Xishun Zhao *

November 14, 2012

Abstract

An universe U, a hypothesis space $\mathcal{S} = \{L_1, \cdots, \}$ of susbets of U. S can be infinite. An unknown susbet L_*. After some queries, learn L_*.

Quesries include:

Memebership query:

Equivalence query

Subset query

Supserset query

Disjointness query

Exhaustiveness query

When an input of a query is a subset of U, then it must be in $\mathcal{S}$.

poker hand

a card: a number and a suit(S,H,D,C))

A hand is a set of five cards, without any order

A pair of hands is an ordered pair of hands with no card in common

The universe U is the set of all pairs of hands.

Exact identification: after some queries to find an index i such that the target notion is exactly L_i in $\mathcal{S}$.

Probablistic identification (by L.G.Valiant 1988).

a disributuion D on U. $\Pr(x)$ is the probability of element x wrt D

Sampling oracle EX( ) which has no input. When EX( ) is called it returns an element x with an identification of whether x is in the target set

*Corresponding author. Tel: 0086-20-84114036, Fax:0086-20-84110298.

difference of two sets L_1, L_2: $d(L_1, L_2) := \Sigma_{x \in L_1 \otimes L_2} Pr(x)$

two parameters ϵ accuracy, δ confidence.

probably approximately correctly(pac) identification: always halts and output an index i such that $Pr(d(L_*, L_i) \leq \epsilon) \geq 1 - \delta$

pac identification is used if EX( ) is available. Otherwise we use exact identification

Equivalence query

exhaustive search: enuerate indeices $i = 1, \cdots$, quering each L_i until one gets an answer of yes, and halts at this point.

If hypothesis space is too large, then exhaustive search may use too many times of equivalence queries.

U binary strings of length n. $S = U$. L_* is a unknown string want to learn. Only equivalece, membership, subset, disjoint queries are available, then in the worst case, may need $2^n - 1$ queries.

Majority strategy (me want to learn some thing from an adversary). an input of the equivalent query can be not a hypothesis. Then $\log N$ queries is sufficient.

Lower bound, and techniques.

input must be a hypothsis.

hypothesis $L_1, \cdots, L_N$ such that $L_i \cap L_j = L_\cap$ for $i \neq j$.

To learn $L_\cap$

Stochastic equivalence

Inequality

$$\Sigma_{i=1}^n (1 - \epsilon)^{q_i} \leq \Sigma_{i=1}^n e^{-\epsilon q_i}$$