Computational Learning Theory
There is no Free Lunch

Dimitris Diochnos
School of Computer Science
University of Oklahoma

Outline

Table of Contents

What is Machine Learning?

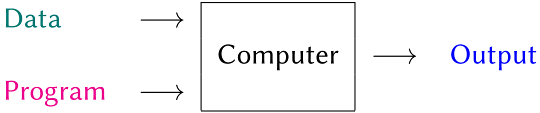
- **Machine learning** is the subfield of computer science that gives "computers the ability to learn without being explicitly programmed".

  - term coined by Arthur Samuel in 1959 while at IBM

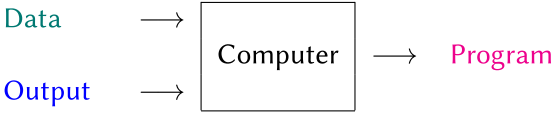
- The study of algorithms that can learn from data.

Another View of Machine Learning

- Learning from historical data to make decisions about unseen data.
- Traditional Programming



- Machine Learning



When is Machine Learning a Good Idea?

- Situations where ...
  - humans can not describe how they do a task
    - character recognition
  - the desired function changes frequently
    - recommend stock transactions
  - each user needs a customized function f
    - email spam / ham
    - email importance (perhaps delete without presenting?)
    - recommendations on Amazon

Can you write a program that recognizes these digits?

What Machine Learning Does

- Class A

- Class B

- Want to be able to generalize the classification to unseen data.

http://ciml.info/ (Credit: Hal Daumé III)

Classify These

(Credit: Hal Daumé III)

Let’s See ...

- Bird vs non-bird
- Flies vs not-flies
- We need **bias** in order to be able to generalize to unseen data.

No Free-Lunch Theorems

Theorem 1

Let $\mathcal{F}$ be the set of all possible Boolean functions on n variables. Let $Acc_G(\mathcal{L})$ be the (generalization) accuracy of $\mathcal{L}$ on non-training examples. Then, for any consistent learner $\mathcal{L}$, it holds

$$\frac{1}{|\mathcal{F}|} \cdot \sum_{\mathcal{F}} Acc_G(\mathcal{L}) = 1/2\,.$$

Proof Sketch.

Let S be the set of training examples.
Let $f \in \mathcal{F}$ such that $Acc_G(f) = \frac{1}{2} + \delta$.
Then, $\exists f' \in \mathcal{F}$ such that $Acc_G(f') = \frac{1}{2} - \delta$.
To see why, note that we can have an $f' \in \mathcal{F}$ that satisfies:

$$\begin{cases} (\forall x \in S)(f'(x) = f(x)) \\ (\forall x \notin S)(f'(x) = \neg f(x)) \end{cases} \quad \square$$

No Free-Lunch Theorems

Theorem 2

Let $\mathcal{F}$ be the set of all possible Boolean functions on n variables. Let $Acc_G(\mathcal{L})$ be the (generalization) accuracy of $\mathcal{L}$ on non-training examples. Then, for any consistent learner $\mathcal{L}$, it holds

$$\frac{1}{|\mathcal{F}|} \cdot \sum_{\mathcal{F}} Acc_G(\mathcal{L}) = {}^1\!/\!_2\,.$$

Corollary 3

For any two learners $\mathcal{L}_1, \mathcal{L}_2$, if there exists a learning problem P such that $Acc_G(\mathcal{L}_1) > Acc_G(\mathcal{L}_2)$, then there exists another learning problem P' such that $Acc_G(\mathcal{L}_1) < Acc_G(\mathcal{L}_2)$.

- You can read more in [1].

References I

[1] David H. Wolpert. The Lack of A Priori Distinctions Between Learning Algorithms. *Neural Computation*, 8(7):1341–1390, 1996.