A Brief Introduction to Teaching Complexity

Farnam Mansouri,
Advisor: Dr. Adish Singla

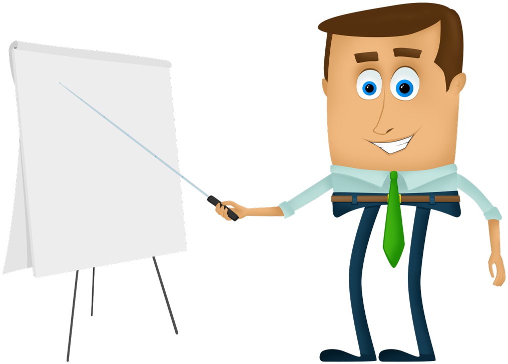
Sharif University of Technology

fmansouri@ce.sharif.edu

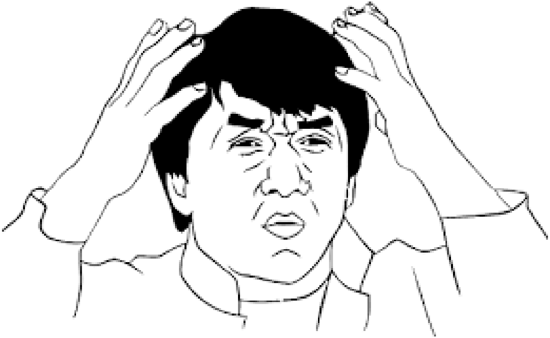
October 24, 2018

Overview

What is teaching?



Why Teaching, though?



Why Teaching?

Some Applications:

- Human Robot Interactions.
  - Develop algorithm for understanding that our system is being taught.
  - gaining insight how to interact with robot instructor.
  - Inverse Reinforcement Learning (IRL).
- Education. (personalized education)
- Trustworthy AI.(Adversarial Attack)

Basic Definitions

Definition (Concept Class)

$X = \{x_1, x_2, ..., x_n\}$ is set of instances; then $c : X \to \{0, 1\}$ is called concept. Also a set of concepts is called concept class.

Definition (incidence matrix of C)

$A^i_j = 1$ iff $x_i \in c_j$

Basic Definitions (cont.)

Definition (labeled sequence)

$z|f = \{(x, f(x)) : x \in z\}$

Definition (version space ($C_{z|f}$))

c is consistent with $z|f$ if $z|f = z|c$

$C_{z|f} = \{c \in C : \text{c is consistent with } z|f\}$

Definition (intersections of z with concept class C)

$\Pi_c(z) = \{c \cap z : c \in C\}$

Basic Definitions (cont.)

Definition (VC dimension (VCD))

C is shattered by z if $\Pi_c(z) = 2^z$
$\Pi_C(d) = max_{|z|=d} \Pi_c(z)$
$VCD(C) = max_d : \Pi_C(d) = 2^d$

Definition

C is (x, y) concept class is a concept class, if $\Pi_C(x) \leq y$.

Basic Definitions (cont.)

Definition

$f(x, y) = max_{C \in (x,y)} TD_{min}(C)$

Definition (monotonic function)

A function on concepts classes is called monotonic if, $\forall C' \subseteq C, K(C') \leq K(C)$, and is called twofold monotonic if K is monotonic and, $\forall X' \subseteq X, K(C_{|X'}) \leq K(C)$.

Teacher-Directed model

- Teaching-Directed complexity for C with respect to c_t is $M_{td}(C, c_t)$, which is defined as $min_{|z|}|C_{z|t}| = 1$
- We call z minimum sequence for c_t
- A teaching set for C with respect to c is $\mathcal{TS}(C, c)$, is a all sets that are only consistent with c and no other concept.

Teacher-Directed model (cont.)

Teaching complexity of C:

- $M_{td-worst}(c) = max_{c_t \in C}(M_{td}(C, c_t))$
- $M_{td-best}(c) = min_{c_t \in C}(M_{td}(C, c_t))$
- $M_{td-average}(c) = \sum_{c \in C} P(c) M_{td}(C, c_t)$ (with respect to distribution P)

Recursive Teaching Dimension

Definition (teaching plan)

P is sequence $((c_1, S_1), ..., (c_N, S_N))$, with following properties:

- $C = \{c_1, ..., c_n\}$
- $\forall 1 \leq t \leq N : S_t \in \mathcal{TS}(c, \{c_t, ..., c_n\})$

- $ord(P) := max_{t=1,...,N-1}|S_t|$ is called order of teaching plan.
- Recursive Teaching Dimension $RTD(C) := min\{ord(P)|P$ is a teaching plan for $C\}$

Properties of RTD

- RTD is monotonic. (Also VCD is monotonic too)
- a teaching plan in canonical form is choosing the easiest to learn concept every time, i.e., $|S_t| = TD_{min}(c_t, \{c_t, ..., c_N\})$.
- RTD is equal to any teaching plan in canonical form.
- $RTD(C) = max_{C' \subseteq C} TS_{min}(C')$

Lemma

if K is monotonic and $\forall C : TD_{min}(C) \leq K(C)$, then $\forall C : RTD(C) \leq K(C)$

Sample Compression Schemes

Definition (Compression Schemes with Information Q)

Let $|X| = n$ and

$$L_C(k_1, k_2) = \{(Y, y) : Y \subseteq X, k_1 \leq |Y| \leq k_2, y \in C_{|Y}\}$$

be the set of labeled samples from C, of sizes between k_1 and k_2. A k-sample compression scheme for C with information Q, for a finite set of Q, consists of tow maps κ, ρ for which the following hold:

- κ (the compression map)

  $$\kappa : L_C(1, n) \to L_C(0, k) \times Q$$

  takes (Y, y) to $((Z, z), q)$ with $Z \subseteq Y$ and $y_{|Z} = z$.

Sample Compression Schemes (cont.)

Definition (Compression Schemes with Information Q (cont.))

- ρ (the reconstruction map)

$$\rho : L_C(o, k) \times Q \to \{0, 1\}^X$$

is so that for all (Y, y) in $L_C(1, n)$,

$$\rho(\kappa(Y, y))_{|Y} = \{(Y, y)\}$$

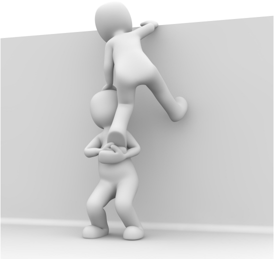
Example of Compression Schemes

Note that: Here we are looking for examples of **k-sample compression scheme with no additional**

rectangles: Consider Class of axis parallel rectangles in $\mathcal{R}^2$; the point within a rectangles are labeled '1', and others '0'. Now compression function only saves the leftmost, rightmost, top and bottom point, so he always saves only 4 points with the labels. Consider the smallest rectangle consistent with all 4, now label every sample according to this rectangle, it is guaranteed to be consistent with original samples. Note that VC-Dimension of this class is also 4.

The Big Motivation

- Classic teaching dimension (TD_{worst}) can be generally exponential of VC-Dimension.
- We'll use RTD instead as a notion of complexity.
- $RTD(C) = O(VCD(C))$?



The Big Motivation (cont.)

- Another ambition for proving above equation is that it'll drive that there exists $O(VCD(C))$-sample compression scheme for C.
- The question above has been open for about 40 years.

Quadratic Upper Bound on RTD

Lemma

For any x,y,z, that $y \leq 2^x - 1$, and $z \leq 2y + 1$ following inequality holds:

$$f(x+1,z) \leq f(x,y) + \lceil \frac{(y+1)(x-1)+1}{2y-z+2} \rceil$$

Proof.

Imagine C is a $(x+1,z)$, We'll define $C_b^Y = \{c \in C : C_{|Y} = b\}$. Also We'll denote $k = \lceil \frac{(y+1)(x-1)+1}{2y-z+2} \rceil$. we'll denote Y^*, b^*, smallest size C_b^Y which is not empty and $|Y| = |b| = k$. Without loss of generality imagine $Y^* = [k]$ and $b^* = 0$. Now, we only need to prove $C_{b^*}^{Y^*}$ is (x,y) class. Assume for the sake of contradiction that previous statement is false, it means that there exist a $|Z| \leq x$, which $|\{c_{|Z} : c \in C_{b^*}^{Y^*}\}| \geq y+1$ (Generally assume that $Z \cap Y^* = \varnothing$, otherwise consider $Z \backslash Y^*$, instead of Z (Y^* has only one pattern which is zero so $Z \backslash Y^*$ cannot be empty)). □

Quadratic Upper Bound on RTD (cont.)

Proof.

Now Define,

$$C_{b^*}^{Y^*,Z} = \{c_{|Z} : c \in C_{b^*}^{Y^*}\}.$$

Since C is $(x+1, z)$ class, the projection of C on the set $Z \cup \{w\}$ has no more than z patterns. Thus

$$|C_{b^*}^{Y^*,Z}| + |C_1^{\{w\},Z}| \leq z$$

we know $|C_{b^*}^{Y^*,Z}| \geq y$, so $|C_1^{\{w\},Z}| \leq z - y - 1$. Now find $\tilde{C}_{b^*}^{Y^*,Z} \subseteq C_{b^*}^{Y^*,Z}$ so that $|\tilde{C}_{b^*}^{Y^*,Z}| = y + 1$. $\forall w \in Y^* : |\tilde{C}_{b^*}^{Y^*,Z} \setminus C_1^{\{w\},Z}| \geq 2y + 2 - z \geq 1$. Thus,

$$\sum_{w \in Y^*} |\tilde{C}_{b^*}^{Y^*,Z} \setminus C_1^{\{w\},Z}| \geq k(2y - z + 2) > (y+1)(x-1)$$

$$= |\tilde{C}_{b^*}^{Y^*,Z}|(x-1) \geq |\tilde{C}_{b^*}^{Y^*,Z}|(|Z| - 1).$$

Proof.

It then follows from the Pigeonhole Principle that there exists $W \subseteq Y^*$ such that $|W| = |Z|$ and $\bigcap_{w \in W}(C_{b^*}^{Y^*,Z} \setminus C_1^{\{w\},Z}) \neq \varnothing$. Pick any string in it called s, now just like previous section $C_{0 \circ s}^{(Y^* \setminus W) \cup Z}$ is a non empty and proper subset of $C_{b^*}^{Y^*}$, which is contradiction. $\square$

Corollary

$\forall 1 < \alpha < 2 : f(X, \lfloor \alpha^x \rfloor) \leq \frac{(x-1)^2}{4-2\alpha} + \frac{3-2\alpha}{4-2\alpha}(x-1) = o(X^2)$

Lemma

For some constant c for every $x > cd$, $(\frac{ex}{d})^d \leq \alpha^x$.

Quadratic Upper Bound on RTD (cont.)

Theorem

$RTD(C) = O(VCD(C)^2)$.

Proof.

By Saur's lemma we know that every concept class with VC-Dimension d, is $(x, (\frac{ex}{d})^d)$, after using previous lemma we'll drive that $\forall x > cd$, every concept class with VC-Dimension d is $(x, \lfloor \alpha^x \rfloor)$ class. Using the corollary we'll drive $TD_{min}(C) \leq O(x^2) = O((cd)^2) = O(d^2)$. Finally using lemma 11 we'll drive $RTD(C) = O(VCD(C)^2)$. $\square$

Teaching Dimension w.r.t Global Preference Function

We call a teaching set of C w.r.t a global preference function $\sigma(c)$ as it follows,

1. Teacher chooses an teaching example $z_t = (x_t, \lambda)$ and updates current version space $C_t = C_{t-1} \cap H$, where H is all the hypotheses which are consistent with z_t.
2. Learner chooses $c_t = argmax_{c \in C_t} \sigma(c)$
3. If $c_t = h^*$, then $Z_t = \{z_1, ..., z_t\}$ is a teaching teaching set of concept class C w.r.t a global preference function $\sigma(c)$

and we call teaching dimension of concept class C w.r.t global preference $\sigma(c)$, $TD_{\sigma(c)} = max_{h^*}|Z_t|$, and best teaching dimension of C w.r.t global preference $TD_{local}(C) = min_{\sigma(c)} TD_{\sigma(c)}$.

Teaching Dimension w.r.t Local Preference Function

We call a teaching set of C w.r.t a local preference function $\sigma(c_t, c_{t-1})$ as it follows,

1. Teacher chooses an teaching example $z_t = (x_t, \lambda)$ and updates current version space $C_t = C_{t-1} \cap H$, where H is all the hypotheses which are consistent with z_t.
2. Learner chooses $c_t = argmax_{c \in C_t} \sigma(c, c_{t-1})$
3. If $c_t = h^*$, then $Z_t = \{z_1, ..., z_t\}$ is a teaching teaching set of concept class C w.r.t a global preference function $\sigma(c)$

and we call teaching dimension of concept class C w.r.t local preference $\sigma(c)$, $TD_{\sigma(c)} = max_{h^*} |Z_t|$, and best teaching dimension of C w.r.t global preference $TD_{local}(C) = min_{\sigma(c)} TD_{\sigma(c)}$.

Teaching Dimension for Global Preference Functions vs. RTD

Theorem

Teaching Dimension w.r.t Global Preference Function is Equivalent to RTD.

Proof.

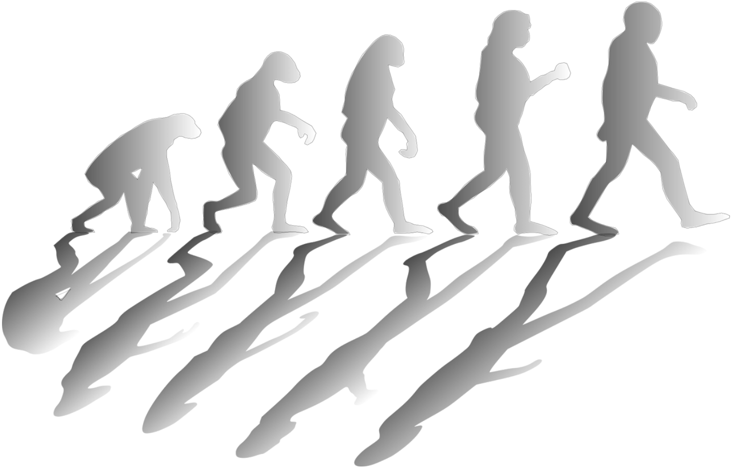
Idea: Order teaching plan in the way that higher preference is first in the list. □

Main Goal

- As explained before, the main motivation is proving $RTD(C) = O(VCD)$.
- Teaching Dimension w.r.t Global Preference Function is obviously less than Teaching Dimension w.r.t Local Preference Function.
- We'll try to prove to prove Teaching Dimension of C w.r.t Local Preference Function is $O(VCD(C))$.

Main Goal (cont.)

- this question is very interesting since Teaching Dimension w.r.t Local Preference Function is more near to human behaviour.



Resources

- Thorsten Doliwa and Gaojian Fan and Hans Ulrich Simon and Sandra Zilles(2014)
  Recursive Teaching Dimension, VC-Dimension and Sample Compression
  Journal of Machine Learning Research 15, 3107-3131.
- Christian Kuhlmann(1998)
  On teaching and learning intersection-closed concept classes
  European Conference on Computational Learning Theory 15, 168–182.
- Lunjia Hu and Ruihan Wu and Tianhong Li and Liwei Wang(2017)
  Quadratic Upper Bound for Recursive Teaching Dimension of Finite VC Classes
  http://arxiv.org/abs/1702.05677
- Shay Moran and Amir Shpilka and Avi Wigderson and Amir Yehudayoff(2015)
  Teaching and compressing for low VC-dimension
  CoRR
- Chen, Yuxin and Singla, Adish and Mac Aodha, Oisin and Perona, Pietro and Yue, Yisong (2018)
  Understanding the Role of Adaptivity in Machine Teaching: The Case of Version Space Learners
  arXiv preprint arXiv:1802.05190

The End