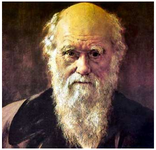
Evolvability in Learning Theory

Dimitris Diochnos

School of Computer Science
University of Oklahoma

Dec 2, 2020
Norman, OK

Evolvability



Evolvability [Valiant, 2009] was based on Darwin's work *On the Origin of Species by Means of Natural Selection* [Darwin, 1859].

Evolvability

Key Points

- Species (**Hypotheses**), Generations (**Iterations**).
- A **fitness** function called **performance**.
  - Estimated through sampling.
- Mutations define the Neighborhood.
- **Tolerance** t **partitions** the Neighborhood:
  - Bene $= \left\{ \mathrm{h}' \mid \mathrm{Perf}_{\mathcal{D}_n}\left(\mathrm{h}', \mathrm{c}\right) > \mathrm{Perf}_{\mathcal{D}_n}\left(\mathrm{h}, \mathrm{c}\right) + t \right\}$.
  - Neut $= \left\{ \mathrm{h}' \mid \mathrm{Perf}_{\mathcal{D}_n}\left(\mathrm{h}', \mathrm{c}\right) \geq \mathrm{Perf}_{\mathcal{D}_n}\left(\mathrm{h}, \mathrm{c}\right) - t \right\} \setminus$ Bene.
  - **Deleterious**, the rest.

Goal

$$\Pr\left(\mathrm{Perf}_{\mathcal{D}_n}\left(\mathrm{h}, \mathrm{c}\right) < \mathrm{Perf}_{\mathcal{D}_n}\left(\mathrm{c}, \mathrm{c}\right) - \varepsilon\right) < \delta. \tag{1}$$

Evolution should proceed from any starting point!

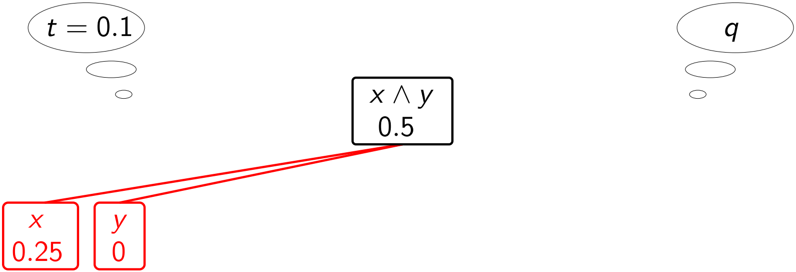
The Swapping Algorithm on Monotone Conjunctions

$t = 0.1$

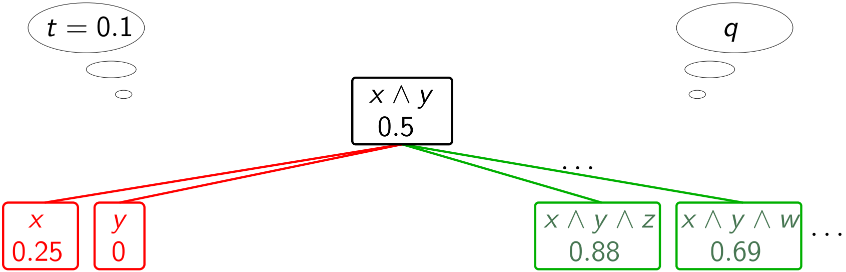
q

$$\boxed{\begin{array}{c} x \wedge y \\ 0.5 \end{array}}$$

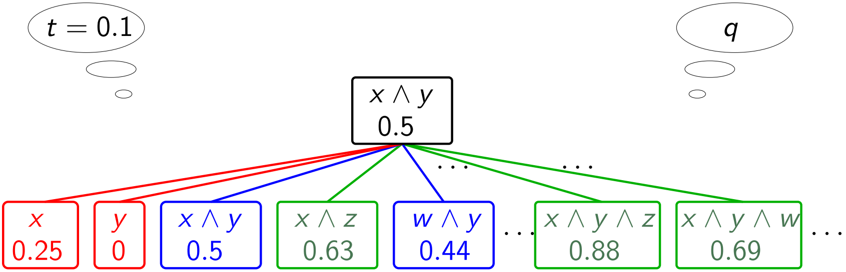
The Swapping Algorithm on Monotone Conjunctions



The Swapping Algorithm on Monotone Conjunctions



The Swapping Algorithm on Monotone Conjunctions



Performance

- $X_n = \{0,1\}^n$.
- $\mathrm{h}(x), \mathrm{c}(x) \in \{+1, -1\}$.

$$
\begin{aligned}
\mathrm{Perf}_{\mathcal{D}_n}(\mathrm{h}, \mathrm{c}) &= \sum_{x \in X_n} \mathrm{h}(x)\mathrm{c}(x)\mathcal{D}_n(x) \\
&= 1 - 2 \cdot \Pr\left(\mathrm{h}(x) \neq \mathrm{c}(x)\right) \\
&= \mathrm{E}\left[\mathrm{h} \cdot \mathrm{c}\right] .
\end{aligned}
$$

- Estimated through sampling,

$$
\mathrm{Perf}_{\mathcal{D}_n}(\mathrm{h}, \mathrm{c}, S) = \frac{1}{|S|} \sum_{x \in S} \mathrm{h}(x) \cdot \mathrm{c}(x) .
$$

Preliminary Remarks

Remark 1 (vs. PAC)

Evolvability is a restricted case of PAC learnability.

Goal 1 (Evolvability)

$$\Pr\left(Perf_{\mathcal{D}_n}(h,c) < Perf_{\mathcal{D}_n}(c,c) - \varepsilon\right) < \delta \ .$$

Goal 2 (PAC Learning)

$$\Pr\left(error_{\mathcal{D}_n}(h,c) > \varepsilon\right) < \delta \ .$$

Preliminary Remarks

Remark 2 (on the *Updates*)

Updates depend only on the positivity and negativity of the examples or experiences, in the sense that there is no dependence on the description of the examples (as is the case in the Statistical Query model); e.g., # of 1's in binary representation.

Remark 3 (vs. *SQ* model, Valiant, 2009)

Evolvable function classes $\subset$ *SQ learnable function classes.*

Preliminary Remarks

Description 1 (The Tool on the *SQ* Model is a Query)

- *Let* $\psi : \{0,1\}^n \times \{-1,1\} \mapsto \{-1,1\}$.
- *A query is a pair* (ψ, τ).
- *Estimate* $\mathrm{E}\left[\psi(x,\ell)\right]$ *within tolerance* τ.

Description 2 (Types of Queries)

- *independent of the target (i.e.* ψ *depends* <u>*only*</u> *on* x*)*
- *correlational if* $\psi(x,\ell) \equiv g(x)c(x)$.

Proposition 1

Any statistical query can be substituted by two statistical queries that are independent of the target and two correlational queries.

A Simulation Result

Remark 4 (*CSQ* Learnability $\Rightarrow$ Evolvability; Feldman 2008)

Let $\mathcal{C}$ be a concept class CSQ learnable over a class of distributions $\mathcal{D}$ by a polynomial time algorithm $\mathcal{A}$. Then, there exists an evolutionary algorithm $N(\mathcal{A})$ such that $\mathcal{C}$ is evolvable by $N(\mathcal{A})$ over $\mathcal{D}$.

Related Results in Evolvability

Feldman
- CSQ → Evolvability algorithm [Feldman, 2008].
- Full conjunctions are evolvable [Feldman, 2009].
- Monotone conjunctions are *not* evolvable distribution-independently using *Boolean loss* [Feldman, 2011].
- Monotone conjunctions *are* evolvable distribution-independently using *quadratic loss* [Feldman, 2012].

D, Turán and D
- Swapping algorithm under $\mathcal{U}_n$ [DT, 2009].
- Swapping algorithm under any $\mathcal{B}_n$ [D, 2016].
- (1+1) EA under some $\mathcal{B}_n$ [D, under submission].

Kanade, Valiant, Vaughan
- Evolvability with drifting targets

Kanade
- Recombination, parallel CSQ learning and general conjunctions [Kanade, 2011].

More Results
- Michael [Michael, 2009], P Valiant [PValiant, 2012], Angelino and Kanade [AK, 2014].

Basic Notation

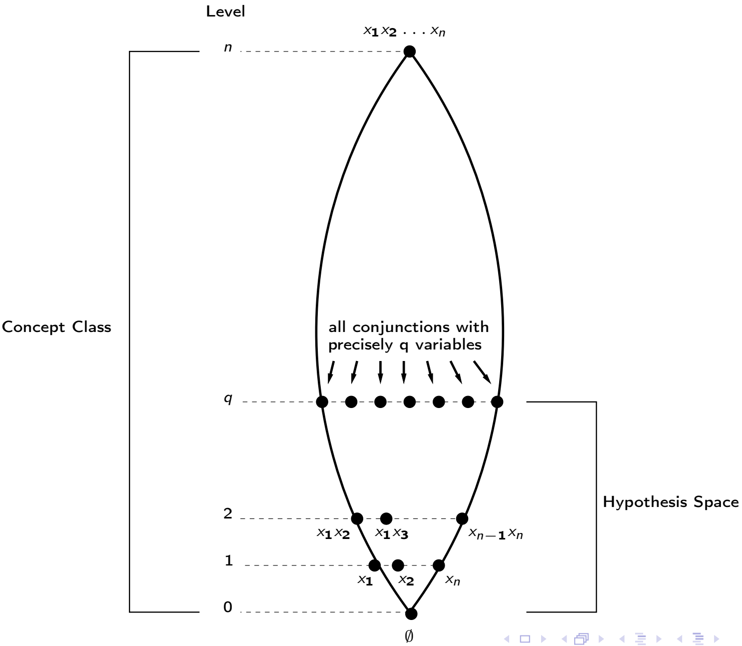
Representation

- Hypotheses are conjunctions of boolean variables; e.g., $\mathrm{h}_1 = x_1 \wedge x_5 \wedge x_8$.
- Size / length: # vars in the conjunction; e.g., $|\mathrm{h}_1| = 3$.
- Represented as a set of indices; e.g., $\mathrm{h}_1 = \{1, 5, 8\}$.
- Also useful: represented by a bitstring; e.g., $\mathrm{h}_1 = 10001001$.
- Hamming distance $d(h_1, h_2)$: # positions where the bitstrings representing h_1 and h_2 differ.

Hypothesis Space

$\mathcal{H} = \mathcal{C}_n^{\leq q}$. Hypotheses such that $0 \leq |\mathrm{h}| \leq q$. ($\leftarrow$ non-realizable)
$\mathcal{H} = \mathcal{C}_n = \mathcal{C}_n^{\leq q} \cup \mathcal{C}_n^{> q}$. Hypotheses such that $0 \leq |\mathrm{h}| \leq n$.

Concept Class and Hypothesis Space



Monotone Conjunctions under the Uniform Distribution are Evolvable

properties	[Valiant, 2007] $\mathcal{H} = \mathcal{C}_n$	[D & Turán, 2009] $\mathcal{H} = \mathcal{C}_n$	[D, 2016] $\mathcal{H} = \mathcal{C}_n^{\leq q}$
q	$\mathcal{O}\left(\lg(n/\varepsilon)\right)$	$\mathcal{O}\left(\lg(1/\varepsilon)\right)$	$\mathcal{O}\left(\lg(1/\varepsilon)\right)$
generations	$\mathcal{O}\left(n \lg(n/\varepsilon)\right)$	$\mathcal{O}\left(n \lg(1/\varepsilon)\right)$	2q
sample size	$\widetilde{\mathcal{O}}\left((n/\varepsilon)^6\right)$	$\widetilde{\mathcal{O}}\left(n^2/\varepsilon^2 + n/\varepsilon^4\right)$	$\widetilde{\mathcal{O}}\left(n/\varepsilon^4\right)$

Theorem 1 (D & Turán, 2009)

Set $q = \lceil \lg(3/\varepsilon) \rceil$. For every target conjunction c and every initial hypothesis h_0 it holds that after $\mathcal{O}\left(q + |h_0| \ln \frac{1}{\delta}\right)$ iterations, each iteration evaluating the performance of $\mathcal{O}\left(nq\right)$ hypotheses, and each performance being evaluated using sample size $\mathcal{O}\left(\left(\frac{1}{\varepsilon}\right)^4 \left(\ln n + \ln \frac{1}{\delta} + \ln \frac{1}{\varepsilon}\right)\right)$ per iteration, the goal is achieved.

Correlation under the Uniform Distribution

$$\mathsf{h} = \overbrace{\bigwedge_{i \in \mathfrak{M}} x_i}^{\textbf{mutual}} \wedge \underbrace{\overbrace{\bigwedge_{\ell \in \mathfrak{R}} x_\ell}^{\textbf{redundant}}}_{\textbf{bad}} \quad \text{and} \quad \mathsf{c} = \underbrace{\bigwedge_{i \in \mathfrak{M}} x_i \wedge \overbrace{\color{red}{\bigwedge_{k \in \mathfrak{U}} x_k}}^{\textbf{undiscovered}}}_{\textbf{good}} \tag{2}$$

$$\begin{aligned} \mathrm{Perf}_{\mathcal{U}_n}(\mathsf{h}, \mathsf{c}) &= 1 - 2^{1-(m+u)} - 2^{1-(m+r)} + 2^{2-(m+r+u)} \\ &= 1 - 2^{1-|c|} - 2^{1-|h|} + 2^{2-|h|-\color{red}{u}} \end{aligned}$$

Strategy

$$\mathsf{h} = \bigwedge_{i \in \mathfrak{M}} x_i \wedge \bigwedge_{\ell \in \mathfrak{R}} x_\ell \quad \text{and} \quad \mathsf{c} = \bigwedge_{i \in \mathfrak{M}} x_i \wedge \bigwedge_{k \in \mathfrak{U}} x_k$$

- Short target $\Rightarrow$ Find target precisely (w.h.p.)
- Long target $\Rightarrow$ Find some good approximation (w.h.p.)

Strategy

$$\mathrm{h} = \bigwedge_{i \in \mathfrak{M}} x_i \wedge \bigwedge_{\ell \in \mathfrak{R}} x_\ell \quad \text{and} \quad \mathrm{c} = \bigwedge_{i \in \mathfrak{M}} x_i \wedge \bigwedge_{k \in \mathfrak{U}} x_k$$

- Short target $\Rightarrow$ Find target precisely (w.h.p.)
- Long target $\Rightarrow$ Find some good approximation (w.h.p.)

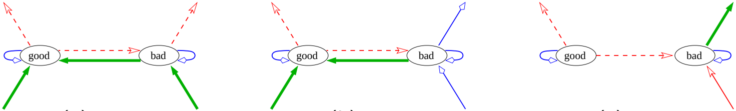
Lemma 2 (Performance Lower Bound)

If $|h| \geq q$ *and* $|c| \geq q + 1$ *then* $Perf_{\mathcal{U}_n}(h, c) > 1 - 3 \cdot 2^{-q}$.

Corollary 3

Let $q \geq \lg(3/\varepsilon)$, $|h| \geq q$, $|c| \geq q + 1 \Longrightarrow Perf_{\mathcal{U}_n}(h, c) > 1 - \varepsilon$.

Guiding the Search



(a) $u \geq 2$ (b) $u = 1$ (c) $u = 0$

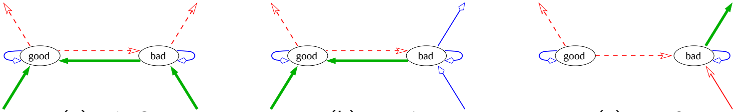
$$\Delta = \mathrm{Perf}_{\mathcal{U}_n}\left(\mathrm{h}', \mathrm{c}\right) - \mathrm{Perf}_{\mathcal{U}_n}\left(\mathrm{h}, \mathrm{c}\right)$$

Theorem 4 (Structure of Best Approximations)

The best q-approximation of a target c is

- *c itself if* $|c| \leq q$
- *any hypothesis formed by q good variables if* $|c| > q$.

Example 1: Short Initial Hypothesis and Short Target



(a) $u \geq 2$ (b) $u = 1$ (c) $u = 0$

Let $X_8 = \{0,1\}^8$ such that $\{g_1, g_2, g_3, b_1, b_2, b_3, b_4, b_5\}$, the target be $c = g_1 \wedge g_2 \wedge g_3$, and require $\varepsilon = 1/5$. $(q = 4)$

Step i	u	Hypothesis h_i	Performance	Neighborhood	Class
0	≥ 2	$\emptyset$	$-3/4$	N^+	Bene
1		b_1	0	$N^+ \cup \{\text{swaps: } b \to g\}$	
2		$b_1 \wedge b_2$	$3/8$	$N^+ \cup \{\text{swaps: } b \to g\}$	
3		$b_1 \wedge b_2 \wedge b_3$	$9/16$	$N^+ \cup \{\text{swaps: } b \to g\}$	
4		$b_1 \wedge b_2 \wedge b_3 \wedge b_4$	$21/32$	$\{\text{swaps: } b \to g\}$	
5		$b_1 \wedge g_3 \wedge b_3 \wedge b_4$	$22/32$	$\{\text{swaps: } b \to g\}$	
6	1	$g_1 \wedge g_3 \wedge b_3 \wedge b_4$	$24/32$	$\{\text{swaps: } b \to g\}$	
7	0	$g_1 \wedge g_3 \wedge g_2 \wedge b_4$	$28/32$	$\{\text{remove } b\}$	
8	0	$g_1 \wedge g_3 \wedge g_2$	1	$\{h_8\}$	Neut

Example 2: Short Initial Hypothesis and Long Target

Let $X_{13} = \{0,1\}^{13}$ such that $\{g_1, g_2, g_3, g_4, g_5, g_6, g_7, b_1, b_2, b_3, b_4, b_5, b_6\}$, the target be $c = g_1 \wedge g_2 \wedge g_3 \wedge g_4 \wedge g_5 \wedge g_6 \wedge g_7$, and require $\varepsilon = 1/5$. $(q = 4)$

Step i	u	Hypothesis h_i	Performance	Neighborhood	Class
0	≥ 2	$\emptyset$	$-63/64$	N^+	Bene
1		b_1	0	$N^+ \cup \{\text{swaps: } b \to g\}$	
2		$b_1 \wedge b_2$	$63/128$	$N^+ \cup \{\text{swaps: } b \to g\}$	
3		$b_1 \wedge b_2 \wedge b_3$	$189/256$	$N^+ \cup \{\text{swaps: } b \to g\}$	
4	≥ 2	$b_1 \wedge b_2 \wedge b_3 \wedge b_4$	$425/512$	$\{\text{all swaps}\} \cup \{h_4\}$	Neut
5		$b_1 \wedge b_6 \wedge b_3 \wedge b_4$	$425/512$	$\{\text{all swaps}\} \cup \{h_5\}$	
6		$b_1 \wedge b_6 \wedge b_3 \wedge b_5$	$425/512$	$\{\text{all swaps}\} \cup \{h_6\}$	
7		$b_1 \wedge b_6 \wedge b_3 \wedge b_5$	$425/512$	$\{\text{all swaps}\} \cup \{h_7\}$	
8	≥ 2	$g_1 \wedge b_6 \wedge b_3 \wedge b_5$	$426/512$	$\{\text{swaps: } b \to g\}$	Bene
9		$g_1 \wedge b_6 \wedge b_3 \wedge g_4$	$428/512$	$\{\text{swaps: } b \to g\}$	
10		$g_1 \wedge b_6 \wedge g_6 \wedge g_4$	$432/512$	$\{\text{swaps: } b \to g\}$	
11	≥ 2	$g_1 \wedge g_3 \wedge g_6 \wedge g_4$	$440/512$	$\{\text{swaps: } g \to g\} \cup \{h_{11}\}$	Neut
12		$g_1 \wedge g_3 \wedge g_5 \wedge g_4$	$440/512$	$\{\text{swaps: } g \to g\} \cup \{h_{12}\}$	
13		$g_1 \wedge g_3 \wedge g_5 \wedge g_4$	$440/512$	$\{\text{swaps: } g \to g\} \cup \{h_{13}\}$	
14		$g_2 \wedge g_3 \wedge g_5 \wedge g_4$	$440/512$	$\{\text{swaps: } g \to g\} \cup \{h_{14}\}$	