Machine Learning 10-601/10-301

Matt Gormley and Tom M. Mitchell
Machine Learning Department
Carnegie Mellon University

February 1, 2021

Today:

- Machine learning is awesome
- What ML is
- What this course will cover
- Course logistics/policies

Suggested Reading:

- "Machine Learning: Trends" perspectives and prospects" Jordan & Mitchell 2015 http://www.cs.cmu.edu/~tom/pubs/Science-ML-2015.pdf

Course Webpage: http://mlcourse.org

Machine Learning:

Study of algorithms that

- improve their performance P
- at some task T
- with experience E

well-defined learning task: <P,T,E>

Example: Learning to play Go

- Task
  - Learn the function
    F: board → move
- Performance
  - Maximize number of games won
- Experience
  - supervised training examples from experts:
    <board, expert_move>

Example: Learning to drive

[www.argo.ai]

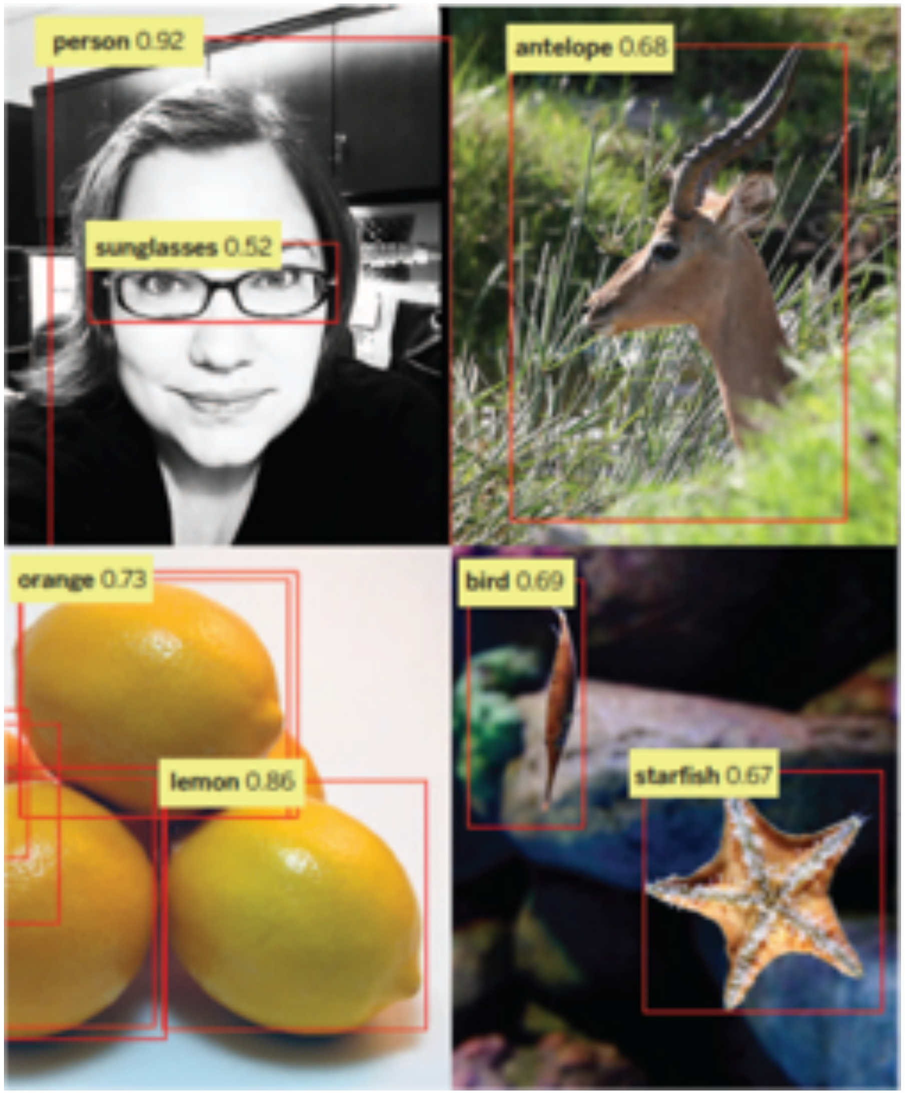
- Task
  - Learn function
    F: image(t) → steering(t), accelerator(t), brake(t)

- Performance
  - Minimize
    [c_1 drive_time_home + c_2 vehicles_hit + c_3 pedestrians_hit]

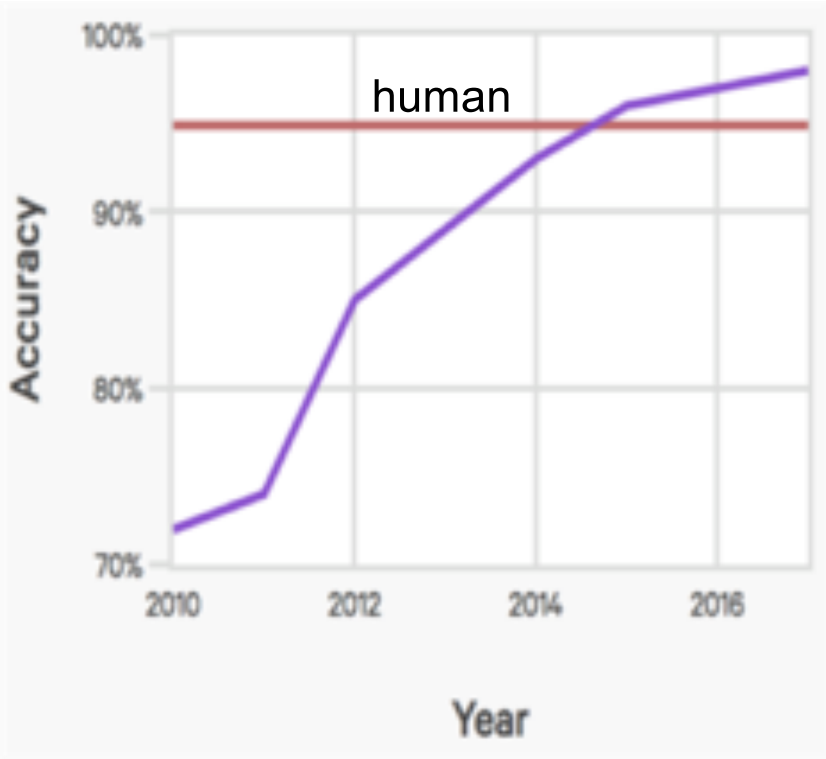
- Experience
  - Self driving experiments in driving simulator

What has machine learning accomplished lately?

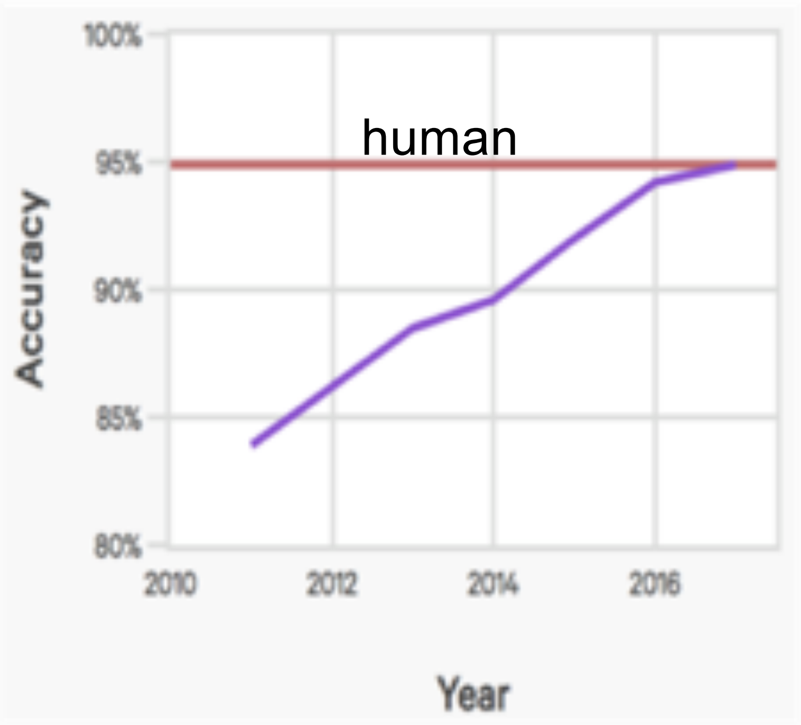
Computer Vision



Imagenet Visual Recognition Challenge



Speech Recognition



Robots

Factories, Land, Air, Sea, Mines, Homes

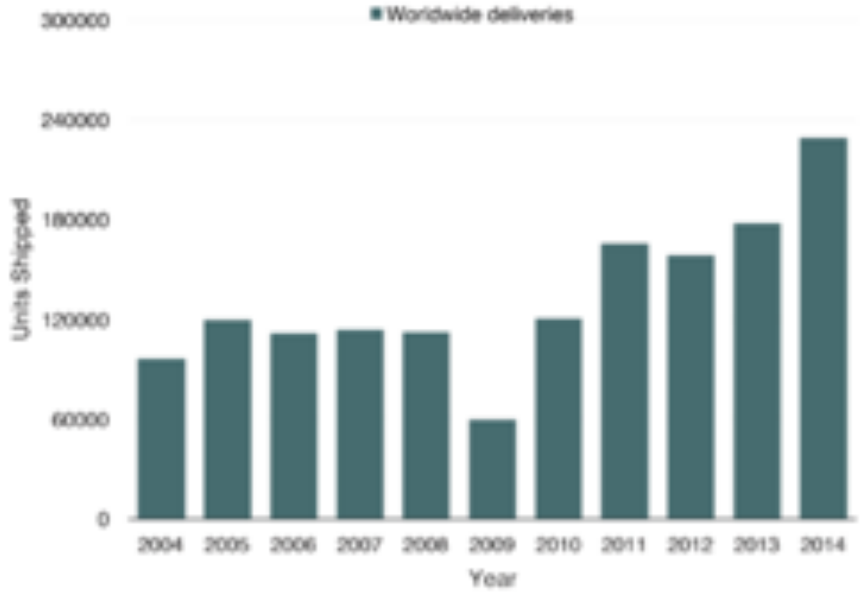


FIGURE 2.4 Worldwide shipping of robots over time. SOURCE: International Federation of Robotics, 2015.

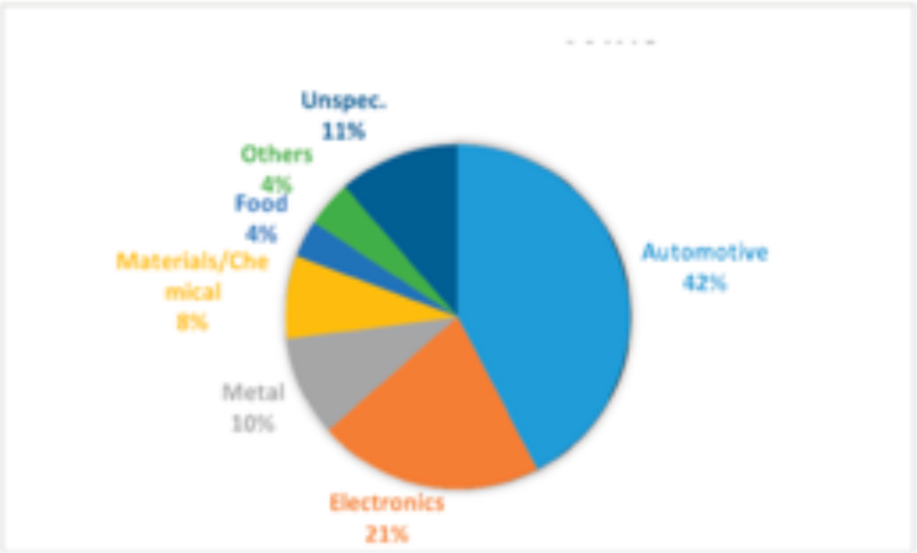


FIGURE 2.5 Robot application areas in 2015. SOURCE: Data from International Federation of Robotics, 2015.

Games and reasoning

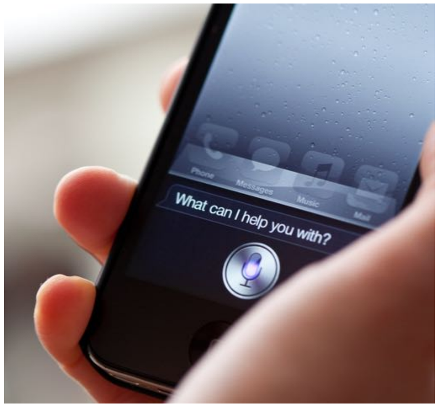
Go

Chess

Jeopardy

Poker

The key: Machine Learning



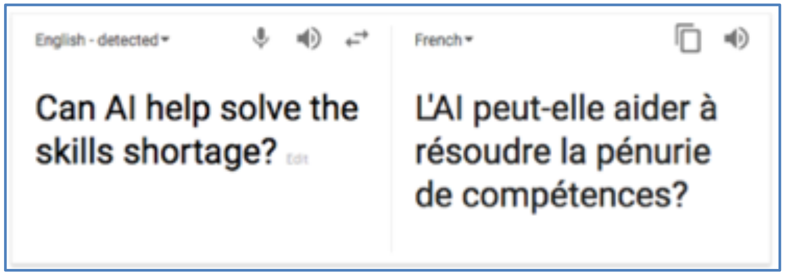
conversational agents

medical diagnosis

fraud detection

Many algorithms:

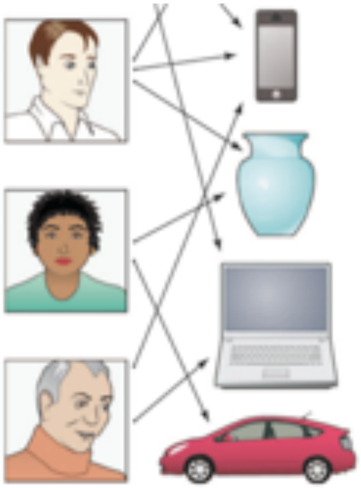
- Decision trees
- Deep neural networks
- Bayesian networks
- Hidden Markov models
- Gaussian mixture model
- Expectation maximization
- ....



Can AI help solve the skills shortage?

L'AI peut-elle aider à résoudre la pénurie de compétences?

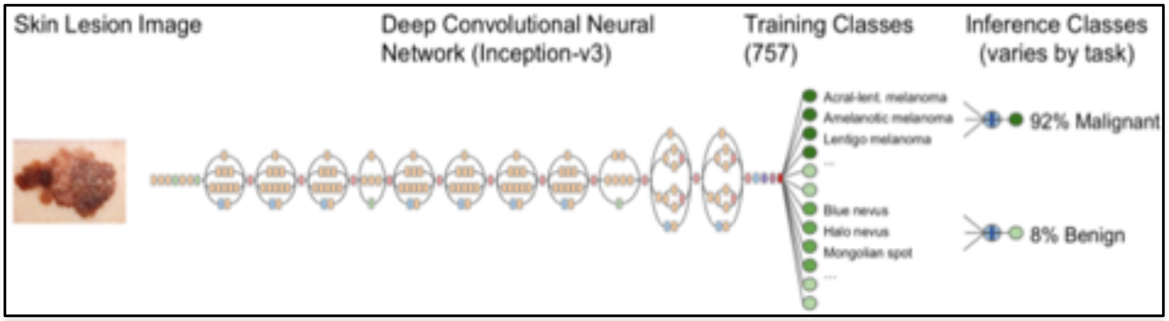
translation



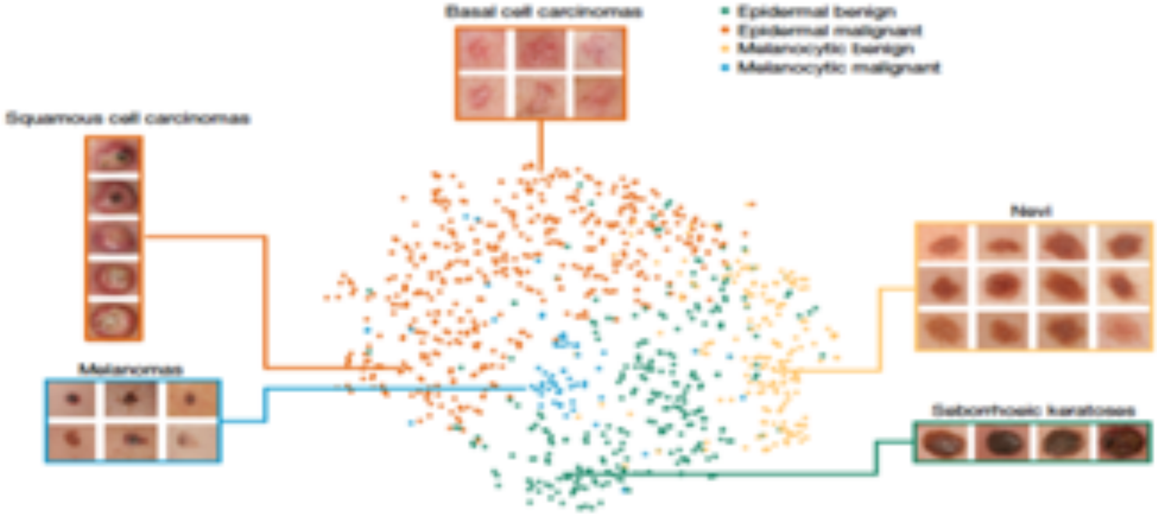
recommendations

Skin Cancer Diagnosis

[Esteva et al., *Nature* 2017]



Trained on 129,450 skin images
plus 1.4 million standard photographs
Deep net Inception v3 architecture
Outperforms doctors



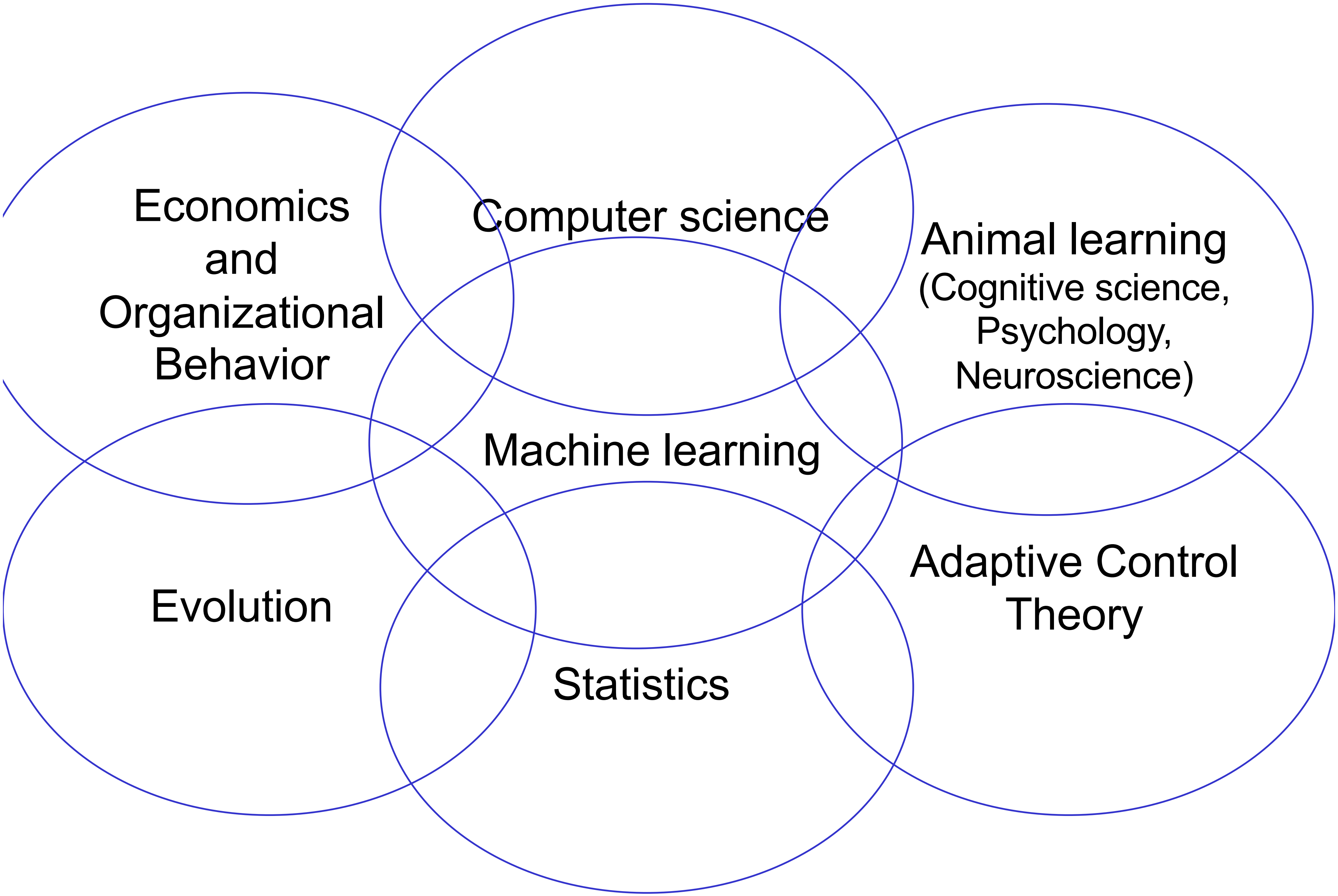
tsne visualization of final hidden layer

Predict Cardiovascular Risk from Retinal Photographs

[Poplin et al., *Nature Biomed Eng.* 2018]

Trained deep net on 284,335 retinal images
New approach to detecting risk factors and biometrics

	Accuracy
Age	within 3.26 years on average
Smoker?	71%
Systolic blood pressure	within 11 mmHg on average
Gender	97%
Major cardiac event within past 5 years?	70%


Economics and Organizational Behavior
Computer science
Animal learning (Cognitive science, Psychology, Neuroscience)
Machine learning
Evolution
Statistics
Adaptive Control Theory

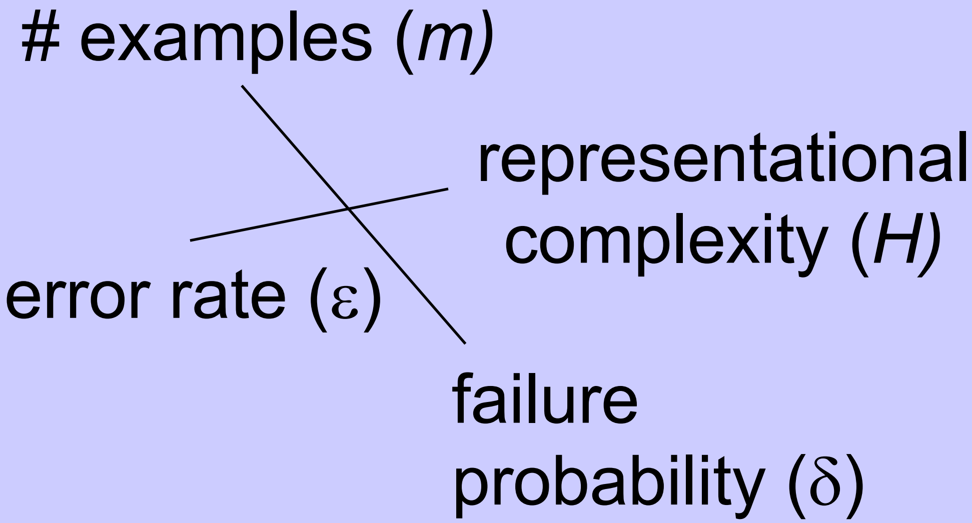
What questions should a “theory” of machine learning answer?

is it even possible?

Machine Learning - Theory

PAC Learning Theory

(supervised concept learning)



# examples (*m*)

representational complexity (*H*)

error rate (ε)

failure probability (δ)

$$m \geq \frac{1}{\epsilon}(\ln|H| + \ln(1/\delta))$$

Other theories for

- Reinforcement skill learning
- Semi-supervised learning
- Active student querying
- …

… also relating:

- # of mistakes during learning
- learner’s query strategy
- convergence rate
- computational demands
- asymptotic performance
- bias, variance, Bayesian priors
- VC dimension

Social impacts of Machine Learning

- ML has produced better, evidence-based, decision making in many domains
  - Medical diagnosis, Credit card fraud detection, Online tutoring, Anticipating equipment failures, …
- and some more controversial domains
  - Jail sentencing, targeted marketing, …

- Raises new issues
  - Explainability
  - Bias
  - If big data is key to successful ML, who controls access to the data?
  - Privacy
  - …

Privacy enhanced machine learning

- Issue: Sometimes need personal data to train needed classifiers
  - E.g., Training diagnosis systems from electronic health records
  - Wish to train on data from all hospitals, but hospitals don’t share personal data

Privacy enhanced machine learning

- Issue: Sometimes need personal data to train needed classifiers
  - E.g., Training diagnosis systems from electronic health records
  - Wish to train on data from all hospitals, but hospitals don't share personal data

- Idea 1: homomorphic encryption
  - Example: Compute average GPA of students in our class (preserving privacy)

  - Encrypt data, but perform math operations without decrypting
  - E.g., given numbers n1, n2, we can calculate sum n1+n2 even if encrypted
  - Encrypt(n1+n2) = Esum(encrypt(n1), encrypt(n2))
  - But sadly, not every function can be done on encrypted data…

Privacy enhanced machine learning

- Issue: Sometimes need personal data to train needed classifiers
  - E.g., Training diagnosis systems from electronic health records
  - Wish to train on data from all hospitals, but hospitals don't share personal data

- Idea 1: homomorphic encryption
  - Example: Compute average GPA of students in our class (preserving privacy)

  - Encrypt data, but perform math operations without decrypting
  - E.g., given numbers n1, n2, we can calculate sum n1+n2 even if encrypted
  - Encrypt(n1+n2) = Esum(encrypt(n1), encrypt(n2))
  - But sadly, not every function can be done on encrypted data…

- Idea 2: differential privacy
  - Add noise to data D or algorithm A, so that an observer of A's output probably cannot determine whether data entry d was even included in D
  - Theoretical results show how to bound this probability
  - Trades off privacy for accuracy

We’ll cover in this course

Algorithms:

- Decision trees
- Bayes classifiers
- Regression
- Deep neural networks
- Convolutional nets
- Transformer networks
- Graphical models
- Expectation maximization
- PCA, Matrix factorization
- Reinforcement learning

Concepts:

- Statistical estimation
- Overfitting
- Cross-validation
- Representation learning
- Probabilistic models
- Probably approximately correct learning
- VC dimension
- Role of unlabeled data
- Transfer learning
- Optimization

What is Machine Learning?

The goal of this course is to provide you with a toolbox:

Machine Learning

Statistics

Probability

Computer Science

Optimization

To solve all the problems above and more

[slide courtesy of Matt Gormley]

Over to Matt…

DEFINING LEARNING PROBLEMS

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*.

Definition from (Mitchell, 1997)

Example Learning Problems

Learning to beat the masters at **chess**

1. Task, *T*:

2. Performance measure, *P*:

3. Experience, *E*:

Problem Formulation

- Often, the same task can be formulated in more than one way:
- Ex: Loan applications
  - creditworthiness/score (regression)
  - probability of default (density estimation)
  - loan decision (classification)

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)

Well-posed Learning Problems

In-Class Exercise

1. Select a **task**, T
2. Identify **performance measure**, P
3. Identify **experience**, E
4. Report ideas back to rest of class

Example Tasks

- Identify objects in an image
- Translate from one human language to another
- Recognize speech
- Assess risk (e.g. in loan application)
- Make decisions (e.g. in loan application)
- Assess potential (e.g. in admission decisions)
- Categorize a complex situation (e.g. medical diagnosis)
- Predict outcome (e.g. medical prognosis, stock prices, inflation, temperature)
- Predict events (default on loans, quitting school, war)
- Plan ahead under perfect knowledge (chess)
- Plan ahead under partial knowledge (Poker, Bridge)

Examples from Roni Rosenfeld

SYLLABUS HIGHLIGHTS

Syllabus Highlights

The syllabus is located on the course webpage:

http://www.cs.cmu.edu/~mgormley/courses/10601

or

http://mlcourse.org

The **course policies** are **required** reading.

Syllabus Highlights

- **Grading**: 50% homework, 15% exam 1, 15% exam 2, 15% final exam, 5% participation
- **Midterm Exam 1**: evening exam
- **Midterm Exam 2**: evening exam
- **Final Exam**: final exam week, date TBD by registrar
- **Homework**: ~3 written and ~6 written + programming
  - 6 grace days for homework assignments
  - Late submissions: 80% day 1, 60% day 2, 40% day 3, 20% day 4
  - No submissions accepted after 4 days w/o extension
  - Extension requests: see syllabus
- **Recitations**: Fridays, same time/place as lecture (optional, interactive sessions)
- **Readings**: required, online PDFs, recommended for after lecture
- **Technologies**:
  - Piazza (discussion),
  - Gradescope (homework),
  - Google Forms (polls),
  - Gather.Town (office hours),
  - Zoom (livestream),
  - Panopto (video recordings)
- **Academic Integrity**:
  - Collaboration encouraged, but must be documented
  - Solutions must always be written independently
  - No re-use of found code / past assignments
  - Severe penalties (i.e.. failure)
- **Office Hours**: posted on Google Calendar on “People” page

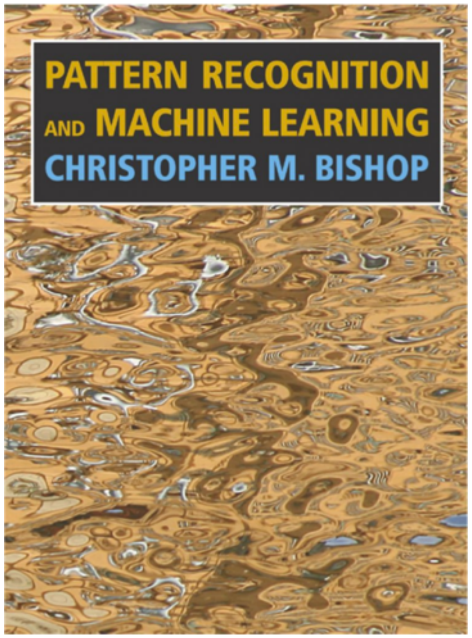
Lectures

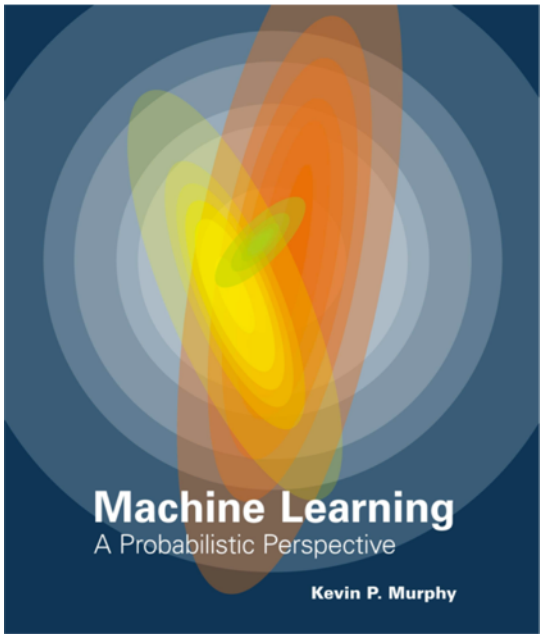
- You should ask lots of questions
  - Interrupting (by raising a hand, turning on your video, and waiting to be called on) to ask your question is strongly encouraged
  - Use the chat to ask questions in real time (TAs will be monitoring the chat and will either answer or interrupt the instructor)
  - Asking questions later on Piazza is also great
- When I ask a question…
  - I want you to answer
  - Even if you don’t answer, think it through as though I’m about to call on you
- Interaction improves learning (both in-class and at my office hours)

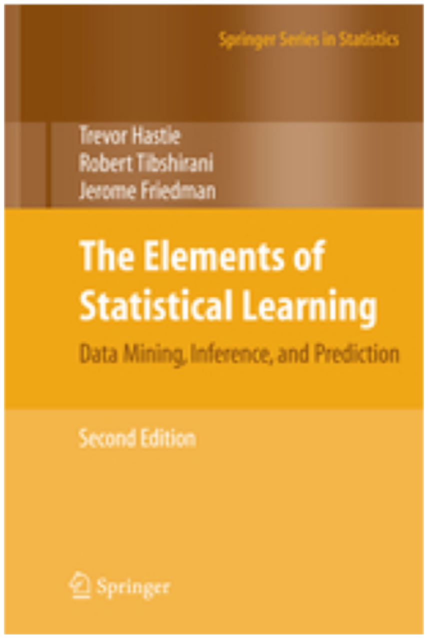
Textbooks

You are not *required* to read a textbook, but it will help immensely!

MACHINE LEARNING
TOM M. MITCHELL






Springer Series in Statistics
Trevor Hastie
Robert Tibshirani
Jerome Friedman
The Elements of Statistical Learning
Data Mining, Inference, and Prediction
Second Edition
Springer

PREREQUISITES

Prerequisites

What they are:

- Significant programming experience (15-122)
  - Written programs of 100s of lines of code
  - Comfortable learning a new language
- Probability and statistics (36-217, 36-225, etc.)
- Mathematical maturity: discrete mathematics (21-127, 15-151), linear algebra, and calculus

Use Probability

Oh, the Places You'll Use Probability!

Supervised Classification

- Naïve Bayes

$$p(y|x_1, x_2, \ldots, x_n) = \frac{1}{Z} p(y) \prod_{i=1}^{n} p(x_i|y)$$

- Logistic regression

$$\begin{aligned} P(Y = y|X = x; \boldsymbol{\theta}) &= p(y|x; \boldsymbol{\theta}) \\ &= \frac{\exp(\boldsymbol{\theta}_y \cdot \mathbf{f}(x))}{\sum_{y'} \exp(\boldsymbol{\theta}_{y'} \cdot \mathbf{f}(x))} \end{aligned}$$

Note: This is just motivation – we'll cover these topics later!

Oh, the Places You'll Use Probability!

ML Theory

(Example: Sample Complexity)

- Goal: h has small error over D.

True error: $err_D(h) = \Pr_{x \sim D}(h(x) \neq c^*(x))$

How often $h(x) \neq c^*(x)$ over future instances drawn at random from D

- But, can only measure:

Training error: $err_S(h) = \frac{1}{m}\sum_i I\big(h(x_i) \neq c^*(x_i)\big)$

How often $h(x) \neq c^*(x)$ over training instances

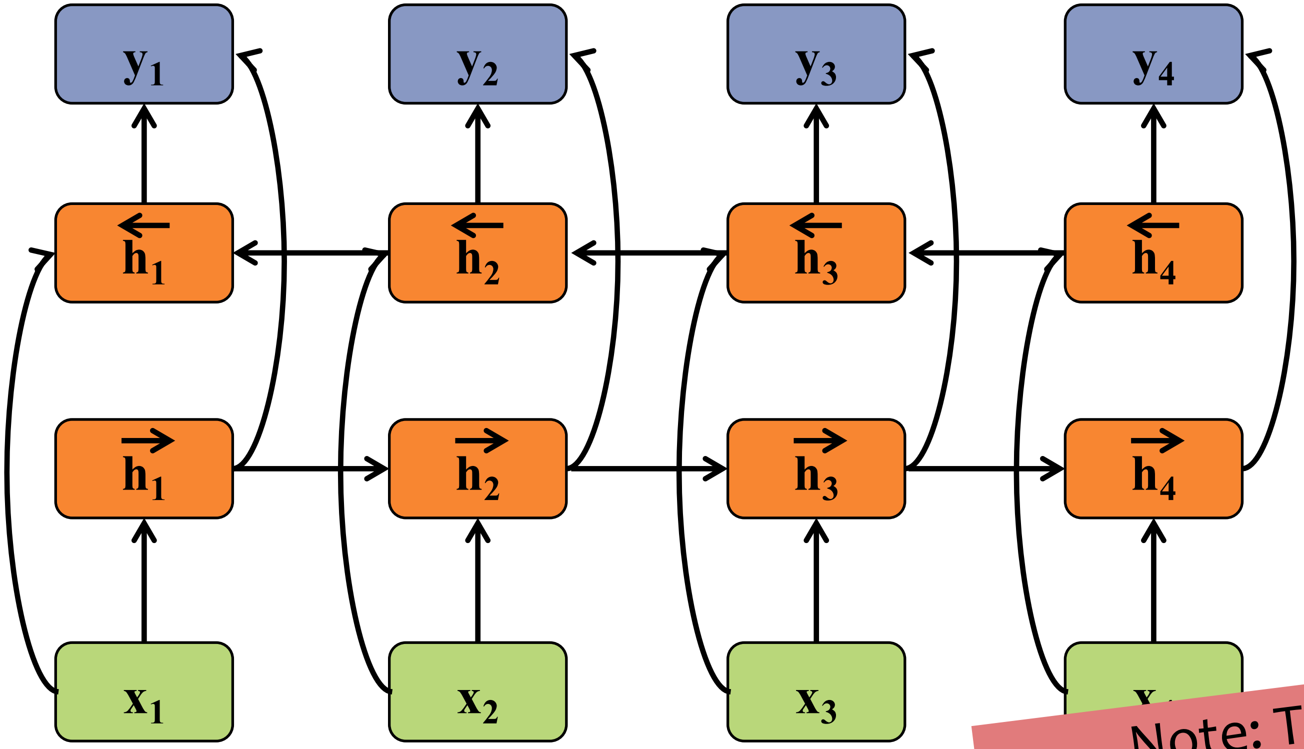
Sample complexity: bound $err_D(h)$ **in terms of** $err_S(h)$

Note: This is just motivation – we'll cover these topics later!

Oh, the Places You'll Use Probability!

Deep Learning

(Example: Deep Bi-directional RNN)

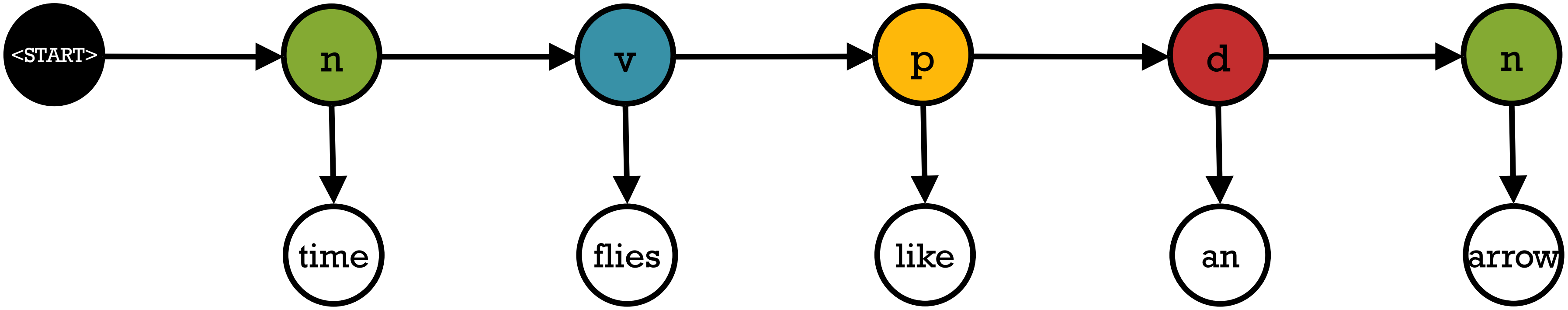

$\mathbf{y}_1$
$\mathbf{y}_2$
$\mathbf{y}_3$
$\mathbf{y}_4$
$\overleftarrow{\mathbf{h}}_1$
$\overleftarrow{\mathbf{h}}_2$
$\overleftarrow{\mathbf{h}}_3$
$\overleftarrow{\mathbf{h}}_4$
$\overrightarrow{\mathbf{h}}_1$
$\overrightarrow{\mathbf{h}}_2$
$\overrightarrow{\mathbf{h}}_3$
$\overrightarrow{\mathbf{h}}_4$
$\mathbf{x}_1$
$\mathbf{x}_2$
$\mathbf{x}_3$

Note: This is just motivation – we'll cover these topics later!

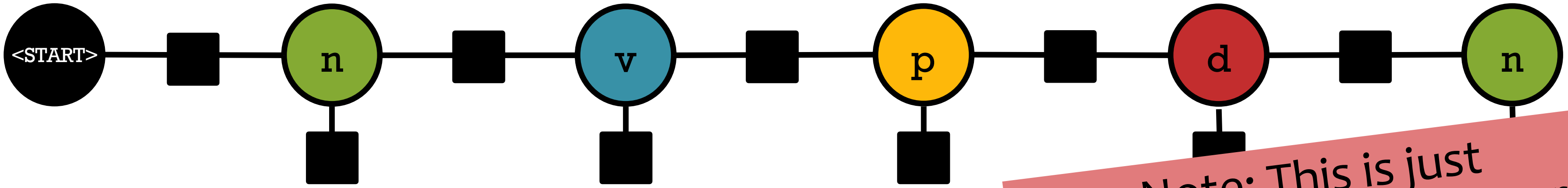
Oh, the Places You'll Use Probability!

Graphical Models

- Hidden Markov Model (HMM)



- Conditional Random Field (CRF)



Note: This is just motivation – we'll cover these topics later!

Prerequisites

What if I'm not sure whether I meet them?

- Don't worry: we're not sure either
- However, we've designed a way to assess your background knowledge so that you know what to study!

  (see instructions of written portion of HW1)

Reminders

- **Homework 1: Background**
  - **Out: Wed, Feb 03 (2nd lecture)**
  - **Due: Wed, Feb 10 at 11:59pm**
  - Two parts:
    1. written part to Gradescope
    2. programming part to Gradescope
  - unique policy for this assignment:
    1. **two submissions** for written (see writeup for details)
    2. **unlimited submissions** for programming (i.e. keep submitting until you get 100%)

Learning Objectives

You should be able to…

1. Formulate a well-posed learning problem for a real-world task by identifying the task, performance measure, and training experience
2. Describe common learning paradigms in terms of the type of data available, when it's available, the form of prediction, and the structure of the output prediction
3. Implement Decision Tree training and prediction (w/simple scoring function)
4. Explain the difference between memorization and generalization [CIML]
5. Identify examples of the ethical responsibilities of an ML expert

Q&A