Inductive Inference

Sanjay Jain and Frank Stephan

Department of Computer Science (Jain and Stephan)

Department of Mathematics (Stephan)

National University of Singapore

sanjay@comp.nus.edu.sg

fstephan@comp.nus.edu.sg

Content

I. Inductive Inference

Learner reading data and outputting hypotheses.

Data	Hypotheses
2	Set of even numbers;
2,3	Set of all numbers;
2,3,5	Set of prime numbers;
2,3,5,13	Set of prime numbers;
2,3,5,13,1	Set of Fibonacci numbers;
2,3,5,13,1,8	Set of Fibonacci numbers.
2,3,5,13,1,8,21	Set of Fibonacci numbers.

Learner outputs a sequence of conjectures which eventually stabilizes on the correct one.

General Setting

Class $\mathbf{C}$ of sets to be learnt; all sets are r.e. subsets of a base set like $\mathbb{N}$ and $\{\mathbf{0},\mathbf{1}\}^*$.

Learner reads more and more data from an infinite sequence (called text) containing all members of some set $\mathbf{L} \in \mathbf{C}$ and perhaps some pause symbols.

Learner conjectures hypothesis $\mathbf{e_n}$ for data $\mathbf{a_0}\,\mathbf{a_1}\,\ldots\,\mathbf{a_n}$.

In general: from some time onwards, all $\mathbf{e_n}$ are the same correct hypothesis $\mathbf{e}$ describing the set $\mathbf{L}$ to be learnt. Gold called this model “explanatory learning”.

Description is index in a chosen hypothesis space $\{\mathbf{L_e} : \mathbf{e} \in \mathbf{I}\}$ where $\mathbf{I}$ is a suitable index set and every member of $\mathbf{C}$ equals to some $\mathbf{L_e}$. Often, the hypothesis space is some fixed acceptable numbering $\mathbf{W_0}, \mathbf{W_1}, \ldots$ of all r.e. sets.

Choices for this talk

I+III: Learners are recursive; II: learners are automatic.

Instead of classes of r.e. sets, one can also look at spaces of recursive functions, co-r.e. sets, regular sets and so on. I+III: Classes of r.e. sets; II: Classes of regular sets.

Besides texts, there are also input presentations like a text plus an upper bound of the index; an informant providing both, positive and negative data; a text plus only some selected negative data; a teacher which answers explicit questions of the learner. I-III: mostly learning from text.

Hypotheses spaces are indexing of all the objects which are contained in the class to be learnt plus perhaps some other ones. Much work has been done on relations between learnability and hypothesis space. I+III: Mostly an acceptable numbering of all r.e. sets; II: automatic families.

I.1. Examples

The following examples use the hypothesis space $\{W_0, W_1, \ldots\}$ of all r.e. sets.

Class of Finite Sets

On input $a_0\, a_1\, \ldots\, a_n$, the learner produces an index for the set $\{a_0, a_1, \ldots, a_n\}$ (repetitions removed) of the data seen so far. The hypothesis is only revised when a new datum has been observed and therefore the learner converges to some correct hypothesis.

Class of Co-Single Sets

On input $a_0\, a_1\, \ldots\, a_n$, the learner computes the value

$$b_n = \min(\mathbb{N} - \{a_0, a_1, \ldots, a_n\})$$

and outputs an index $f(b_n)$ with $W_{f(b_n)} = \mathbb{N} - \{b_n\}$.

Learning by Enumeration

Assume that the hypothesis space $\{L_e : e \in I\}$ is uniformly recursive and that $L_d \not\subset L_e$ for all d, e. Then one can learn the class C by the following algorithm called "Learning by Enumeration".

On input $a_0\, a_1\, \ldots\, a_n$, the hypothesis e_n is the first e in I such that either $\{a_0, a_1, \ldots, a_n\} \subseteq L_e$ or $e > n$.

In the case that the set L_e has to be learnt, the algorithm will converge to the first d with $L_d = L_e$; the indices below d will all be abandoned eventually when some counter example is observed in the input.

Learning by Self-Reference

Consider the class $\mathbf{C} = \{\mathbf{W_e} : \mathbf{e} = \mathbf{min}(\mathbf{W_e})\}$.

This class has an easy learner with respect to the hypothesis space $\{\mathbf{W_0}, \mathbf{W_1}, \ldots\}$: On input $\mathbf{a_0}\, \mathbf{a_1}\, \ldots\, \mathbf{a_n}$, the learner conjectures $\mathbf{min}\{\mathbf{a_0}, \mathbf{a_1}, \ldots, \mathbf{a_n}\}$.

This learner converges to the miminum of the data observed which is the minimum of the set to be learnt and by choice of $\mathbf{C}$ then also the index of the set to be learnt.

Wiehagen observed that the class $\mathbf{C}$ is quite large: It contains for every r.e. set $\mathbf{E}$ a finite variant of $\mathbf{E}$. For this, let $\mathbf{f}$ be chosen such that

$$\mathbf{W_{f(e)}} = \{\mathbf{e}\} \cup \mathbf{E} - \{\mathbf{d} : \mathbf{d} < \mathbf{e}\}.$$

By Kleene's Fixed-Point Theorem there is an $\mathbf{e}$ with $\mathbf{W_e} = \mathbf{W_{f(e)}}$ and then $\mathbf{W_e} \in \mathbf{C}$.

Gold's Unlearnable Class

Theorem [Gold 1967]
Let $\mathbf{C}$ contain all finite sets and $\mathbb{N}$. Then $\mathbf{C}$ is not learnable.

Proof
To see this, assume that $\mathbf{M}$ learns every finite set in $\mathbf{C}$. Now construct inductively a text $\sigma_\mathbf{0}\sigma_\mathbf{1}\ldots$ such that each $\sigma_\mathbf{n}$ consists of so many repetitions of $\mathbf{n}$ that $\mathbf{M}$ on input $\sigma_\mathbf{0}\sigma_\mathbf{1}\ldots\sigma_\mathbf{n}$ outputs a conjecture for $\{\mathbf{0},\mathbf{1},\ldots,\mathbf{n}\}$.

Each $\sigma_\mathbf{n}$ has to exist as $\mathbf{M}$ on the text $\sigma_\mathbf{0}\sigma_\mathbf{1}\ldots\sigma_\mathbf{n-1}\,\mathbf{n}\,\mathbf{n}\,\mathbf{n}\,\mathbf{n}\,\ldots$ learns the set $\{\mathbf{0},\mathbf{1},\ldots,\mathbf{n}\}$ and therefore also outputs an index for that set. As a consequence, $\mathbf{M}$ outputs on the text $\sigma_\mathbf{0}\sigma_\mathbf{1}\ldots$ for $\mathbb{N}$ infinitely many wrong indices and so does not learn $\mathbb{N}$.

A Computationally Difficult Class

Theorem

Let $\mathbf{C}$ contain all r.e. sets which, for every $\mathbf{n}$, contain exactly one of $\mathbf{2n}, \mathbf{2n+1}$. Then $\mathbf{C}$ is not explanatorily learnable.

Proof

Assume now by way of contradiction that $\mathbf{M}$ is an explanatory learner for $\mathbf{C}$. Now one constructs a text $\mathbf{T}$ such that $\mathbf{M}$ makes infinitely many mind changes on $\mathbf{T}$ and $\mathbf{T}$ is a text for a set in $\mathbf{C}$. Let $\sigma_\mathbf{0}$ be $\mathbf{0\,2}$. For $\mathbf{n} = \mathbf{0}, \mathbf{1}, \ldots$ let $\sigma_{\mathbf{n+1}}$ such that

- $\sigma_{\mathbf{n+1}}$ contains either $\mathbf{2m}$ or $\mathbf{2m+1}$ and, for some $\mathbf{k}$, $\mathbf{2m{+}2\;2m{+}4\;2m{+}6\;\ldots\;2k}$ for the least $\mathbf{m}$ such that $\mathbf{2m}, \mathbf{2m+1}$ do both not occur in $\sigma_\mathbf{0}\,\sigma_\mathbf{1}\ldots\sigma_\mathbf{n}$;
- $\mathbf{M}(\sigma_\mathbf{0}\,\sigma_\mathbf{1}\ldots\sigma_\mathbf{n}\,\sigma_{\mathbf{n+1}}) \neq \mathbf{M}(\sigma_\mathbf{0}\,\sigma_\mathbf{1}\ldots\sigma_\mathbf{n})$.

Each extension can be found as otherwise $\mathbf{M}$ would converge on two different sets from $\mathbf{C}$ to the same index.

I.2. Locking Sequences

Theorem [Blum and Blum 1975, Fulk 1990]
Let $\mathbf{M}$ be an explanatory learner learning a set $\mathbf{L}$. Then there exists a sequence σ of data in $\mathbf{L}$ such that for all further sequences τ from data in $\mathbf{L}$ it holds that $\mathbf{M}(\sigma\tau) = \mathbf{M}(\sigma)$.

Such a sequence is called a stabilising sequence. If it furthermore holds that $\mathbf{W}_{\mathbf{M}(\sigma)} = \mathbf{L}$ then σ is called a locking sequence for $\mathbf{L}$.

Observation [Fulk 1990]
Whenever $\mathbf{M}$ learns $\mathbf{L}$ then every stabilising sequence for $\mathbf{L}$ is also a locking sequence for $\mathbf{L}$.

Properties of Learners

Let $\mathbf{M}$ be a learner for a class $\mathbf{C}$.

Order Independence (Blum and Blum 1975): $\mathbf{M}$ is order independent iff for every $\mathbf{L} \in \mathbf{C}$ and every two texts for $\mathbf{L}$, $\mathbf{M}$ converges on both texts to the same index for $\mathbf{L}$.

Rearrangement Independence (Fulk 1990): $\mathbf{M}$ is rearrangement independent iff for all finite sequences σ, τ with $|\sigma| = |\tau| \wedge \mathbf{content}(\sigma) = \mathbf{content}(\tau)$ it holds that $\mathbf{M}(\sigma) = \mathbf{M}(\tau)$.

Set-Driven (Osherson, Stob and Weinstein 1982): $\mathbf{M}$ is set-driven iff for all finite sequences σ, τ with $\mathbf{content}(\sigma) = \mathbf{content}(\tau)$ it holds that $\mathbf{M}(\sigma) = \mathbf{M}(\tau)$.

Every learner can be replaced by an equivalent learner which is order independent and rearrangement independent, but one cannot always obtain set-drivenness.

Hunting Locking-Seqences

Let $\mathbf{M}$ be a learner. Furthermore, for any $\mathbf{L}$ and any finite sequence σ of some elements in $\mathbf{L}$, let τ_σ be the length-lexicographic least string with

- $\mathbf{content}(\tau_\sigma) \subseteq \mathbf{content}(\sigma)$;
- $\mathbf{M}(\tau_\sigma \eta) = \mathbf{M}(\tau_\sigma)$ for all η with $\mathbf{content}(\eta) \subseteq \mathbf{content}(\sigma) \wedge |\tau_\sigma| + |\eta| = |\sigma|$.

Let $\mathbf{N}(\sigma) = \mathbf{M}(\tau_\sigma)$.

If $\mathbf{M}$ learns $\mathbf{L}$ then there is some sequence ϑ such that every σ with $|\vartheta| \leq |\sigma|$ and $\mathbf{content}(\vartheta) \subseteq \mathbf{content}(\sigma) \subseteq \mathbf{L}$ satisfies $\tau_\sigma = \tau_\vartheta$. $\mathbf{N}$ converges on every text for $\mathbf{L}$ to $\mathbf{M}(\tau_\vartheta)$ and τ_ϑ is the length-lexicographic least locking sequence of $\mathbf{L}$ (or very similar to it).

$\mathbf{N}$ is rearrangement-independent and order-independent.

Set-Drivenness is Restrictive

Let $b_0, b_1, b_2, \ldots$ be a recursive one-one enumeration of K. Now let C contain all sets of the form

- $\{2e\} \cup \{1, 3, 5, 7, \ldots\}$;
- $\{2e\} \cup \{1, 3, 5, \ldots, 2s+1\}$ with $\exists t > s\, [b_t = e]$.

This class has a recursive learner but not a set-driven learner. If M would be a set-driven learner then one can search for the first s with

$$M(2e\,1\,3\,5\,\ldots\,2s{+}1) = M(2e\,1\,3\,5\,\ldots\,2s{+}1\,2s{+}3).$$

and would have that $e \in K \Leftrightarrow e \in \{b_0, b_1, \ldots, b_s\}$.

I.3. Angluin's Tell-Tale Sets

An indexed family is a class given by a hypothesis space $\{L_e : e \in I\}$ such that the mapping $e, x \to L_e(x)$ is recursive.

Theorem [Angluin 1980]
An indexed family $\{L_e : e \in I\}$ is explanatory learnable iff there is a uniformly r.e. family $\{H_e : e \in I\}$ of finite sets such that all d, e satisfy $H_e \subseteq L_e$ and $H_e \subseteq L_d \subseteq L_e \Rightarrow L_d = L_e$.

The finite sets H_e are called tell-tale sets for L_e.

Given a learner M, one can enumerate H_e by searching among the finite sequences σ over L_e the first stabilising sequence to be found and enumerating the range of every candidate considered into H_e.

For the converse direction, consider a learner N which on input $a_0\, a_1\, \ldots\, a_n$ outputs the first e with $e \geq n$ or $H_{e,n} \subseteq \{a_0, a_1, \ldots, a_n\} \subseteq L_e$.

Generalising Angluin's Criterion

Let $\{L_e : e \in I\}$ be a uniformly r.e. class.

Theorem [de Jongh and Kanazawa 1996]
The class $\{L_e : e \in I\}$ is explanatorily learnable iff there are limit-recursive functions mapping each index e to indices i, j such that W_i is finite and W_j is a set of canonical indices of finite sets and for all indices d it holds that

- $W_i \subseteq L_e$;
- $W_i \subseteq L_d \subseteq L_e \Rightarrow L_d = L_e$;
- $\forall k \in W_j\,[D_k \not\subseteq L_e]$;
- $W_i \subseteq L_d \wedge L_d \neq L_e \Rightarrow \exists k \in W_j\,[D_k \subseteq L_d]$.

Conservative Learning

Definition [Angluin 1980]
A learner $\mathbf{M}$ is conservative iff whenever $\mathbf{M}(\sigma\tau) \neq \mathbf{M}(\sigma)$ then some datum $\mathbf{x}$ occurring in $\sigma\tau$ is not inside the hypothesis conjectured by $\mathbf{M}(\sigma)$.

Example
(a) The class of all finite sets is conservatively learnable.
(b) The class of all sets $\mathbb{N} - \{\mathbf{b}\}$ is conservatively learnable.
(c) Every inclusion-free indexed family is conservatively learnable; the learner follows the algorithm “Learning by Enumeration”.

Theorem [Zeugmann, Lange and Kapur 1992]
An indexed family $\{\mathbf{L_e} : \mathbf{e} \in \mathbf{I}\}$ has a conservative learner iff there is a recursive function $\mathbf{S}$ such that $\mathbf{L_e} \cap \{\mathbf{0}, \mathbf{1}, \ldots, \mathbf{S}(\mathbf{e})\}$ is a tell-tale set for each set $\mathbf{L_e}$.

Explanatory versus Conservative Learning

Theorem [Angluin 1980]
There is an indexed family which is explanatorily learnable but not conservatively learnable.

Let $\mathbf{C}$ contain all sets $\{\mathbf{x}, \mathbf{x}+\mathbf{1}, \ldots\}$ and, whenever $\mathbf{x}$ goes into the halting problem $\mathbf{K}$ at time $\mathbf{s}$ then let $\mathbf{C}$ also contain every set which has minimum $\mathbf{x}$ and contains, besides perhaps some other elements, those numbers which are smaller than $\mathbf{x}+\mathbf{s}$.

The class $\mathbf{C}$ has a class-preserving indexed family $\{\mathbf{L_e} : \mathbf{e} \in \mathbf{I}\}$ as hypothesis space.

There are no recursive functions $\mathbf{f}, \mathbf{S}$ such that for all $\mathbf{x}$, $\{\mathbf{x}\} \subseteq \mathbf{L_{f(x)}} \subseteq \{\mathbf{x}, \mathbf{x}+\mathbf{1}, \ldots\}$ and $\mathbf{L_{f(x)}} \cap \{\mathbf{0}, \mathbf{1}, \ldots, \mathbf{S(x)}\}$ is a finite tell-tale set for $\mathbf{L_{f(x)}}$. Otherwise, $\mathbf{S}$ would satisfy $\mathbf{x} \in \mathbf{K} \Leftrightarrow \mathbf{x} \in \mathbf{K_{S(x)}}$, a contradiction.

Uniformly Co-r.e. Classes

A hypothesis space $\{L_e : e \in I\}$ is uniformly co-r.e. iff $\{\langle e, x\rangle : x \notin L_e\}$ is recursively enumerable.

Theorem [Gao 2010: Angluin's criterion for co-r.e.]
A class C of co-r.e. sets is conservatively learnable (with respect to co-r.e. indices) iff there exists a uniformly co-r.e. hypothesis space $\{L_e : e \in I\}$ containing C and a uniformly r.e. family $\{H_e : e \in I\}$ such that all indices $d, e \in I$ satisfy $H_e \subseteq L_e$ and $H_e \subseteq L_d \subseteq L_e \Rightarrow L_d = L_e$.

Theorem [Gao 2010]
An indexed family is conservatively learnable using a uniformly co-r.e. hypothesis space iff it is explanatorily learnable using a uniformly recursive hypothesis space.

Theorem [Gao 2010]
There is a uniformly co-r.e. family which is explanatorily but not conservatively learnable.

Consistent Learning

Definition

A learner $\mathbf{M}$ is called consistent iff for every sequence σ of data it holds that $\mathbf{M}(\sigma)$ outputs a hypothesis containing all data-items from σ.

Proposition

Every learnable indexed family $\mathbf{C}$ has a consistent learner.

If $\mathbf{M}$ is a learner for $\mathbf{C}$, one can make a new consistent learner $\mathbf{N}$ such that $\mathbf{N}$ outputs an index for $\mathbb{N}$ whenever the original hypothesis $\mathbf{M}$ outputs a conjecture not containing all the data observed so far. Note that the learner need not be class-preserving, as $\mathbb{N}$ might not be in $\mathbf{C}$.

Example

In general, consistency is restrictive. The class $\{\mathbf{K}, \mathbb{N}\}$ does not have a consistent learner, as the halting problem $\mathbf{K}$ is undecidable.

I.4. Convergence

Up to now, the main learning criterion was learning in the limit where the learner converges syntactically to a correct hypothesis. In the following, more restrictive notions of convergence (finite, confident) and more relaxed notions of convergence (vacillatory, behaviourally correct) will be introduced.

Note that conservativeness is also related to convergence constraints, as it says that a hypothesis generating all the data seen so far cannot be revised. This implies that correct hypotheses are never revised. In contrast to this, behaviourally correct learning (as defined later) will permit to revise a correct hypothesis infinitely often.

Finite Learning

Definition [Gold 1967]
A learner $\mathbf{M}$ is called finite iff it outputs on every sequence σ of data either a special symbol "$?$" for "no hypothesis" or an index $\mathbf{e}$ which correctly describes the set to be learnt.

Examples
(a) The class of all sets with $\mathbf{8}$ elements is finitely learnable; the learner outputs "$?$" as long as less than $\mathbf{8}$ elements have been seen and conjectures the set of the first $\mathbf{8}$ elements seen, otherwise.
(b) The class of all sets $\mathbf{W_e}$ where $\mathbf{2e}$ is the only unique even element of the set is finitely learnable.
(c) The class $\{\emptyset, \mathbb{N}\}$ is not finitely learnable; in general, every finitely learnable class has to be inclusion-free.
(d) There are inclusion-free classes which are not finitely learnable, for example, $\{\mathbb{N} - \{\mathbf{m}\} : \mathbf{m} \in \mathbb{N}\}$.

Confident Learning

Definition [Osherson, Stob and Weinstein 1986]
A learner $\mathbf{M}$ is confident iff it does on every input text — whatever set this text is for — converge to some index.

Examples
(a) Every finitely learnable class is confidently learnable.
(b) Every finite class is confidently learnable; hence there are confidently learnable classes which are not finitely learnable.
(c) The class $\{\mathbf{E} : \exists \mathbf{e}\, [\{\mathbf{e}\} \subseteq \mathbf{E} \subseteq \{\mathbf{e}, \mathbf{e}+\mathbf{1}, \ldots, \mathbf{2e}\}]\}$ is confidently learnable.
(d) The class $\{\mathbb{N} - \{\mathbf{m}\} : \mathbf{m} \in \mathbb{N}\}$ is not confidently learnable.
(e) The class of all finite sets is not confidently learnable.
(f) If a class is confidently learnable then it does not contain an infinite ascending chain $\mathbf{L_0} \subset \mathbf{L_1} \subset \mathbf{L_2} \subset \ldots$ of sets.

Behavioural Correct Learning

Definition [Barzdins 1974]
A learner $\mathbf{M}$ is behaviourally correct iff it outputs on every text for a set $\mathbf{L}$ to be learnt a sequence of indices such that from some point onwards each index is an index for $\mathbf{L}$.

Remarks
(a) The learner $\mathbf{M}$ converges only semantically, and need not converge syntactically. Note that every expla- natory learner is behaviourally correct, but not vice versa.
(b) The class $\{\mathbf{A} \cup \mathbf{D} : \mathbf{D} \text{ is finite}\}$ where $\mathbf{A}$ is r.e. and not recursive is behaviourally correctly learnable but not explantorily learnable. The behaviourally correct learner conjectures on input $\mathbf{a_0\, a_1\, \ldots\, a_n}$ an index for the set $\mathbf{A} \cup \{\mathbf{a_0, a_1, \ldots, a_n}\}$.
(c) If a hypothesis space is uniformly recursive then every behaviourally correct learner can be replaced by an explantory one learning the same class.

Vacillatory Learning

Definition [Case 1999]
A learner $\mathbf{M}$ is vacillatory iff it is a behaviourally correct learner of the given class which outputs on every text for a language to be learnt only finitely many different indices.

Examples
(a) Every explanatorily learnable class is vacillatorily learnable.
(b) The class of all sets $\mathbf{L}$ for which there is $\mathbf{e} < \mathbf{min}(\mathbf{L})$ with $\mathbf{W_e} = \mathbf{L}$ is vacillatorily learnable but not explanatorily learnable.
(c) The class of all sets $\mathbf{L}$ with at most $\mathbf{2}$ even elements $\mathbf{2i}, \mathbf{2j}$ with $\mathbf{L} \in \{\mathbf{W_i}, \mathbf{W_j}\}$ is vacillatory learnable in a way that the learner on every text eventually vacillates among at most two indices.
(d) The class $\{\mathbf{A} \cup \mathbf{D} :$ is finite$\}$ where $\mathbf{A}$ is r.e. and not recursive is not vacillatorily learnable.

Partial Learning

A partial learner $\mathbf{M}$ outputs on every text for a language $\mathbf{L}$ to be learnt exactly one index infinitely often and all other indices only finitely often; the infinitely often output index is an index for $\mathbf{L}$.

Theorem [Osherson, Stob and Weinstein 1986]
There is a partial learner which learns the whole class of all r.e. languages.

Remark
Partial learners for the class of all r.e. sets are not consistent, that is, every partial learner which learns the class of all r.e. sets outputs on some data $\mathbf{a_0\, a_1\, \ldots\, a_n}$ a hypothesis $\mathbf{e}$ with $\{\mathbf{a_0}, \mathbf{a_1}, \ldots, \mathbf{a_n}\} \not\subseteq \mathbf{W_e}$.

I.5. Memory-Restrictions

An iterative learner is a learner $\mathbf{M}$ such that $\mathbf{M}$ bases its new hypothesis only on the previous hypothesis and the current datum. That is, there is an initial value $\mathbf{o}$ and an update function $\mathbf{f}$ such that $\mathbf{e_0 = f(o, a_0)}$, $\mathbf{e_1 = f(e_0, a_1)}$, $\ldots$, $\mathbf{e_n = f(e_{n-1}, a_n)}$.

Example
The class of finite sets has an iterative learner as each hypothesis $\mathbf{e_D}$ codes in a easily checkable way the set $\mathbf{D}$ of data seen so far and $\mathbf{f(e_D, x) = e_{D \cup \{x\}}}$.

Example
If $\mathbf{L_0 = \{1, 2, \ldots\}}$ and, for $\mathbf{e > 0}$, $\mathbf{L_e = \{0, 1, \ldots, e\}}$ then the resulting class does not have an iterative learner.

More General Memory-Restrictions

The learner $\mathbf{M}$ is described by two recursive functions $\mathbf{f}, \mathbf{g}$ and an initial memory $\mathbf{o}$.

- $\mathbf{f}$ updates the memory by $\mathbf{m_n = f(m_{n-1}, a_n)}$ or $\mathbf{f(o, a_0)}$ when $\mathbf{n = 0}$.
- $\mathbf{g}$ produces the current hypothesis by $\mathbf{e_n = g(m_n, a_n)}$.

The memory $\mathbf{m_n}$ can be the current hypothesis (iterative learning), some selected data observed (bounded example memory), some data bounded by some function $\mathbf{b(a_0, a_1, \ldots, a_n)}$ and so on. Many different restrictions are possible, also conjecture "$?$" might be permitted after convergence when memory restrictions do not permit to memorise the full hypothesis.

The Memory Hierarchy

Theorem [Lange and Zeugmann 1996]
The class consisting of the set $\{1, 2, \ldots\}$ and all sets $\{0, 1, \ldots, b_0\} \cup \{b_1, b_2, \ldots, b_k\}$ where $b_0 < b_1 < \ldots < b_k$ can be learnt with bounded example memory $k+1$ plus current hypothesis but not with bounded example memory k plus current hypothesis.

Proof. The learner with example memory converges on texts not containing 0 to some hypothesis and must then rely on its bounded example memory to archive the $k+1$ largest elements of the data seen so far. When 0 shows up, it makes a mind change and conjectures from then onwards always the finite set generated by the members of its memory. The hypothesis stores whether 0 has been seen. If the bounded example memory contains only k members then the learner cannot store enough information to code up the largest $k+1$ elements seen so far.

Summary

A learner is a machine which reads more and more data and outputs — in parallel — a sequence of hypotheses such that $\mathbf{e_n}$ is based on $\mathbf{a_0\, a_1\, \ldots\, a_n}$. In the basic setting of explantory learning, the $\mathbf{e_n}$ converge syntactically to one index $\mathbf{e}$ such that $\mathbf{W_e}$ is the set to be learnt.

Many variants of this basic settings have been investigated, with different convergence conditions, additional constraints on the quality of each hypothesis, the usage of preassigned hypothesis spaces and constraints on how learners form and update hypotheses.

II. Automatic Structures and Learning

Regular Languages

Languages accepted by finite automaton.
Alternative definition: smallest class of languages which contains all subsets of a finite alphabet Σ and is closed under concatenation, union, intersection, star-operation and complementation.
Concatenation: $\mathbf{00} \cdot \mathbf{1} = \mathbf{001}$; $\mathbf{A} \cdot \mathbf{B} = \{\alpha \cdot \beta : \alpha \in \mathbf{A} \wedge \beta \in \mathbf{B}\}$; $\mathbf{A}^* = \{\lambda\} \cup \mathbf{A} \cup (\mathbf{A} \cdot \mathbf{A}) \cup (\mathbf{A} \cdot \mathbf{A} \cdot \mathbf{A}) \cup \ldots$ (Kleene star).

Convolution

Given $\alpha, \beta \in \Sigma^*$ and $\# \notin \Sigma^*$, $\mathbf{conv}(\alpha, \beta)$ is the sequence all pairs made of $\alpha_\mathbf{n}$ and $\beta_\mathbf{n}$ for $\mathbf{n} = \mathbf{0}, \mathbf{1}, \ldots, \mathbf{max}\{|\alpha|, |\beta|\} - \mathbf{1}$ with $\alpha_\mathbf{n} = \#$ for $\mathbf{n} \geq |\alpha|$ and $\beta_\mathbf{n} = \#$ for $\mathbf{n} \geq |\beta|$.
Relation $\mathbf{R}$ is regular iff $\{\mathbf{conv}(\alpha, \beta) : \mathbf{R}(\alpha, \beta)\}$ is regular.

Automatic Structures

Domain and all relations of the structure are regular.

Examples

Functions

A function is automatic iff its graph is an automatic relation; all functions which are first-order definable in a given set of automatic relations are automatic.

Fibonacci Numbers and Addition [Tan 2008]

Domain $(\mathbf{0}^*\mathbf{01})^*$, addition $+$, comparison $<$; predicate $\mathbf{F}$. Here $\mathbf{a_1 a_2 \ldots a_n}$ represents $\mathbf{F_1 \cdot a_1 + F_2 \cdot a_2 + \ldots + F_n \cdot a_n}$ where $\mathbf{F_1 = 1, F_2 = 1, F_3 = 2, F_4 = 3, F_5 = 5}$.

Various Algebras

There are automatic presentations of the algebra of the eventually constant functions from $\mathbb{N}$ into a given finite field. Similarly for the algebra of finite and cofinite sets.

Rationals [Tsankov 2009]

The group $(\mathbb{Q}, +)$ has no automatic presentation.

Automatic Classes

Definition

An automatic class is a class of languages each contained in a regular set $\mathbf{S}$ such that there is an indexing with regular domain $\mathbf{I}$ for which $\{\mathbf{conv}(\mathbf{i}, \mathbf{x}) : \mathbf{x} \in \mathbf{S} \wedge \mathbf{i} \in \mathbf{I} \wedge \mathbf{x} \in \mathbf{L_i}\}$ is a regular relation.

Examples

1. The class of all sets $\mathbf{x}\Sigma^*$ is automatic where the parameter $\mathbf{x}$ could be used as the index.

2. The class $\{\mathbf{z} \in \Sigma^* : \mathbf{x} \leq_{\mathbf{lex}} \mathbf{z} \leq_{\mathbf{lex}} \mathbf{y}\}$ is automatic where the convolution of $\mathbf{x}$ and $\mathbf{y}$ can be used as an index for the corresponding interval.

3. The finite subsets of $\{\mathbf{2}\}^*$ form an automatic class with the indices ranging over $\mathbf{I} = (\mathbf{0}^*\mathbf{1})^*$ and $\mathbf{2^n} \in \mathbf{L_i}$ iff $\mathbf{n} < |\mathbf{i}|$ and the symbol in $\mathbf{i}$ at position $\mathbf{n}$ is $\mathbf{1}$.

4. The finite subsets of $\{\mathbf{2}, \mathbf{3}\}^*$ do not form an automatic class.

Learning of Automatic Classes

Let $\{L_i : i \in I\}$ be a given automatic class.

General Model
Learner reads data $a_0, a_1, a_2, \ldots$ from L such that every element of L appears in this list.
Learner produces hypothesis e_n after reading a_n.
Learner succeeds to learn L iff almost all e_n are the same index i with $L_i = L$.
The class is learnable iff some recursive learner learns every language in the class.

Theorem [Angluin 1980 adapted to autom. classes]
The given class is learnable iff for every $i \in I$ there is a finite set $D_i \subseteq L_i$ such that there is no $j \in I$ with $D_i \subseteq L_j \subset L_i$.
D_i is called a tell-tale set for L_i.

II.1. Automatic Learners

An automatic learner [Jain, Luo and Stephan 2009] $\mathbf{M}$ works by a sequence of updates.
The learner maintains a long term memory where $\mathbf{E_n}$ is the long term memory before reading $\mathbf{a_n}$.
$\mathbf{M}$ maps $\mathbf{conv(E_n, a_n)}$ to $\mathbf{conv(E_{n+1}, e_n)}$ and this function has to be automatic.

Restrictions
Iterative: Long term memory is last hypothesis.
Bounded Example-Memory: Long term memory consists of up to $\mathbf{c}$ selected input data.
Example-Bounded: Long term memory is a string of length bounded by the length of the longest example seen so far plus a constant.
Hypothesis-Bounded: Long term memory is a string of length bounded by the length of the hypothesis plus a constant.

Examples

1. The class of all $\{z : x \leq_{lex} z \leq_{lex} y\}$ can be learnt by an automatic learner which always remembers the minimum x and maximum y of the data seen so far. The learner works with each of the four types of restrictions on the long-term memory given above.

2. For given n, let the given class contain all sets $L_{x,y_1,y_2,\ldots,y_k} = x\Sigma^* y_1 \Sigma^* y_2 \Sigma^* \ldots \Sigma^* y_k$ with $|y_1 y_2 \ldots y_k| \leq n$. This class is called the class of automatic pattern languages and is automatically learnable. The long-term memory can be bounded by hypothesis-size. [Ong]

3. The class of all sets $L_i = \{0,1\}^* - \{i\}$ with $i \in \{0,1\}^*$ does not have an automatic learner.

4. The class of all sets $L_i = \{x \in \{0,1\}^* : |x| = |i| \wedge x \neq i\}$ with $i \in \{0,1\}^*$ does not have an automatic learner.

Additional Properties

Let a learner produce output $\mathbf{e_n}$ based on $\mathbf{a_0, a_1, \ldots, a_n}$. Let $\{\mathbf{L_i} : \mathbf{i} \in \mathbf{I}\}$ be the class to be learnt.

Confidence

A learner is confident if it converges on any sequence of data, even if this sequence of data does not belong to any language to be learnt.

Consistency

A learner is consistent if $\mathbf{L_{e_n}}$ contains $\mathbf{a_0, a_1, \ldots, a_n}$.

Conservativeness

A learner is conservative if $\{\mathbf{a_0, a_1, \ldots, a_n, a_{n+1}}\} \not\subseteq \mathbf{L_{e_n}}$ whenever $\mathbf{e_{n+1}} \neq \mathbf{e_n}$.

Example

The class of all sets with up to $\mathbf{5}$ elements is confidently, consistently, conservatively and iteratively learnable.

Unary Classes

Theorem [Jain, Luo, Stephan 2009]
If an automatic class $\{L_i : i \in I\}$ satisfies Angluin's tell-tale condition and $L_i \subseteq \{0\}^*$ for all $i \in I$ then it has an automatic, consistent and conservative learner.

Theorem [Jain, Luo, Stephan 2009]
The class consisting of $\{0^m : m \geq 2\}$ and of all $\{0^m : m \leq n\}$ with $n \geq 1$ has an automatic learner for which the long term memory is bounded by example-size. But this class has no iterative learner.

Theorem [Jain, Luo, Stephan 2009]
There is a unary class having a one-one indexing such that it is not conservatively iteratively learnable in this indexing although the class has a conservative learner and a different iterative learner.

Consistency

Theorem [Jain, Luo, Stephan 2009]
Let $\mathbf{L}_\lambda = \{\mathbf{0},\mathbf{1}\}^*$, $\mathbf{L_2} = \{\mathbf{0},\mathbf{1},\mathbf{2}\}^*$ and
$\mathbf{L_y} = \{\mathbf{x} : \mathbf{x} = \mathbf{2}^{|\mathbf{y}|} \vee (\mathbf{x} \in \{\mathbf{0},\mathbf{1}\}^* \wedge \mathbf{y}$ is not a prefix of $\mathbf{x})\}$ for
$\mathbf{y} \in \{\mathbf{0},\mathbf{1}\}^+$.
This class has an automatic iterative learner but no consistent automatic learner.

Theorem [Jain, Luo, Stephan 2009]
Let $\mathbf{L}_\lambda = \{\mathbf{0},\mathbf{1}\}^*$ and $\mathbf{L_y} = \{\mathbf{2}\} \cup \{\mathbf{x} : \mathbf{x} \leq_{\mathbf{lex}} \mathbf{y0}^\omega\}$ for
$\mathbf{y} \in \{\mathbf{0},\mathbf{1}\}^*\mathbf{1}$.
This class has an automatic iterative learner, a consistent automatic learner but no iterative consistent learner.

Fat Texts

Fat text is a form of input where every data item occurs infinitely often and not only once. Such texts are helpful as they permit to overcome problems caused by forgetting data.

Theorem [Jain, Luo, Stephan 2009]
An automatic class can be learnt from fat text using an automatic learner with the long term memory bounded by example-size iff the class satisfies Angluin's tell-tale condition.

Theorem [Jain, Luo, Stephan 2009]
Every automatic class can be partially identified from fat text by an automatic learner with the long term memory bounded by example-size.

II.2. Translations

Jain, Martin and Stephan [upcoming] study the question:
When are all images learnable under a given criterion?
Counterpart to research on robust learning of functions.

What Operators Φ?
Φ given by first-order definition using regular parameters.
Φ is positive: $L \subseteq L' \Rightarrow \Phi(L) \subseteq \Phi(L')$.
Φ preserves noninclusions from the given class:
$L_i \not\subseteq L_j \Rightarrow \Phi(L_i) \not\subseteq \Phi(L_j)$.
Such operators are called translators.

Basic Property
If $\{L_i : i \in I\}$ is an automatic class so is $\{\Phi(L_i) : i \in I\}$.

Text-Preserving Translators
A translator is text-preserving iff
$\forall L \forall x \in \Phi(L) \exists$ finite $F \subseteq L\, [x \in \Phi(F)]$.

Examples

Ascending Chain

Let the class contain all $\mathbf{L_x} = \{\mathbf{y} \in \{\mathbf{0}, \mathbf{1}\}^* : \mathbf{y} <_{\mathbf{ll}} \mathbf{x}\}$ where $\mathbf{x} \in \{\mathbf{0}, \mathbf{1}\}^*$ as well. Every translation of this class is learnable.

The learner conjectures $\mathbf{\Phi(L_x)}$ for the least $\mathbf{x}$ such that $\mathbf{\Phi(L_x)}$ contains all the data seen so far.

One top element with antichain below

Let the class contain $\{\mathbf{0}, \mathbf{1}\}^*$ and all singleton sets $\{\mathbf{x}\}$. Then the class is learnable but the following text-preserving translation destroys learnability:

$$\mathbf{\Phi(L)} = \{\mathbf{y} \in \{\mathbf{0}, \mathbf{1}\}^* : \exists \mathbf{x} \neq \mathbf{y}\, [\mathbf{x} \in \mathbf{L}]\}.$$

The translation contains the full set and all co-single sets; hence Angluin's tell-tale condition fails.

Learnability

Theorem [Jain, Martin, Stephan]
The following is equivalent for an automatic class $\{L_i : i \in I\}$:
(a) Every translation is learnable.
(b) Every text-preserving translation is learnable.
(c) $\forall i \exists b_i \forall j \; [L_j \subset L_i \Rightarrow \exists k \leq_{ll} b_i \; [L_j \subseteq L_k \wedge L_i \not\subseteq L_k]]$.

To see that (c) implies (a), fix Φ and fix the mapping i to b_i. Now the learner conjectures the length-lexicographic least i such that $\Phi(L_i)$ is consistent with the data seen so far such that for all $k \leq_{ll} b_i$ either $\Phi(L_i) \subseteq \Phi(L_k)$ or $\Phi(L_k)$ is inconsistent with the data seen so far.

Note that the learner abstains from conjecturing if no hypothesis qualifies. When learning a set $\Phi(L_i)$, the finitely many $\Phi(L_k)$ with $k \leq_{ll} \max_{ll}\{i, b_i\} \wedge \Phi(L_i) \not\subseteq \Phi(L_k)$ will eventually all be inconsistent with the input data and from that onwards i is conjectured.

Strong-Monotonic Learning

A learner is strong-monotonic iff every new hypothesis is a superset of the old one.

Theorem [Jain, Martin, Stephan]
The following is equivalent for an automatic class $\{L_i : i \in I\}$:
(a) Every translation is strong-monotonically learnable.
(b) Every text-preserving translation is strong-monotonically learnable.
(c) $\forall i \exists b_i \forall j\ [L_j \not\supseteq L_i \Rightarrow \exists k \leq_{ll} b_i\ [L_j \subseteq L_k \wedge L_i \not\subseteq L_k]]$.

Learner as before: Learner conjectures i iff $\Phi(L_i)$ is consistent with the data seen so far and for all $k \leq_{ll} b_i$ either $L_i \subseteq L_k$ or $\Phi(L_k)$ is inconsistent with the data seen so far.

Note that the learner abstains from conjecturing if no hypothesis qualifies; by assumption on the mapping $i \mapsto b_i$ any two subsequent hypotheses i, j satisfy $L_i \subseteq L_j$.

Consistency

Theorem [Jain, Martin, Stephan]
The following is equivalent for an automatic class:
(a) Every translation is consistently strong-monotonically learnable.
(b) Every text-preserving translation is consistently strong-monotonically learnable.
(c) Any two sets in the class are comparable and every set has only finitely many subsets within the class.

Theorem [Jain, Martin, Stephan]
The following is equivalent for an automatic class:
(a) Every translation is consistently conservatively learnable.
(c) Every translation is learnable and there is no infinite descending chain.

Condition on text-preserving translations cannot be added, there is a counter example.

II.3. Uncountable Classes

Jain, Luo, Semukhin and Stephan [ALT 2009] considered the following formalisation of the learning of uncountable classes.

- All sets in the class are subsets of a given regular set;
- The indices are ω-words;
- The indexing is an ω-automatic structure;
- The learner reads more and more positive data from a set $\mathbf{L}_\alpha$ and longer and longer parts of an index β;
- In parallel to this, the learner outputs a sequence $\mathbf{A_0}, \mathbf{A_1}, \mathbf{A_2}, \ldots$ of Muller automata such that, for almost all $\mathbf{n}$, $\mathbf{A_n}$ accepts β iff $\mathbf{L}_\alpha = \mathbf{L}_\beta$.

Formalizing Convergence

Convergence of $\mathbf{A_0}, \mathbf{A_1}, \mathbf{A_2}, \ldots$ according to one of the following criteria.

Explanatory: Almost all $\mathbf{A_n}$ are the same $\mathbf{A}$ such that $\mathbf{A}$ accepts β iff $\mathbf{L}_\alpha = \mathbf{L}_\beta$.

Vacillatory with size k: Almost all $\mathbf{A_n}$ belong to a set $\{\mathbf{B_1}, \mathbf{B_2}, \ldots, \mathbf{B_k}\}$ such that each $\mathbf{B}$ in this set accepts β iff $\mathbf{L}_\alpha = \mathbf{L}_\beta$.

Behaviourally correct: For almost all $\mathbf{n}$, $\mathbf{A_n}$ accepts β iff $\mathbf{L}_\alpha = \mathbf{L}_\beta$.

Partial identification: There is exactly one automaton $\mathbf{A}$ such that $\mathbf{A}$ equals to infinitely many $\mathbf{A_n}$. This automaton $\mathbf{A}$ accepts β iff $\mathbf{L}_\alpha = \mathbf{L}_\beta$.

Indexing-Independent Results

Theorem [Jain, Luo, Semukhin, Stephan 2009]
An ω-automatic class $\{\mathbf{L}_\alpha : \alpha \in \mathbf{I}\}$ is behaviourally correctly learnable iff it satisfies Angluin's tell-tale condition:
$\forall\alpha\ \exists$ finite $\mathbf{D}\ \forall\beta\ \neg[\mathbf{D} \subseteq \mathbf{L}_\beta \subset \mathbf{L}_\alpha]$.

Theorem [Jain, Luo, Semukhin, Stephan 2009]
Every behaviourally correctly learnable ω-automatic class is also vacillatory learnable with some parameter $\mathbf{k}$.

Theorem [Jain, Luo, Semukhin, Stephan 2009]
Every ω-automatic class is partially identifiable.

Indexing-Dependent Results

Theorem [Jain, Luo, Semukhin, Stephan 2009]
There is an ω-automatic and inclusion-free class such that for each $\mathbf{k}$ this class has an indexing such that it is vacillatory learnable with parameter $\mathbf{k}$ but not with any parameter $\mathbf{h} < \mathbf{k}$.

Theorem [Jain, Luo, Semukhin, Stephan 2009]
Every behaviourally correctly learnable ω-automatic class has an indexing such that it is explanatorily learnable with respect to that new indexing.

Blind Learning

A learner is blind iff it does not see the index β but only the data for $\mathbf{L}_\alpha$.

Theorem [Jain, Luo, Semukhin, Stephan 2009]
A class has a blind explanatory learner iff the class satisfies Angluin's tell-tale condition and is countable.

Theorem [Jain, Luo, Semukhin, Stephan 2009]
A class has a blind behaviourally correct learner iff it satisfies Angluin's tell-tale condition.

Theorem [Jain, Luo, Semukhin, Stephan 2009]
A class has a blind partial learner iff it is countable.

Summary

Research with the goal to combine automatic structures with learning.

Basic Principle

Class of languages to be learnt needs to have an automatic indexing, that is, the relation saying which element is in which language (represented by an index) is automatic.

Topics considered today:

1. Automatic learners;
2. Learnability of translations;
3. Learning of uncountable classes.

III. Intrinsic Complexity

Consider the following three collections of languages:
$\mathbf{SINGLE} = \{\mathbf{L} : \mathbf{L}$ is a singleton $\}$.
$\mathbf{COINIT} = \{\mathbf{L} : (\exists \mathbf{n})[\mathbf{L} = \{\mathbf{x} : \mathbf{x} \geq \mathbf{n}\}]\}$.
$\mathbf{FIN} = \{\mathbf{L} :$ cardinality of $\mathbf{L}$ is finite$\}$.
All these classes are explantorily learnable.

However, the “complexity” of learning them is different.

$\mathbf{SINGLE}$: immediately after seeing one example in the input, the language is known.
$\mathbf{COINIT}$: after seeing first input, a bounded amount of uncertainty in the input language.
$\mathbf{FIN}$: will never know when we are at the final language. Moreover, uncertainity is same all the time.

Comparing Classes

Reductions: Similar to that in complexity theory.
$\mathbf{C} \rightarrow \mathbf{D}$.

$$\text{Text for } \mathbf{A} \in \mathbf{C} \quad \Rightarrow \quad \text{Text for } \mathbf{B} \in \mathbf{D}$$
$$\text{Seq. of Hypotheses for } \mathbf{A} \quad \Leftarrow \quad \text{Seq. of Hypotheses for } \mathbf{B}$$

Goal: If $\mathbf{D}$ is explanatorily learnable and $\mathbf{C}$ is reducible to $\mathbf{D}$ then $\mathbf{C}$ is also explanatorily learnable.

First done for learning functions by Freivalds, Kinber and Smith (1995).

We will be mainly concentrating on language identification.

Operators and Sequences

$\mathbf{SEQ}$: finite sequences
An enumeration operator, $\mathbf{\Theta}$, is an algorithmic mapping from $\mathbf{SEQ}$ into $\mathbf{SEQ}$ such that for all $\sigma, \tau \in \mathbf{SEQ}$, if $\sigma \subseteq \tau$, then $\mathbf{\Theta}(\sigma) \subseteq \mathbf{\Theta}(\tau)$.

$\mathcal{T}$ denotes the collection of all texts for all r.e. languages.

We further assume that $\mathbf{\Theta}$ defines a mapping from $\mathcal{T}$ into $\mathcal{T}$.

Let $\mathbf{G}$ range over infinite sequences of grammars.

$\mathbf{G = g_0, g_1, g_2, \ldots}$ converges to a grammar for $\mathbf{content(T)}$ iff there is an $\mathbf{m}$ such that $\mathbf{g_m}$ is a grammar for $\mathbf{content(T)}$ and $\mathbf{g_n = g_m}$ for all $\mathbf{n \geq m}$.

Reductions

Definition [Jain and Sharma 1996]
Weak Reducibility: $\mathbf{C}$ is weak reducible to $\mathbf{D}$ iff there is a recursive operators $\mathbf{\Theta}$ and $\mathbf{\Psi}$ such that for every text $\mathbf{T}$ for a language in $\mathbf{C}$:

(a) $\mathbf{\Theta}$ translates $\mathbf{T}$ to a text $\mathbf{\Theta(T)}$ for a set in $\mathbf{D}$;
(b) $\mathbf{\Psi}$ translates every sequence of grammars converging to $\mathbf{content(\Theta(T))}$ to a sequence converging to a grammar for $\mathbf{content(T)}$.

Strong Reducibility: As weak reducibility with the additional constraint that whenever $\mathbf{T, T'}$ are texts for the same language, so are $\mathbf{\Theta(T), \Theta(T')}$.

Very Strong Reducibility: As strong reducibility with the additional constraint that $\mathbf{\Psi}$ is given as a recursive function $\mathbf{f}$ on indices, that is, $\mathbf{\Psi(g_0, g_1, \ldots, g_k) = f(g_0), f(g_1), \ldots, f(g_k)}$.

III.1. Hardness and Completeness

A class $\mathbf{C}$ is weak-hard for explanatory learning iff every explanotorily learnable class is weak-reducible to $\mathbf{C}$.
A class $\mathbf{C}$ is weak-complete for explanatory learning iff it is explanatorily learnable and weak-hard for explanatory learning.

Similarly for other learning criteria and reducibilities.

Theorem
$\mathbf{C} \leq_{\mathbf{vs}} \mathbf{D} \Rightarrow \mathbf{C} \leq_{\mathbf{strong}} \mathbf{D} \Rightarrow \mathbf{C} \leq_{\mathbf{weak}} \mathbf{D}$.
$\mathbf{D}$ is explanatorily learnable $\Rightarrow$ every $\mathbf{C} \leq_{\mathbf{weak}} \mathbf{D}$ is explanatorily learnable.
$\mathbf{D}$ is behaviourally correctly learnable $\Rightarrow$ every $\mathbf{C} \leq_{\mathbf{vs}} \mathbf{D}$ is behaviourally correctly learnable.
The last statement holds also with “finitely learnable” and “explanatorily learnable” in both places of “behaviourally corrrectly learnable”.

Concrete Classes

Let $\mathbf{FIN} = \{\mathbf{A} : \mathbf{A}$ is finite$\}$;
$\mathbf{COSINGLE} = \{\mathbb{N} - \{\mathbf{x}\} : \mathbf{x} \in \mathbb{N}\}$;
$\mathbf{INIT} = \{\{\mathbf{y} : \mathbf{y} < \mathbf{x}\} : \mathbf{x} \in \mathbb{N}\}$;
$\mathbf{WIEHAGEN} = \{\mathbf{L} : \mathbf{W}_{\min(\mathbf{L})} = \mathbf{L}\}$;
$\mathbf{PATTERN} = \{\mathbf{L} : \mathbf{L}$ is generated by a non-erasing pattern$\}$.

Theorem [Jain and Sharma 1996]
$\mathbf{SINGLE} <_{\mathbf{weak}} \mathbf{COINIT} \equiv_{\mathbf{weak}} \mathbf{PATTERN} <_{\mathbf{weak}} \mathbf{FIN} \equiv_{\mathbf{weak}} \mathbf{COSINGLE} \equiv_{\mathbf{weak}} \mathbf{INIT}$.

$\mathbf{FIN}, \mathbf{COSINGLE}, \mathbf{INIT}$ are weak-complete for explatory learning.

$\mathbf{COSINGLE} <_{\mathbf{strong}} \mathbf{INIT} \equiv_{\mathbf{strong}} \mathbf{FIN} <_{\mathbf{strong}} \{\mathbf{L} \oplus \mathbf{H} : \mathbf{L} \in \mathbf{FIN} \wedge \mathbf{H} \in \mathbf{COINIT}\}$.

Sample Proof

It is shown that $\mathbf{INIT}$ is weak-complete.

Clearly $\mathbf{INIT}$ is explanatorily learnable.

Let $\mathbf{M}$ be a learner for the class $\mathbf{C}$.

$\mathbf{\Phi}$ translates every text on which $\mathbf{M}$ converges to $\mathbf{e}$ to a text for some set $\{\mathbf{0}, \mathbf{1}, \ldots, \langle \mathbf{d}, \mathbf{e} \rangle\}$ for a suitable $\mathbf{d}$; note that $\mathbf{d}$ depends on the text $\mathbf{T}$ and $\mathbf{M}$'s intermediate hypotheses on $\mathbf{T}$.

$\mathbf{\Psi}$ translates every sequence converging to an index $\mathbf{g}$ enumerating $\{\mathbf{0}, \mathbf{1}, \ldots, \langle \mathbf{d}, \mathbf{e} \rangle\}$ to a sequence converging to $\mathbf{e}$; note that $\mathbf{e}$ can be found in the limit from the index $\mathbf{g}$.

A vs-Complete Class

Let $\mathbf{C} = \{\{\langle \mathbf{i}, \mathbf{x} \rangle : \mathbf{x} \in \mathbf{L}\} : \mathbf{M_i}$ learns $\mathbf{L}\}$

(a) Easily seen to be learnable.

(b) Reduction from $\mathbf{C}$ which is learnt by M_i:
$\mathbf{\Theta}$: map every element x in the input to $\langle i, x \rangle$.
$\mathbf{\Psi}$: map $\mathbf{j}$ to $\mathbf{f(j)}$ such that $\mathbf{W_{f(j)}} = \{\pi_\mathbf{2}(\mathbf{x}) : \mathbf{x} \in \mathbf{W_j}\}$.

A Strong-Complete Class

$\mathbf{X_r} = \{\mathbf{code(r)} : \mathbf{0} \leq \mathbf{x} \leq \mathbf{r}, \mathbf{x} \in$ rationals $\}$

$\mathbf{RINIT} = \{\mathbf{X_r} : \mathbf{0} \leq \mathbf{r} \leq \mathbf{1}\}$.

Theorem [Jain, Kinber, Wiehagen 2001]
$\mathbf{RINIT}$ is strong-complete.

The proof is based on some preparing results.

Proposition
There exist a recursive functions $\mathbf{F}$ and ε, from $\mathbf{R_{0,1}}$ to $\mathbf{R_{0,1}}$ such that
(a) $\forall \mathbf{x} \in \mathbf{R_{0,1}}\, [\varepsilon(\mathbf{x}) > \mathbf{0}]$;
(b) $\forall \mathbf{x}\, [\mathbf{0} \leq \mathbf{F}(\mathbf{x}) \leq \mathbf{1}]$;
(c) $\forall \mathbf{x}, \mathbf{y}\, [\mathbf{0} \leq \mathbf{x} < \mathbf{y} \leq \mathbf{1} \Rightarrow \mathbf{F}(\mathbf{x}) + \varepsilon(\mathbf{x}) < \mathbf{F}(\mathbf{y}) \leq \mathbf{1}]$.

Proof Continued

Let $\mathbf{code}(\mathbf{S}) = \sum_{\mathbf{x} \in \mathbf{S}} \mathbf{2}^{-\mathbf{x}-\mathbf{1}}$.

Let $\mathbf{G}(\langle \mathbf{S}, \ell \rangle) = \mathbf{F}(\mathbf{code}(\mathbf{S})) + \varepsilon(\mathbf{code}(\mathbf{S})) - \frac{\varepsilon(\mathbf{code}(\mathbf{S}))}{\ell + \mathbf{2}}$.

Note that, if $\min(\mathbf{S} - \mathbf{S}') < \min(\mathbf{S}' - \mathbf{S})$ or $\mathbf{S} = \mathbf{S}'$ and $\ell > \ell'$, then $\mathbf{G}(\langle \mathbf{S}, \ell \rangle) > \mathbf{G}(\langle \mathbf{S}', \ell' \rangle)$.

Definition

$\langle \mathbf{S}, \ell \rangle$ is full stabilizing sequence for $\mathbf{M}$ on $\mathbf{L}$ iff the following three conditions hold:

(a) $\ell > \max(\mathbf{S})$;

(b) $\forall \mathbf{x} < \ell \, [\mathbf{x} \in \mathbf{L} \Leftrightarrow \mathbf{x} \in \mathbf{S}]$;

(c) $\langle \mathbf{S}, \mathbf{2}\ell \rangle$ is a stabilizing sequence for $\mathbf{M}$ on $\mathbf{L}$ (with respect to content and length).

It can be shown that if a rearrangement independent machine explanatorily learns $\mathbf{L}$, then it has a full stabilizing sequence on $\mathbf{L}$.

Proof Continued

Proposition

Given a rearrangement independent and order independent machine $\mathbf{M}$, there exists a recursive function $\mathbf{H}$ from $\mathbf{SEQ}$ to $\mathbb{N}$ such that the following conditions hold:

(a) $\mathbf{H}(\sigma) = \langle \mathbf{S}, \ell \rangle$ implies $\max(\mathbf{S}) < \ell$;

(b) For σ, τ with $\sigma \subseteq \tau$, $\mathbf{G}(\mathbf{H}(\sigma)) \leq \mathbf{G}(\mathbf{H}(\tau))$;

(c) For all texts $\mathbf{T}$, $\mathbf{H}(\mathbf{T})$ converges to least full stabilizing sequence (the one which minimizes ℓ) for $\mathbf{M}$ on $\mathbf{content}(\mathbf{T})$, if any.

Characterising Strong Completeness

Definition [Freivalds 1975]
$\mathbf{C}$ is limiting standardisable iff there exists a partial limiting recursive function $\mathbf{F}$ such that (a) For all $\mathbf{i}$ with $\mathbf{W_i} \in \mathbf{C}$, $\mathbf{F(i)}$ is defined; (b) For all $\mathbf{i}, \mathbf{j}$ with $\mathbf{W_i}, \mathbf{W_j} \in \mathbf{C}$, $\mathbf{F(i)} = \mathbf{F(j)}$ iff $\mathbf{W_i} = \mathbf{W_j}$.

Theorem [Jain, Kinber, Wiehagen 2001]
$\mathbf{C}$ is strong-complete iff there exists a recursive function $\mathbf{H}$ from $\mathbf{R_{0,1}}$ to $\mathbb{N}$ such that
(a) $\{\mathbf{W_{H(r)}} : \mathbf{r} \in \mathbf{R_{0,1}}\} \subseteq \mathbf{C}$;
(b) If $\mathbf{0} \leq \mathbf{r} < \mathbf{r'} \leq \mathbf{1}$, then $\mathbf{W_{H(r)}} \subseteq \mathbf{W_{H(r')}}$;
(c) $\{\mathbf{W_{H(r)}} : \mathbf{r} \in \mathbf{R_{0,1}}\}$ is limiting-standardisable.

Similar characterisations for strong-reduciblity comparison of classes like $\mathbf{INIT}, \mathbf{COINIT}, \mathbf{SINGLE}, \mathbf{COSINGLE}$ were done by Jain, Kinber, Wiehagen (2001).

Characterising Weak Completeness

Definition [Jain, Kinber, Wiehagen 2001]
A non-empty r.e. class $\mathcal{T}$ of texts is called quasi-dense iff
(a) For distinct $\mathbf{T}, \mathbf{T}' \in \mathcal{T}$, $\mathbf{content}(\mathbf{T}) \neq \mathbf{content}(\mathbf{T}')$.
(b) For each σ, either there exists no text in $\mathcal{T}$ which extend σ, or there exist infinitely many texts in $\mathcal{T}$ which extend σ.

Theorem [Jain, Kinber, Wiehagen 2001]
For any class $\mathbf{C}$ which is explanatorily learnable, $\mathbf{C}$ is weak-complete iff there exists an r.e. quasi dense class of texts $\mathcal{T}$ representing a subclass of $\mathbf{C}$ such that $\{\mathbf{content}(\mathbf{T}) : \mathbf{T} \in \mathcal{T}\}$ is limiting standardisable.

Inclusion-Structure

Theorem [Jain and Sharma 1997]
If $\mathbf{C} \leq_{\mathbf{strong}} \mathbf{D}$ via $\mathbf{\Theta}, \mathbf{\Psi}$ and $\mathbf{A}, \mathbf{B} \in \mathbf{C}$ then
$\mathbf{A} \subseteq \mathbf{B} \Rightarrow \mathbf{\Theta}(\mathbf{A}) \subseteq \mathbf{\Theta}(\mathbf{B})$.

For example, $\mathbf{SINGLE} \leq_{\mathbf{strong}} \mathbf{INIT}$; however, by above result, $\mathbf{INIT} \not\leq_{\mathbf{strong}} \mathbf{SINGLE}$.

Theorem [Jain and Sharma 1997]
Every finite acyclic graph can be embedded into the reducibility structure. That is, if $(\{\mathbf{1}, \mathbf{2}, \ldots, \mathbf{n}\}, \mathbf{E})$ is a finite acyclic graph then there are explanatorily learnable classes $\mathbf{C_1}, \mathbf{C_2}, \ldots, \mathbf{C_n}$ such that $\mathbf{C_i} \leq_{\mathbf{strong}} \mathbf{C_j}$ iff there is a path in the graph from i to j.

Density

Theorem [Jain and Sharma 1997]
The reduction structure is not dense. That is, there are $\mathbf{C}, \mathbf{C}'$ such that $\mathbf{C} <_{\mathbf{strong}} \mathbf{C}'$ and no class $\mathbf{D}$ satisfies $\mathbf{C} <_{\mathbf{strong}} \mathbf{D} <_{\mathbf{strong}} \mathbf{C}'$.

Proof
Let $\mathbf{C} = \{\{\mathbf{0}\}, \{\mathbf{1}\}\}$ and $\mathbf{C}' = \{\{\mathbf{0}\}, \{\mathbf{0}, \mathbf{1}\}\}$.

$\mathbf{C} \leq_{\mathbf{strong}} \mathbf{C}'$ by $\boldsymbol{\Theta}, \boldsymbol{\Psi}$ with $\boldsymbol{\Theta}(\{\mathbf{0}\}) = \{\mathbf{0}\}$ and $\boldsymbol{\Theta}(\{\mathbf{1}\}) = \{\mathbf{0}, \mathbf{1}\}$.

$\mathbf{C}' \not\leq_{\mathbf{strong}} \mathbf{C}$ as $\mathbf{C}$ is inclusion-free.

Let $\mathbf{D} = \{\mathbf{A}, \mathbf{B}\}$ and $\mathbf{C} \leq_{\mathbf{strong}} \mathbf{D} \leq_{\mathbf{strong}} \mathbf{C}'$.
Case $\mathbf{D}$ inclusion-free: there are $\mathbf{x} \in \mathbf{A} - \mathbf{B}$ and $\mathbf{y} \in \mathbf{B} - \mathbf{A}$.
Make $\boldsymbol{\Theta}$ such that sets containing $\mathbf{x}$ are mapped to $\{\mathbf{0}\}$ and sets containing $\mathbf{y}$ are mapped to $\{\mathbf{1}\}$. Hence $\mathbf{D} \leq_{\mathbf{strong}} \mathbf{C}$.
Case $\mathbf{D}$ not inclusion-free, say $\mathbf{A} \subset \mathbf{B}$: Make $\boldsymbol{\Theta}$ such that $\boldsymbol{\Theta}(\{\mathbf{0}\}) = \mathbf{A}$ and $\boldsymbol{\Theta}(\{\mathbf{0}, \mathbf{1}\}) = \mathbf{B}$. Hence $\mathbf{C}' \leq_{\mathbf{strong}} \mathbf{D}$.

III.2. Mitoticity

In recursion theory, a recursively enumerable set $\mathbf{A}$ is called mitotic if it is the union of two disjoint infinite r.e. sets $\mathbf{B_1, B_2}$ such that $\mathbf{A, B_1, B_2}$ are Turing equivalent. Ladner [1973] introduced the notion and showed that some but not all r.e. sets are mitotic.

Ambos-Spies [1984] transferred the concept to complexity theory. A set $\mathbf{A}$ is p-mitotic if it is the disjoint union of two infinite sets $\mathbf{B_1, B_2}$ such that $\mathbf{A, B_1, B_2}$ are polynomial time many-one equivalent.

Our Goal is a theory of mitoticity in inductive inference.

Mitotic Classes

Definition [Jain and Stephan 2008]
A splitting of an infinite class $\mathbf{C}$ are disjoint infinite subclasses $\mathbf{C_0, C_1}$ whose union is $\mathbf{C}$ such that some classifier $\mathbf{M}$ converges on every text of some $\mathbf{L \in C}$ to the index $\mathbf{a \in \{0, 1\}}$ with $\mathbf{L \in C_a}$.
A class $\mathbf{C}$ is weak / strong mitotic iff there is a splitting $\mathbf{C_0, C_1}$ of $\mathbf{C}$ such that $\mathbf{C}$, $\mathbf{C_0}$, $\mathbf{C_1}$ are all weak / strong equivalent.

Remark
Strong mitotic classes are also weak mitotic, but converse does not hold.

Example
$\mathbf{INIT}$ is strong mitotic by splitting into the class of sets with even maximum and the class of sets with odd maximum.

Examples

(1) Let $\mathbf{C}$ contain all sets $\mathbb{N} - \{\mathbf{x}\}$. Then $\mathbf{C}$ is strong mitotic by splitting into $\{\mathbb{N} - \{\mathbf{2x+a}\} : \mathbf{x} \in \mathbb{N}\}$ for $\mathbf{a = 0, 1}$.

(2) Let $\mathbf{D}$ contain $\{\mathbf{0}\}$, $\{\mathbf{0, 1}\}$ and $\{\mathbf{0, 1, 2}\}$ as well as $\{\mathbf{x}\}$ for all $\mathbf{x > 2}$. Then $\mathbf{D}$ is not weak mitotic.
One half contains $\mathbf{A, B}$ with $\mathbf{A \subset B}$ while other half has only disjoint finite sets. This makes a reduction from the first half into the second impossible.

(3) Let $\mathbf{E}$ contain $\mathbb{N}$ and all finite subsets of $\{\mathbf{1, 2, 3, 4, \ldots}\}$.
Then $\mathbf{E}$ is weak mitotic but not strong mitotic.
Idea of not being strong mitotic: If $\mathbf{E_0, E_1}$ is a splitting with $\mathbf{E_0}$ containing $\mathbb{N}$ then the $\mathbf{\Theta}$ of a strong reduction maps $\mathbb{N}$ to a finite set $\mathbf{A}$ and all other sets in $\mathbf{E_0}$ to subsets of $\mathbf{A}$ in a one-one manner, a contradiction.
Weak reductions need not to preserve set-inclusion.

Finite Learning

Theorem [Jain and Stephan 2008]
If $\{L_0, L_1, L_2, \ldots\}$ is a uniformly recursive class (= indexed family) and finitely learnable then $\{L_0, L_1, L_2, \ldots\}$ is strong mitotic.

Example
Let A be a maximal set and let $L_a = A$ for $a \in A$ and $L_a = \{a\}$ for $a \notin A$. Then $\{L_0, L_1, L_2, \ldots\}$ is finitely learnable and uniformly r.e. but not weak mitotic.

Complete Classes

Theorem [Jain and Stephan 2008]
Every strong complete class is strong mitotic.
Every weak complete class is weak mitotic.
Some weak complete class $\mathbf{C}$ is not strong mitotic.

Construction of $\mathbf{C}$
Let $\mathbf{A}$ be a maximal set, that is, $\mathbf{A}$ is r.e. and every r.e. superset of $\mathbf{A}$ is either cofinite or a finite variant of $\mathbf{A}$.
Let $\mathbf{C}$ contain all sets $\{\mathbf{x}, \mathbf{x}+\mathbf{1}, \mathbf{x}+\mathbf{2}, \ldots\}$ with $\mathbf{x} \notin \mathbf{A}$ and all sets $\{\mathbf{x}, \mathbf{x}+\mathbf{1}, \mathbf{x}+\mathbf{2}, \ldots, \mathbf{x}+\mathbf{y}\}$ with $\mathbf{x} \in \mathbf{A}, \mathbf{y} \in \mathbb{N}$.

Types of Splittings

Theorem [Li and Stephan 2010]
An infinite explanatorily learnable class $\mathbf{C}$ has either no splitting or has a splitting $\mathbf{C_0}, \mathbf{C_1}$ such that $\mathbf{C_0}$ and $\mathbf{C_1}$ are not weak-equivalent.

Theorem [Li and Stephan 2010]
There is an infinite explantorily learnable class $\mathbf{C}$ having a splitting such that every splitting $\mathbf{C_0}, \mathbf{C_1}$ satisfies either $\mathbf{C_0} \leq_{\mathbf{vs}} \mathbf{C_1}$ or $\mathbf{C_1} \leq_{\mathbf{vs}} \mathbf{C_0}$.

Sacks’ Splitting Theorem

In recursion theory, Sacks’ Splitting Theorem says that every nonrecursive r.e. set is the disjoint union of two Turing-incomparable r.e. sets.

Theorem (Jain and Stephan 2008)
Every infinite recursively enumerable and explanatorily learnable class has a splitting into two subclasses which are incomparable with respect to weak reducibility.

Remark
The property “r.e. class” is necessary as every infinite class $\mathbf{C}$ has an infinite subclass $\mathbf{D}$ which cannot be split into two infinite classes $\mathbf{D_0, D_1}$ by any recursive classifier.

Autoreducibility

In recursion theory, Takhtenbrot [1970] defined that a set $\mathbf{A}$ is autoreducible iff one can compute $\mathbf{A}(\mathbf{x})$ relative to $\mathbf{A}$ without asking the oracle $\mathbf{A}$ at $\mathbf{x}$. Ladner [1973] showed that an r.e. set is mitotic iff it is autoreducible.

Definition [Jain and Stephan 2008]
A class $\mathbf{C}$ is weak / strong autoreducible iff there is a weak / strong reduction $(\boldsymbol{\Theta}, \boldsymbol{\Psi})$ from $\mathbf{C}$ to $\mathbf{C}$ such that for all $\mathbf{L} \in \mathbf{C}$ and all texts $\mathbf{T}$ for $\mathbf{L}$, the content of $\boldsymbol{\Theta}(\mathbf{T})$ differs from $\mathbf{L}$.

Example
Let $\mathbf{A}$ be a maximal set and $\mathbf{C}$ contain all sets $\{\mathbf{3x}\}, \{\mathbf{3x+1}\}, \{\mathbf{3x+2}\}$ with $\mathbf{x} \notin \mathbf{A}$ as well as $\{\mathbf{3y} : \mathbf{y} \in \mathbf{A}\}$, $\{\mathbf{3y+1} : \mathbf{y} \in \mathbf{A}\}$ and $\{\mathbf{3y+2} : \mathbf{y} \in \mathbf{A}\}$. Then $\mathbf{C}$ is strong autoreducible but not weak mitotic.

Open Problem
Is every strong mitotic set strong autoreducible?

Behaviourally Correct Learning

Theorem [Jain and Stephan 2008]
If $\mathbf{C}$ is strong-complete for behaviourally correct learning and $\mathbf{C_0, C_1}$ is a splitting of $\mathbf{C}$ then either $\mathbf{C \equiv_{strong} C_0}$ or $\mathbf{C \equiv_{strong} C_1}$.
This also holds with “vs” in place of “strong”.

Theorem [Jain and Stephan 2008]
There is an r.e. behaviourally correct learnable class $\mathbf{C}$ which is not weak mitotic such that for every splitting $\mathbf{C_0, C_1}$ of $\mathbf{C}$ either $\mathbf{C_0 <_{strong} C_1}$ or $\mathbf{C_1 <_{strong} C_0}$.

Open Problem
Is every class which is weak complete for behaviourally correct learning also weak mitotic?

Symmetric Splittings

Theorem [Li and Stephan 2010]
There is a behaviourally correct learnable class $\mathbf{C}$ which admits some splittings such that every splitting $\mathbf{C_0}, \mathbf{C_1}$ of $\mathbf{C}$ satisfies $\mathbf{C} \equiv_{\mathbf{vs}} \mathbf{C_0} \equiv_{\mathbf{vs}} \mathbf{C_1}$.

Here one can choose the class $\mathbf{C}$ as both, vs-complete for behaviourally correct learning and vs-incomplete for it.

As mentioned before, such a result is impossible for explanatory learning.

Open Problem
Is there a vacillatorily learnable class $\mathbf{C}$ which admits some splittings such that every splitting $\mathbf{C_0}, \mathbf{C_1}$ of $\mathbf{C}$ satisfies $\mathbf{C_0} \equiv_{\mathbf{vs}} \mathbf{C_1}$.

Summary

Weak, strong and very-strong reducibilities are used to compare the degree of learnability of classes; the explanatorily learnable classes are closed downward with respect to these reducibilities.

Strong and very strong reducibility preserve inclusions.

There are weak-complete, strong-complete and vs-complete classes for explanatory learning.

Splittings and mitoticity can be introduced for learning theory and it can be shown that weak-complete classes for explanatory learning are weak-mitotic and strong-complete classes are strong-mitotic.

Some behaviourally learnable class, which admits a splitting, satisfies that every splitting of it consists of two vs-equivalent classes.