Sample compression schemes for VC classes

Shay Moran Amir Yehudayoff

Technion, Haifa, Israel

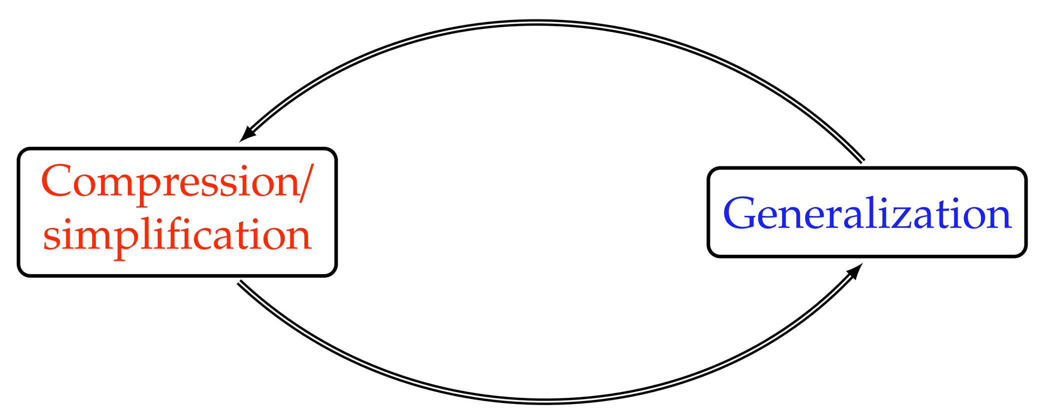
Learning

Two aspects:

1. Generalization:
Infer new knowledge from existing knowledge

2. Simplification/compression:
Provide simple(r) explanations for existing knowledge

Interrelations



Example (math):

theorem $\xrightarrow{\text{simplification}}$ simpler proof $\xrightarrow{\text{generalization}}$ more general theorem

Simplification $\Longrightarrow$ generalization

Aim for the simplest valid explanation.
[Occam's razor – William of Ockham ≈ 1300].

Generalization $\Longrightarrow$ simplification

"If we can't reduce it to the freshman level then we don't understand it." [Richard Feynman 1980's].

When presented with a complicated proof, Erdos replied:
"Now, let's find the book's proof..." [Paul Erdos]

Can these interrelations be manifested as theorems in learning theory?

Setup

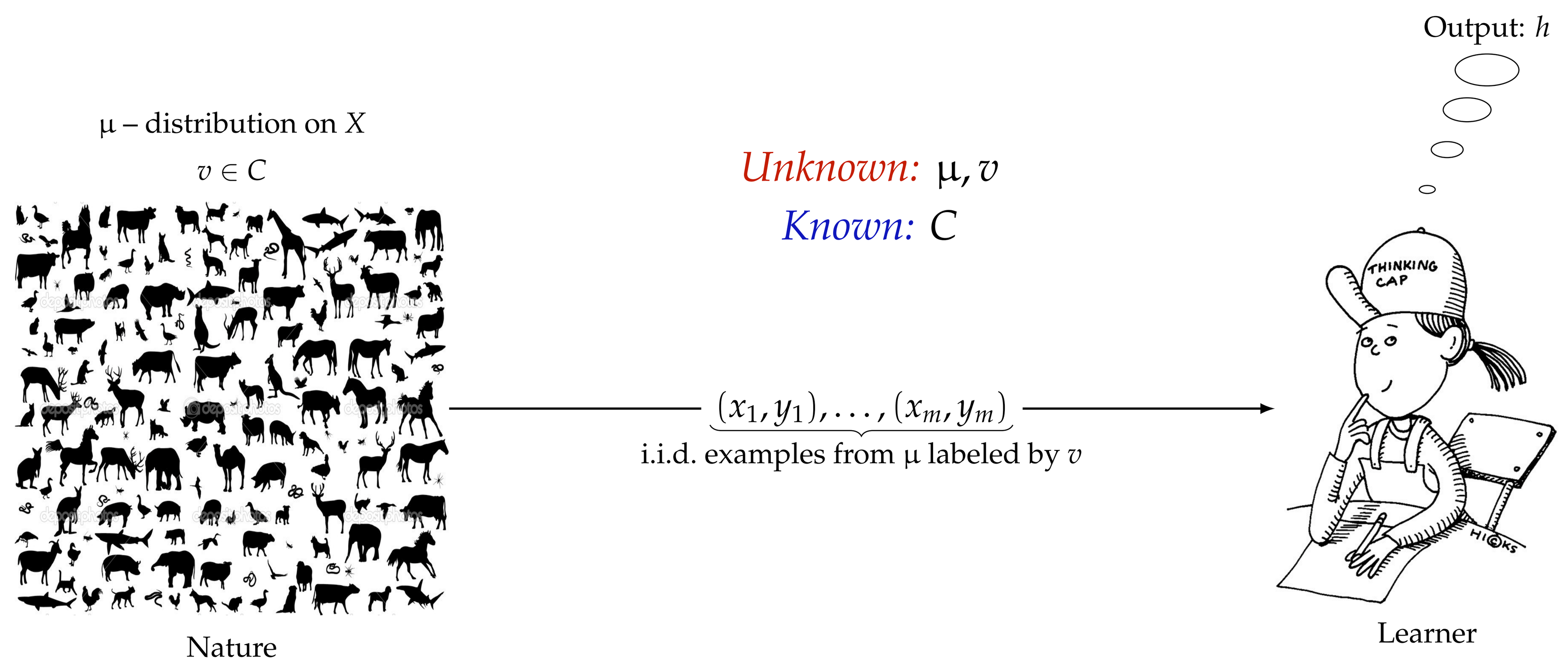
Concept: unknown concept $v : X \to \{0,1\}$

Given: a sample $(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)$ labeled by v

Goal: find an hypothesis $h : X \to \{0,1\}$ that is "close" to the unknown v

Assumption: v belongs to a known concept class $C \subseteq \{0,1\}^X$

PAC [Vapnik, Chervonenkis '71, Valiant '84]



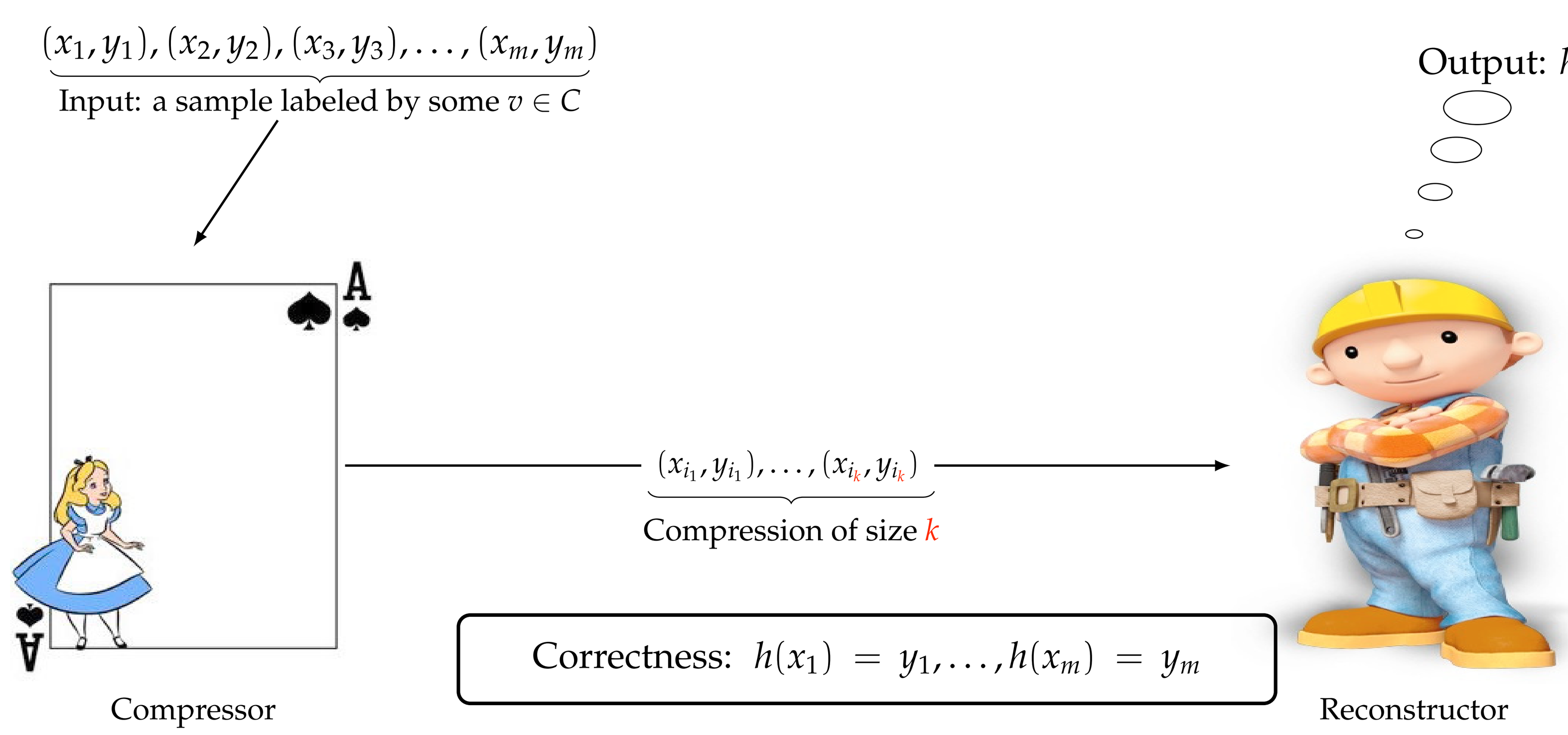
Correctness: with high μ^m-probability (confidence) h is μ-close (error) to v

$dim(C)$ = minimum number of examples for learning with confidence 2/3, error 1/3

Theorem
[Vapnik,Chervonenkis], [Blumer,Ehrenfeucht,Haussler,Warmuth], [Ehrenfeucht,Haussler,Kearns,Valiant]:

$$dim(C) \approx \text{VC dimension of } C$$

Sample Compression Schemes [Littlestone, Warmuth '86]



Correctness: $h(x_1) = y_1, \ldots, h(x_m) = y_m$

Generalization versus Compression

Theorem (simplification $\Longrightarrow$ generalization) ***[Littlestone, Warmuth '86]*:**

If C has a compression scheme of size k then $dim(C) = O(k)$

A manifestation of Occam's razor

Question (generalization $\Longrightarrow$ simplification?):

Is there a compression scheme of size depending only on $dim(C)$?

A manifestation of Feynman and Erdos's statements

Previous works

Boosting: $dim(C) \log m$ compression scheme
[Freund, Schapire 90's]

Compression schemes for special well-studied concept classes
[Floyd,Warmuth '95],[Floyd '89],[Helmbold,Sloan,Warmuth '92], [Ben-David,Litman '98],[Chernikov,Simon '13],[Kuzmin,Warmuth '07], [Rubinstein,Bartlett,Rubinstein '09],[Rubinstein,Rubinstein '12], [Livni,Simon '13], [M,Warmuth '15] ...

Connections with model theory
[Chernikov,Simon '13],[Livni,Simon '13],[Johnson '09],...

Connections with algebraic topology
[Rubinstein,Bartlett,Rubinstein '09],[Rubinstein,Rubinstein '12]

Enough to compress finite classes (compactness)
[Ben-David,Litman '98]

Compression scheme of size $\log|C|$
[Floyd,Warmuth '95]

compression scheme of size $\exp(dim(C)) \log\log|C|$
[M,Shpilka,Wigderson,Yehudayoff '15]

Result

Theorem (generalization $\Longrightarrow$ simplification) *[M-Yehudayoff]:*

There exists a compression scheme of size $\exp(dim(C))$

Proof uses:
(i) von Neumann's minimax theorem
(ii) duality
(iii) ϵ-approximations

Future research

1. Replace $\exp(dim(C))$ by $O(dim(C))$ (Manfred Warmuth offers 600$!)

2. Extend to other learning models

http://www.cs.technion.ac.il/~shaymrn/ shaymoran1@gmail.com