instances X of length n, and a learner L using hypothesis space H. C is **PAC-learnable** by L using H if for all $c \in C$, distributions $\mathcal{D}$ over X, ϵ such that $0 < \epsilon < 1/2$, and δ such that $0 < \delta < 1/2$, learner L will with probability at least $(1 - \delta)$ output a hypothesis $h \in H$ such that $error(h) \leq \epsilon$, in time that is polynomial in $1/\epsilon$, $1/\delta$, n and $size(c)$.

Consistent Learners

Definition: Let L be a learner that considers hypothesis space H and is given a sample of training examples D. L is said to be a **consistent learner** if it outputs a hypothesis h that correctly classifies every training example in D, whenever such a hypothesis exists within H.

Question: how many training examples are sufficient for *any* consistent learner to ϵ, δ learn the target concept??

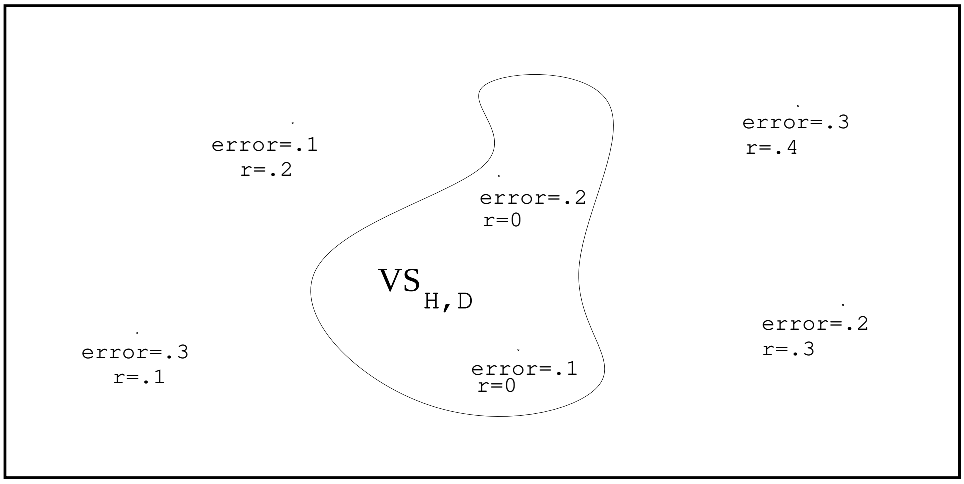
Note, *every* consistent learner must output some h from $VS_{H,D}$

$$VS_{H,D} = \{h \in H | (\forall \langle x, c(x) \rangle \in D)(h(x) = c(x))\}$$

Rephrased question: how many training examples are sufficient to probably $(1 - \delta)$ assure that no hypotheses with $error(h) \geq \epsilon$ remain in the $VS_{H,D}$?

Exhausting the Version Space

Hypothesis Space H



($r =$ training error, $error(h) = Pr_{x \in \mathcal{D}}[c(x) \neq h(x)]$)

Consider a hypothesis space H, target concept c, instance distribution $\mathcal{D}$, and sequence of examples D of c.

> **Definition:** The version space $VS_{H,D}$ is said to be ϵ-**exhausted** with respect to c and $\mathcal{D}$, if every hypothesis h in $VS_{H,D}$ has error less than ϵ with respect to c and $\mathcal{D}$.

PAC Learning by Consistent Learners

ϵ-exhausting the version space [Haussler, 1988]. If the hypothesis space H is finite, and D is a sequence of $m \geq 1$ independent random examples of some target concept c, then for any $0 \leq \epsilon \leq 1$, the probability that the version space with respect to H and D is not ϵ-exhausted (with respect to c) is less than

$$|H|e^{-\epsilon m}$$

Interesting! This bounds the probability that any consistent learner will output a hypothesis h with $error(h) \geq \epsilon$

Suppose we want to reduce this probability below some level δ

$$|H|e^{-\epsilon m} \leq \delta$$

then

$$m \geq \frac{1}{\epsilon}(\ln |H| + \ln(1/\delta))$$