COMS 4252: Introduction to Computational Learning Theory

Rocco Servedio

Overview of today's lecture:

- ✓ Introductions (you, me, topic)
- ✓ Admin. overview (web page)
- → High level course overview (learning models) ←
- Some technical material: Key basic notions + terminology

$\boxed{\text{Questions?}}$

CLT: study of mach. learning (ML) questions from a theor. CS perspective

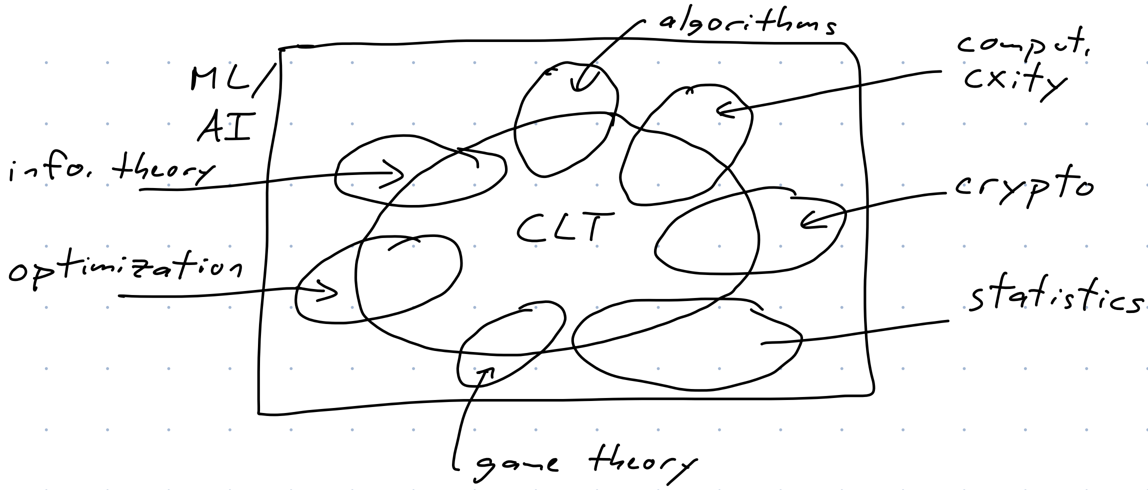
- well defined models
- proving results
- centrally concerned w/ comput. efficiency

ML:
- study of systems that improve perf. given more data;
- deriving programs from experience/data.

Applic./examples:

- classification/clustering
- prediction
- highly cx software: AI, self-driving cars, etc.

CLT: theory of ML. Lots of conn. to TCS topics:



Admin:

- for admin: coms4252columbiaF25@gmail.com
- for course content: Ed Discussion

Course Overview

CLT: 2 parts:

1) defining comput. models of learning — 1% "rules of chess"

2) Proving results abt model: giving algs, lower bds (lb), developing a theory. — 99% learning how to play

Learning model: specifies framework for learning. Some aspects:

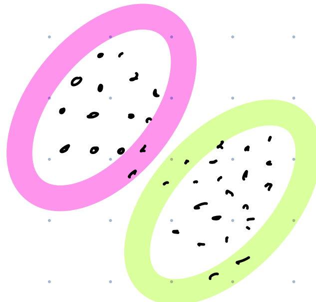
* Supervised vs Unsupervised

example, label

$(x^1, f(x^1))$
$(x^2, f(x^2))$
$\vdots$

labeled data;
classification

no labels!
Clustering



This class: Supervised learning

* What is being learned?

→ us: Binary classif. rules, aka
Boolean functions

$$f: X \longrightarrow \{0,1\}$$

$$f(x) = 0 \text{ or } 1.$$

"concept learning"; f is a "concept" (elephant)

* How does learner get info?

- passive models: learner is given examples
- active models: " can actively

make queries

- quality of info: noise in data?

* Prior knowledge/assumptions?

We'll typically assume that f is "simple" — has a simple syntactic represention.

* Performance criteria: what does it mean to learn successfully?
  - Comput. efficiency?
  - data efficiency?
  - memory ?
  - accuracy/convergence?

We'll discuss several models:

- Online mistake bound learning
- Probably Approximately Correct learning
- learning from "statistical queries"
- Exact learning from membership + equivalence queries

Things we'll do:

- Describe + analyze specific alg's for specific problems
- general techniques for designing learning alg's
- how much data needed to learn?
- Comput. barriers to efficient learning
- Noise?
- Boosting?
- Comparisons between models

Basic Notions/notation (occur across all our learning models!)

Notation:

X = domain of fn f we're learning. "instance space"

$f: X \to \{0,1\}$
$\{-1,1\}$

E.g. X = all cars in the worlds.

For us, $X = \{0,1\}^n$ or $X = \mathbb{R}^n$

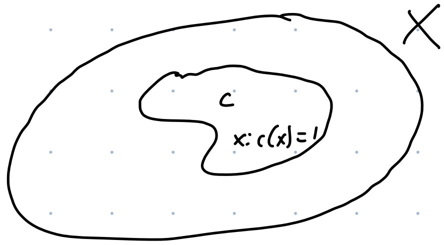
$>\$50{,}000$

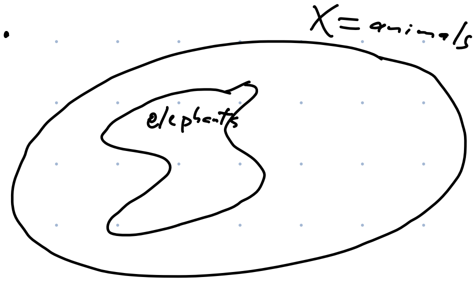
$x = (x_1, \ldots, x_n) \in X$

red American

A **concept** c or f is $c : X \to \{0,1\}$

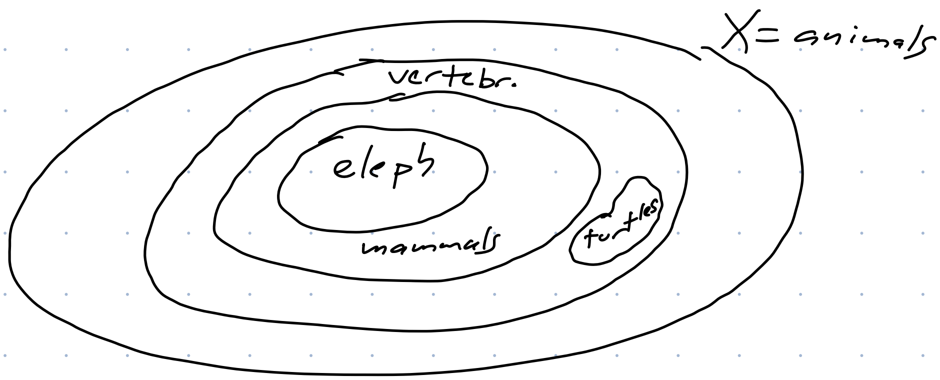
or equiv., $c \subseteq X$.





A **concept class** $\mathcal{C}$ is a set of concepts.

$\mathcal{C} = \{$ elephants, mammals, vertebrates, turtles $\}$



Basic idea of all our learning models:

- There's a **known** $\mathcal{C}$, & an **unknown** **target concept** $c \in \mathcal{C}$.

- Learner has some source of info. about how c labels examples $x \in X$. Want to find/approximate c.

Some examples of the kinds of $\mathcal{C}$'s we'll study:

- $X = \underline{\{0,1\}^n}$ $\qquad 0 \leftrightarrow$ False
  $x = (x_1, \dots, x_n)$ $\qquad 1 \leftrightarrow$ True

  $\mathcal{C}$ = all Boolean $\underline{\overset{\text{disjunctions}}{\text{conjunctions}}}$, of any length

  A $\underline{\text{literal}}$ is a variable x_i or $\overline{x_i}$ (negated variable)

  A $\underline{\overset{\text{disjunction}}{\text{conjunction}}}$ is an AND ($\wedge$) $\overset{\text{OR } (\vee)}{}$ of literals over disjoint var's.

  $c(x) = x_3 \vee \overline{x_4} \vee x_5$

  e.g. $c(x) = \underline{x_2 \wedge x_7 \wedge \overline{x_{14}}}$

  $c : \{0,1\}^n \to \{0,1\}$

  $|\mathcal{C}| = 3^n$ $\qquad$ $\{\,—,\ x_i,\ \overline{x_i}\,\}$ $\qquad i = 1, \dots, n.$

  too big for brute force....

- $X = \{0,1\}^n$, $\mathcal{C}$ = all $\underline{\text{monotone}}$ $\overset{\text{disj.}}{\text{conj.}}$

  no negations

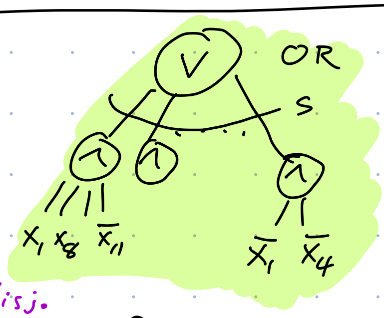
$c(x) = x_2 \wedge x_8 \wedge x_9 \wedge x_{10}$

$|\mathcal{C}| = 2^n$

- $X = \{0,1\}^n$, $\mathcal{C}$ = all s-term DNF

"term" = "conjunction"

Disjunctive Normal Form formulas:



disj.

an OR of s conjunctions

$|\mathcal{C}| \approx 3^{ns}$ ....

harder to learn than conj...

- $X = \{0,1\}^n$, $\mathcal{C}$ = all k-DNF formulas

an OR of any # of conjunctions of length $\leq k$

Ex: $k = 3$: