Android Epistemology

Renjie Yang

Capital Normal University
Department of Philosophy

May 2020

The Story of Epistemology So Far

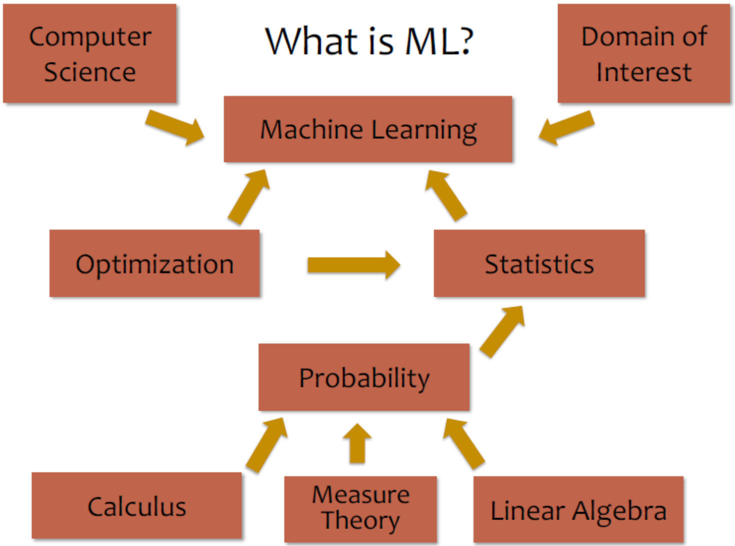
- Kant claimed to had explained why arithmetic and geometry can be known *a priori* with certainty.
- He also claimed that he provided a refutation of Hume's skepticism about induction. Although the metaphysical skepticism is still there.
- Kant's argument follows the idea of "transcendental argument": analyze the condition of possibility of human knowledge.
- But what about machines? Can we analyze the condition of possibility of machine knowledge?

Outline

Artificial Intelligence

The basic goal of AI is to develop intelligent machines.

- Perception
- Reasoning
- Control / Motion / Manipulation
- Planning
- Communication
- Creativity
- Learning

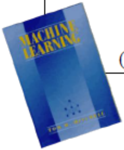

Computer Science
What is ML?
Domain of Interest
Machine Learning
Optimization
Statistics
Probability
Calculus
Measure Theory
Linear Algebra

Speech Recognition

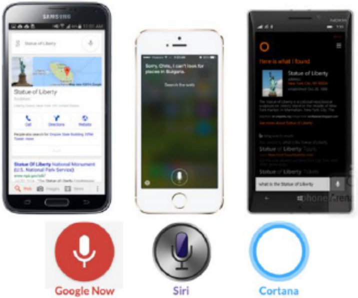
THEN

"…the SPHINX system (e.g. Lee 1989) learns speaker-specific strategies for recognizing the primitive sounds (phonemes) and words from the observed speech signal…neural network methods…hidden Markov models…"

(Mitchell, 1997)

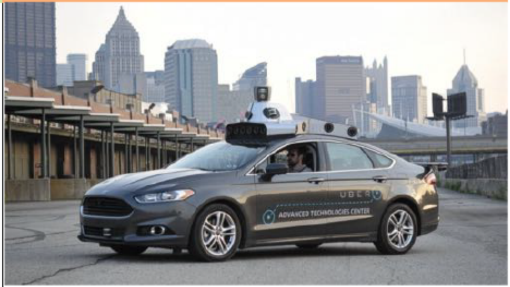


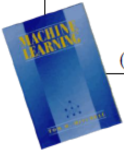
NOW



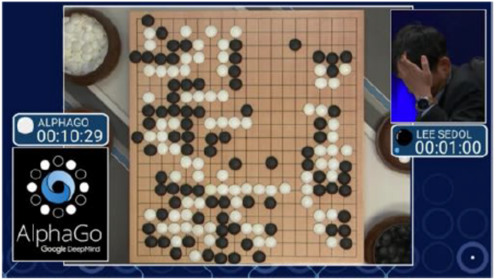
Source: https://www.stonetemple.com/great-knowledge-box-showdown/#VoiceStudyResults

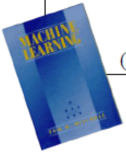
Robotics: Autonomous Vehicle

THEN	NOW
"…the ALVINN system (Pomerleau 1989) has used its learned strategies to drive unassisted at 70 miles per hour for 90 miles on public highways among other cars…" (Mitchell, 1997)	 https://www.geek.com/wp-content/uploads/2016/03/uber.jpg



Games / Reasoning

THEN	NOW
"…the world's top computer program for backgammon, TD-GAMMON (Tesauro, 1992, 1995), learned its strategy by playing over one million practice games against itself…" (Mitchell, 1997)	

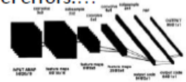


Computer Vision

THEN	NOW

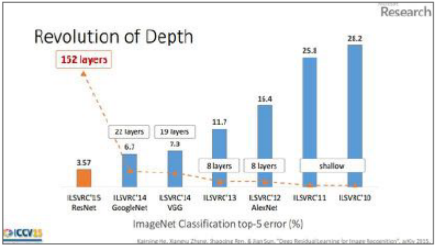
THEN

"…The recognizer is a convolution network that can be spatially replicated. From the network output, a hidden Markov model produces word scores. The entire system is globally trained to minimize word-level errors…."



(LeCun et al., 1995)

NOW



Images from https://blog.openai.com/generative-models/

Learning Theory

Sample Complexity Results

Definition 0.1. The **sample complexity** of a learning algorithm is the number of examples required to achieve arbitrarily small error (with respect to the optimal hypothesis) with high probability (i.e. close to 1).

Four Cases we care about...

	Realizable	Agnostic
Finite $\lvert\mathcal{H}\rvert$	$N \geq \frac{1}{\epsilon}\left[\log(\lvert\mathcal{H}\rvert) + \log(\frac{1}{\delta})\right]$ labeled examples are sufficient so that with probability $(1-\delta)$ all $h \in \mathcal{H}$ with $R(h) \geq \epsilon$ have $\hat{R}(h) > 0$.	$N \geq \frac{1}{2\epsilon^2}\left[\log(\lvert\mathcal{H}\rvert) + \log(\frac{2}{\delta})\right]$ labeled examples are sufficient so that with probability $(1-\delta)$ for all $h \in \mathcal{H}$ we have that $\lvert R(h) - \hat{R}(h)\rvert < \epsilon$.
Infinite $\lvert\mathcal{H}\rvert$	$N = O(\frac{1}{\epsilon}\left[\mathrm{VC}(\mathcal{H})\log(\frac{1}{\epsilon}) + \log(\frac{1}{\delta})\right])$ labeled examples are sufficient so that with probability $(1-\delta)$ all $h \in \mathcal{H}$ with $R(h) \geq \epsilon$ have $\hat{R}(h) > 0$.	$N = O(\frac{1}{\epsilon^2}\left[\mathrm{VC}(\mathcal{H}) + \log(\frac{1}{\delta})\right])$ labeled examples are sufficient so that with probability $(1-\delta)$ for all $h \in \mathcal{H}$ we have that $\lvert R(h) - \hat{R}(h)\rvert \leq \epsilon$.

Two Types of Error

① True Error (aka. expected risk) (aka. Generalization Error)

$R(h) = P_{x \sim p^*(x)}(c^*(x) \neq h(x))$ ← always unknown.

② Train Error (aka. empirical risk)

$\hat{R}(h) = P_{x \sim S}(c^*(x) \neq h(x))$ $\quad S = \{x^{(1)}, \dots, x^{(N)}\}$

$= \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}(c^*(x^{(i)}) \neq h(x^{(i)}))$ ← known, computable

$= \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}(y^{(i)} \neq h(x^{(i)}))$

PAC Learning

Q: Can we bound $R(h)$ in terms of $\hat{R}(h)$?
A: Yes!

PAC stands for Probably Approximately Correct

PAC learner yields hypothesis h, which is approximately correct $R(h) \approx 0$ with high probability $\Pr(R(h) \approx 0) \approx 1$

Def: PAC Criterion

$\Pr(\forall h, \lvert R(h) - \hat{R}(h)\rvert \leq \epsilon) \geq 1 - \delta$

1. How many examples do we need to learn?
2. How do we quantify our ability to generalize to unseen data?
3. Which algorithms are better suited to specific learning settings?

Machine Learning Big Picture

Learning Paradigms:
What data is available and when? What form of prediction?

- supervised learning
- unsupervised learning
- semi-supervised learning
- reinforcement learning
- active learning
- imitation learning
- domain adaptation
- online learning
- density estimation
- recommender systems
- feature learning
- manifold learning
- dimensionality reduction
- ensemble learning
- distant supervision
- hyperparameter optimization

Problem Formulation:
What is the structure of our output prediction?

boolean	Binary Classification
categorical	Multiclass Classification
ordinal	Ordinal Classification
real	Regression
ordering	Ranking
multiple discrete	Structured Prediction
multiple continuous	(e.g. dynamical systems)
both discrete & cont.	(e.g. mixed graphical models)

Application Areas
Key challenges?
NLP, Speech, Computer Vision, Robotics, Medicine, Search

Theoretical Foundations:
What principles guide learning?

- ❑ probabilistic
- ❑ information theoretic
- ❑ evolutionary search
- ❑ ML as optimization

Facets of Building ML Systems:
How to build systems that are robust, efficient, adaptive, effective?

1. *Data prep*
2. *Model selection*
3. *Training (optimization / search)*
4. *Hyperparameter tuning on validation data*
5. *(Blind) Assessment on test data*

Big Ideas in ML:
Which are the ideas driving development of the field?

- *inductive bias*
- *generalization / overfitting*
- *bias-variance decomposition*
- *generative vs. discriminative*
- *deep nets, graphical models*
- *PAC learning*
- *distant rewards*

Well-Posed Learning Problems

- Three components $< T, P, E >$:
  1. Task, T
  2. Performance measure, P
  3. Experience, E
- **Definition of learning**: A computer program learns if its performance at tasks in T, as measured by P, improves with experience E.

Outline

1. Machine Learning
2. **Demo I**
3. Agent-based Modeling
4. Demo II

Demo I

1. Regression
2. Recommendation System
3. Deep Learning

Outline

What is an Agent-Based Model?

- An agent is an autonomous individual element with properties and actions in a computer simulation
- Agent-Based Modeling (ABM) is the idea that the world can be modeled using agents, an environment, and a description of agent-agent and agent environment interactions

Complex Systems

- Complex systems are composed of many interacting parts
- Those parts are often connected in complex ways
- Complex Systems can be difficult to predict, control and manage, which in many ways is the goal of public policy
- Agent-Based Modeling and Complex Systems analysis is to provide a "flight simulator" rather than a perfect prediction

A Third Way of Doing Science(Axelrod, 1997)

- Two traditional ways of doing science
  1. Induction - inferring from particular data a general theory
  2. Deduction - reasoning from first principles to a general theory
- Third Way
  - Generative - using first principles to generate a particular set of data that can create a general theory

Outline

Demo II

1. The Boids Model
2. Fire Model
3. Tipping Model
4. El Feral Model

References

- Matt Gormley's Slides on "Machine Learning"
- Uri Wilensky, William Rand. *An Introduction to Agent-Based Modeling*. The MIT Press, 2015