Lecture 19: Conclusion

Roadmap

Summary of CS221

Next courses

History of AI

Food for thought

Paradigm

Modeling

Inference **Learning**

Course plan

Search problems
Markov decision processes
Adversarial games

Constraint satisfaction problems
Bayesian networks

Reflex **States** **Variables** **Logic**

"Low-level intelligence" "High-level intelligence"

Machine learning

Machine learning

Objective: loss minimization

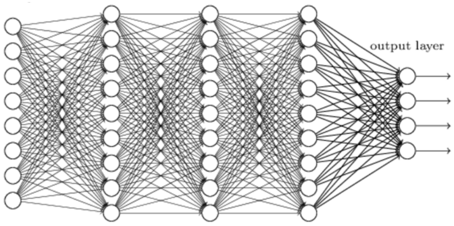
$$\min_{\mathbf{w}} \sum_{(x,y)\in\mathcal{D}_{\text{train}}} \text{Loss}(x, y, \mathbf{w})$$

Algorithm: stochastic gradient descent

$$\mathbf{w} \to \mathbf{w} - \eta_t \underbrace{\nabla \text{Loss}(x, y, \mathbf{w})}_{\text{prediction}-\text{target}}$$

Applies to wide range of models!

Reflex-based models

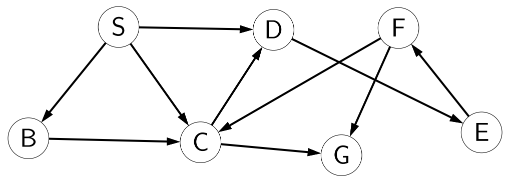


Models: linear models, neural networks, nearest neighbors

Inference: feedforward

Learning: SGD, alternating minimization

State-based models



Key idea: state

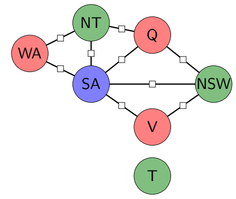
A **state** is a summary of all the past actions sufficient to choose future actions **optimally**.

Models: search problems, MDPs, games

Inference: UCS/A*, DP, value iteration, minimax

Learning: structured Perceptron, Q-learning, TD learning

Variable-based models



Key idea: factor graphs

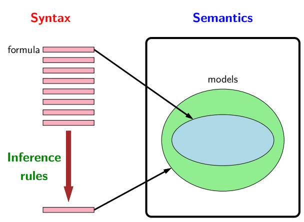
Graph structure captures conditional independence.

Models: CSPs, Bayesian networks

Inference: backtracking, forward-backward, beam search, Gibbs sampling

Learning: maximum likelihood (closed form, EM)

Logic-based models



Key idea: logic

Formulas enable more powerful models (infinite).

Models: propositional logic, first-order logic

Inference: model checking, modus ponens, resolution

Learning: ???

Tools

- CS221 provides a set of tools

- Start with the problem, and figure out what tool to use
- Keep it simple!

Roadmap

Summary of CS221

Next courses

History of AI

Food for thought

Other AI-related courses

http://ai.stanford.edu/courses/

Foundations:

- CS228: Probabilistic Graphical Models
- CS229: Machine Learning
- CS229T: Statistical Learning Theory
- CS230: Deep Learning
- CS334A: Convex Optimization
- CS238: Decision Making Under Uncertainty
- CS257: Logic and Artificial Intelligence
- CS246: Mining Massive Data Sets

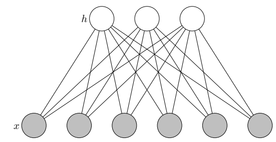
Other AI-related courses

http://ai.stanford.edu/courses/

Applications:

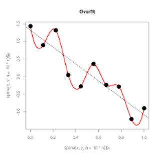
- CS224N: Natural Language Processing (with Deep Learning)
- CS224U: Natural Language Understanding
- CS231A: From 3D Reconstruction to Recognition
- CS231N: Convolutional Neural Networks for Visual Recognition
- CS223A: Introduction to Robotics
- CS237A-B: Robot Autonomy
- CS227B: General Game Playing

Probabilistic graphical models (CS228)



- Forward-backward, variable elimination ⇒ belief propagation, variational inference
- Gibbs sampling ⇒ Markov Chain Monte Carlo (MCMC)
- Learning the structure

Machine learning (CS229)



- Discrete ⇒ continuous
- Linear models ⇒ kernel methods, decision trees
- Boosting, bagging, feature selection
- K-means ⇒ mixture of Gaussians, PCA, ICA

Statistical learning theory (CS229T)

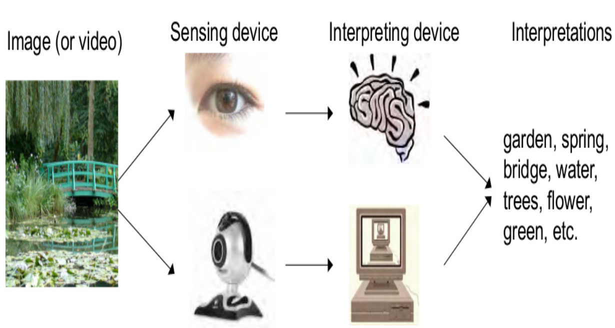
Question: what are the mathematical principles behind learning?

Uniform convergence: with probability at least 0.95, your algorithm will return a predictor $h \in \mathcal{H}$ such that

$$\text{TestError}(h) \le \text{TrainError}(h) + \sqrt{\frac{\text{Complexity}(\mathcal{H})}{n}}$$

[figure credit: Fei-Fei Li]

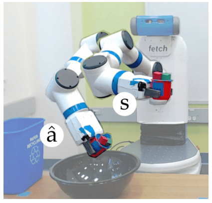
Vision (CS231A, CS231N)



- Challenges: variation in viewpoint, illumination, intra-class variation
- Tasks: object recognition/detection/segmentation, pose estimation, 3D reconstruction, image captioning, visual question answering, activity recognition

Robotics (CS223A, CS225A)

- Tasks: manipulation, grasping, navigation



- Applications: self-driving cars, medical robotics
- Physical models: kinematics, control

Robotics (CS237A, CS237B)

- Tasks: interaction, robot learning, autonomy

- Applications: mobile manipulation
- Term: Winter 2020, (Marco Pavone, Jeannette Bohg, Dorsa Sadigh)

Language (CS224N, CS224U)

- Designed by humans for communication
- World: continuous, words: discrete, meanings: continuous
- Properties: compositionality, grounding

- Tasks: syntactic parsing, semantic parsing, information extraction, coreference resolution, machine translation, question answering, summarization, dialogue

Cognitive science

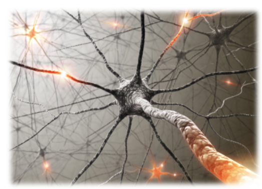
Question: How does the human mind work?

- Cognitive science and AI grew up together
- Humans can learn from few examples on many tasks

Computation and cognitive science (PSYCH204, CS428):

- Cognition as Bayesian modeling — probabilistic program [Tenenbaum, Goodman, Griffiths]

Neuroscience



- Neuroscience: hardware; cognitive science: software
- Artificial neural network as computational models of the brain
- Modern neural networks (GPUs + backpropagation) not biologically plausible
- Analogy: birds versus airplanes; what are principles of intelligence?

Roadmap

Summary of CS221

Next courses

History of AI

Food for thought

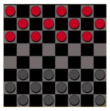
Birth of AI

1956: Workshop at Dartmouth College; attendees: John McCarthy, Marvin Minsky, Claude Shannon, etc.

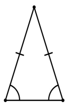
Aim for **general principles**:

Every aspect of learning or any other feature of intelligence can be so precisely described that a machine can be made to simulate it.

Birth of AI, early successes



Checkers (1952): Samuel's program learned weights and played at strong amateur level



Problem solving (1955): Newell & Simon's Logic Theorist: prove theorems in Principia Mathematica using search + heuristics; later, General Problem Solver (GPS)

Overwhelming optimism...

Machines will be capable, within twenty years, of doing any work a man can do. —Herbert Simon

Within 10 years the problems of artificial intelligence will be substantially solved. —Marvin Minsky

I visualize a time when we will be to robots what dogs are to humans, and I'm rooting for the machines. —Claude Shannon

...underwhelming results

Example: machine translation

The spirit is willing but the flesh is weak.

↓

(Russian)

↓

The vodka is good but the meat is rotten.

1966: ALPAC report cut off government funding for MT

AI is overhyped...

We tend to overestimate the effect of a technology in a short run and underestimate the effect in a long run. —Roy Amara (1925-2007)

Implications of early era

Problems:

- **Limited computation**: search space grew exponentially, outpacing hardware ($100! \approx 10^{157} > 10^{80}$)
- **Limited information**: complexity of AI problems (number of words, objects, concepts in the world)

Contributions:

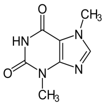
- Lisp, garbage collection, time-sharing (John McCarthy)
- Key paradigm: separate **modeling** (declarative) and **inference** (procedural)

Knowledge-based systems (70-80s)

Expert systems: elicit specific domain knowledge from experts in form of rules:

```
if [premises] then [conclusion]
```

Knowledge-based systems (70-80s)



DENDRAL: infer molecular structure from mass spectrometry

MYCIN: diagnose blood infections, recommend antibiotics

XCON: convert customer orders into parts specification; save DEC $40 million a year by 1986

Knowledge-based systems

Contributions:

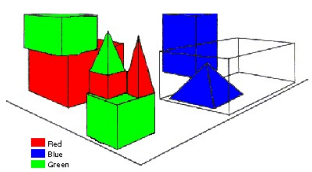
- First **real application** that impacted industry
- Knowledge helped curb the exponential growth

Problems:

- Knowledge is not deterministic rules, need to model **uncertainty**
- Requires considerable **manual effort** to create rules, hard to maintain

SHRDLU [Winograd 1971]

Person: Pick up a big red block.
Computer: OK.
Person: Grasp the pyramid.
Computer: I don't understand which pyramid you mean.
Person (changing their mind): Find a block which is taller than the one you are holding and put it into the box.
Computer: By "it", I assume you mean the block which is taller than the one I am holding.
Computer: OK.
Person: What does the box contain?
Computer: The blue pyramid and the blue block.
Person: What is the pyramid supported by?
Computer: The box.



The Complexity Barrier

A number of people have suggested to me that large programs like the SHRDLU program for understanding natural language represent a kind of dead end in AI programming. Complex interactions between its components give the program much of its power, but at the same time they present a formidable obstacle to understanding and extending it. In order to grasp any part, it is necessary to understand how it fits with other parts, presents a dense mass, with no easy footholds. Even having written the program, I find it near the limit of what I can keep in mind at once.

— Terry Winograd (1972)

Modern AI (90s-present)

- **Probability**: Pearl (1988) promote Bayesian networks in AI to **model uncertainty** (based on Bayes rule from 1700s)

  model → predictions

- **Machine learning**: Vapnik (1995) invented support vector machines to **tune parameters** (based on statistical models in early 1900s)

  data → **model**

A melting pot

- Bayes rule (Bayes, 1763) from **probability**
- Least squares regression (Gauss, 1795) from **astronomy**
- First-order logic (Frege, 1893) from **logic**
- Maximum likelihood (Fisher, 1922) from **statistics**
- Artificial neural networks (McCulloch/Pitts, 1943) from **neuroscience**
- Minimax games (von Neumann, 1944) from **economics**
- Stochastic gradient descent (Robbins/Monro, 1951) from **optimization**
- Uniform cost search (Dijkstra, 1956) from **algorithms**
- Value iteration (Bellman, 1957) from **control theory**

Roadmap

Summary of CS221

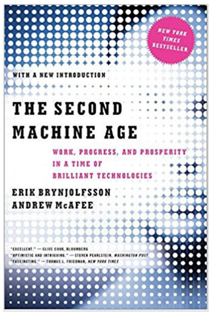
Next courses

History of AI

Food for thought

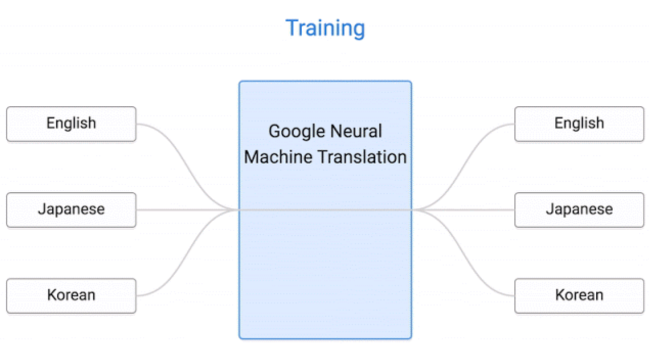
Outlook

AI is everywhere: consumer services, advertising, transportation, manufacturing, etc.



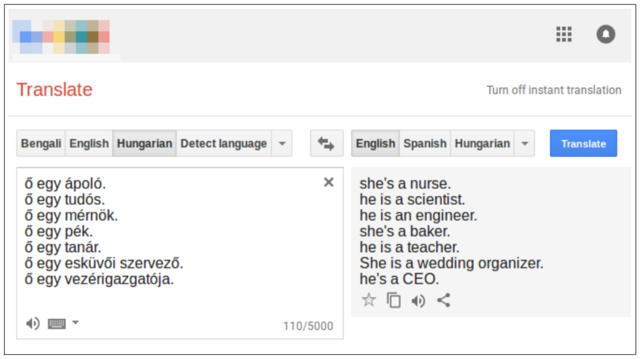
AI being used to make decisions for: education, credit, employment, advertising, healthcare and policing

Google Machine Translation (2016)

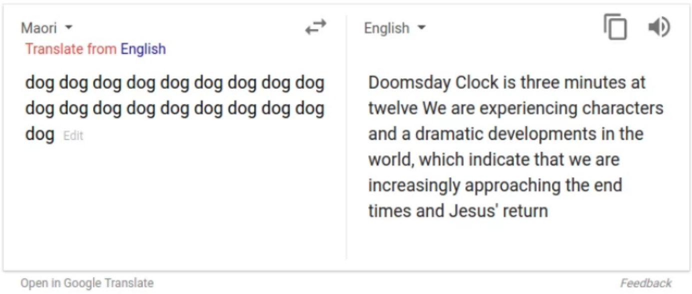


[Prates+ 2018]

Biases

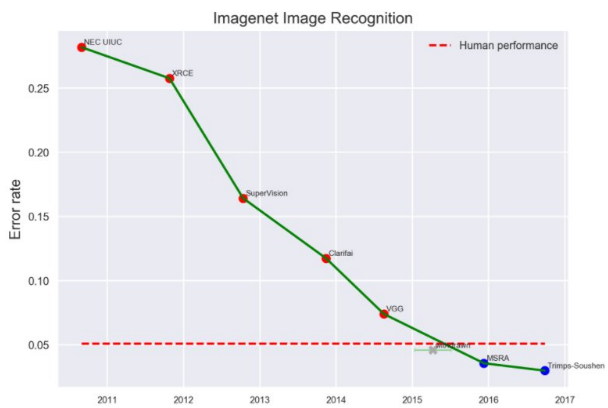


Craziness



[figure credit: EFF]

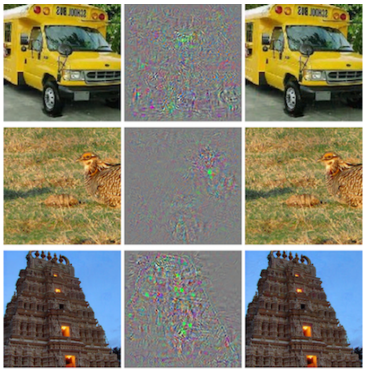
Image classification



[Szegedy et al., 2013; Goodfellow et al., 2014]

Adversaries

AlexNet predicts correctly on the left

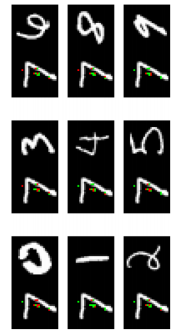


AlexNet predicts **ostrich** on the right

[Shamir et al., 2019]

Adversaries

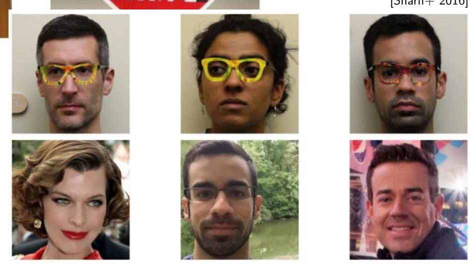
A Simple Explanation for Existence of Adversarial Examples with Small Hamming Distance



Security

[Evtimov+ 2017]

[Sharif+ 2016]



[with Robin Jia; EMNLP 2017]

Reading comprehension

Individual Huguenots settled at the Cape of Good Hope from as early as 1671 with the arrival of Francois Villion (Viljoen). The first Huguenot to arrive at the Cape of Good Hope was however Maria de la Queillerie, wife of commander Jan van Riebeeck (and daughter of a Walloon church minister), who arrived on 6 April 1652 to establish a settlement at what is today Cape Town. The couple left for the Far East ten years later. On 31 December 1687 the first organised group of Huguenots set sail from the Netherlands to the Dutch East India Company post at the Cape of Good Hope. The largest portion of the Huguenots to settle in the Cape arrived between 1688 and 1689 in seven ships as part of the organised migration, but quite a few arrived as late as 1700; thereafter, the numbers declined and only small groups arrived at a time. The number of old Acadian colonists declined after the year 1675.

The number of new Huguenot colonists declined after what year?

↓

BERT [Google]

↓

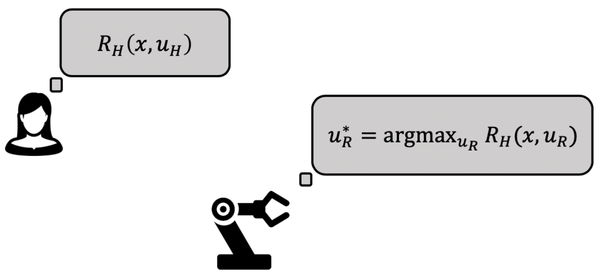
1675

Optimizing for clicks

Is this a good objective function for society?

How to model human objectives?

Write a reward function:



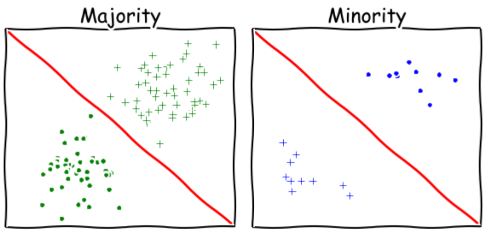
Is this a good objective function for the human?

Generating fake content

Can build it $\neq$ should build it?

Figure from Moritz Hardt

Fairness

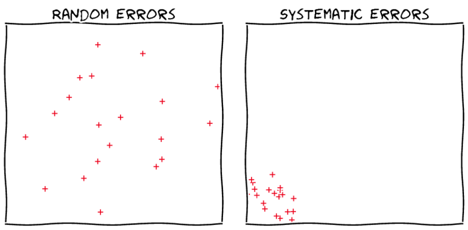


- Most ML training objectives will produce model accurate for majority class, at the expense of the minority one.

Figure from Moritz Hardt

Fairness

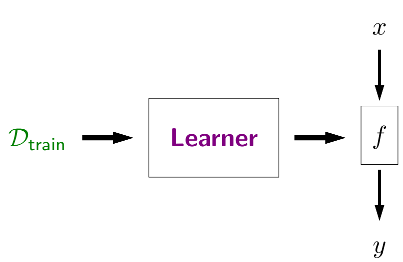
Two classifiers with 5% error:



Fairness in criminal risk assessment

- Northpointe: COMPAS predicts criminal risk score (1-10)
- ProPublica: given that an individual did not reoffend, blacks 2x likely to be (wrongly) classified 5 or above
- Northpointe: given a risk score of 7, 60% of whites reoffended, 60% of blacks reoffended

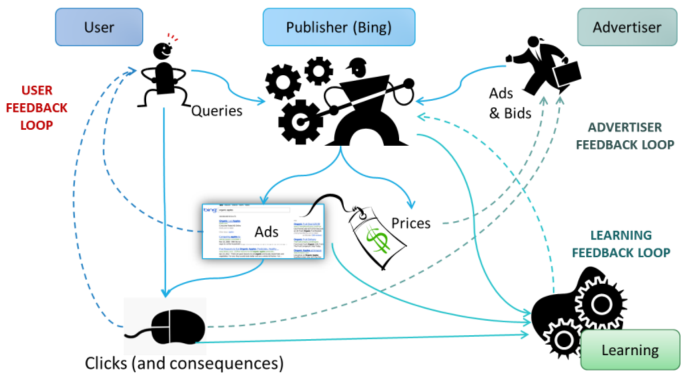
Are algorithms neutral?



By design: picks up patterns in training data, including biases

[Leon Bottou]

Feedback loops



Privacy

- Not reveal sensitive information (income, health, communication)
- Compute average statistics (how many people have cancer?)

yes no yes no no

no no yes no yes

[Warner, 1965]

Privacy: Randomized response

Do you have a sibling?

Method:

- Flip two coins.
- If both heads: answer yes/no randomly
- Otherwise: answer yes/no truthfully

Analysis:

true-prob $= \frac{4}{3} \times$ (observed-prob $- \frac{1}{8}$)

Causality

Goal: figure out the effect of a treatment on survival

Data:

For untreated patients, 80% survive
For treated patients, 30% survive

Does the treatment help?

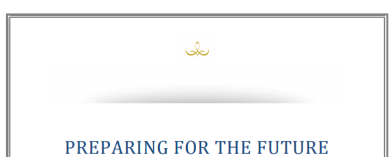
Who knows? Sick people are more likely to undergo treatment...

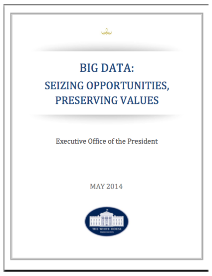
[Mykel Kochdorfer]

Interpretability versus accuracy

- For air-traffic control, threshold level of safety: probability 10^{-9} for a catastrophic failure (e.g., collision) per flight hour
- Move from human designed rules to a numeric Q-value table?

yes





..big data analytics have the potential to eclipse longstanding civil rights protections in how personal information is used in housing, credit, employment, health, education and the marketplace. Americans relationship with data should expand, not diminish, their opportunities..

There is always a human ultimately responsible for decisions made or informed by an algorithm. "The algorithm did it" is not an acceptable excuse if algorithmic systems make mistakes or have undesired consequences, including from machine-learning processes

Societal and industrial impact

Enormous potential for positive impact, use responsibly!

Search problems
Markov decision processes
Adversarial games

Constraint satisfaction problems
Bayesian networks

Reflex — **States** — **Variables** — **Logic**

"Low-level intelligence" → "High-level intelligence"

Machine learning

Please fill out course evaluations on Axess.

Thanks for an exciting quarter!