

~~AI~~ / ML

# Outline

1. What is ML / ML research?
2. What's hot?
3. What are particular institutions known for?
4. What is UCSD known for?

# What is machine learning (ML)?

“Scientific discipline that is concerned with the design and development of algorithms that allow computers to change behavior based on data.” – Wikipedia.

- Data
  - Handwritten digits from USPS.
  - User interaction logs from Google.
  - Sensor readings from a robot.
- Behavior
  - Optical character recognition.
  - Targeted advertising.
  - Be like a human.

# How can machines learn?

- Data = examples of (situation,behavior,outcome) triples.
- Infer a policy for what to do in different situations in order to effect the desired outcomes.
- Hopefully, policy generalizes to situations not previously encountered.

# What is machine learning research?

- “Core” ML: develop algorithms for common learning tasks (e.g., classification, ranking, regression, reinforcement learning).
  - Often focus on different issues encountered in applications (e.g., prior knowledge, multiple tasks, sparse structure).
- Applied ML: develop algorithms for real applications (e.g., bio, information retrieval, NLP, robotics, speech, vision).
  - Sometimes also leads to advances in core ML.
- Theoretical ML: mathematical study of learning (e.g., generalization ability of learning algorithms, computational and representational complexity, game-theoretic analysis).
  - Goal is to inform development of core ML and applied ML.

# What's hot?

Some ML-related symposia and workshops from 23rd Annual Conference on Neural Information Processing Systems (NIPS):

- Analyzing networks and learning with graphs
- Applications for topic models: text and beyond
- Causality and time series
- Large-scale machine learning: parallelism and massive datasets
- Partially observable reinforcement learning
- Transfer learning for structured data

# Topic #1: Exploiting low-dimensional structure

- Data increasingly high-dimensional.
  - Dimension = # features/variables describing each data point.  
e.g., NLP: features = words; Bio: features = SNPs.
  - Curse of dimensionality: statistical and computational.
- But features often exhibit statistical dependencies or other structure that can be exploited.
  - Sparsity: only a few relevant features among many irrelevant ones.
  - Latent structure: the high-dimensional observed variables are the noisy / redundant manifestation of a few latent unobserved variables.
- What are the algorithms for discovering such structure? What statistical efficiency gains can be achieved?

## Topic #2: Large-scale machine learning

- Lots of data being generated and collected.
  - Old days:  $\sim 10^4$  examples,  $10^2$ - $10^3$  features (say).
  - Now: e.g., Wikipedia,  $3M$  articles,  $25M$  links,  $2M$  word vocab.
  - Cannot even compute a correlation matrix ( $\Omega(n^2)$  time/space). Many basic optimization routines (e.g. Newton's method) similar or worse.
- Which algorithms dominate in the data-laden regime?
  - e.g. First-order gradient descent vs second-order Newton method?
- How do we parallelize learning algorithms?
  - Split up examples across cores  $\rightarrow$  statistical efficiency suffers.
  - Split up variables across cores  $\rightarrow$  computational efficiency suffers.



# What are particular institutions known for?

Top four schools in AI according to US News & World Report 2008:

1. MIT: the “AI” in “CSAIL” is “artificial intelligence” (see <http://www.csail.mit.edu/node/3#AI>).
2. CMU: machine learning department (see <http://www.ml.cmu.edu/>). Also departments for compbio, NLP, and robotics; Google on campus.
3. Stanford: compbio, NLP, neuroscience, optimization, robotics, ...
4. Berkeley: Bayesian methods, NLP, optimization, theory, vision, ...

The gist: top schools have research in both core ML and applications.

Rest of top ten: UT Austin, UW, GIT, UIUC, UMD, UMass Amherst  
Also industry (Facebook, Google, IBM, Microsoft, NEC, Yahoo!, ...) and  
abroad (MPI, NICTA, Toronto, UCL, ...)

# What is UCSD known for?

ML spans several departments (CSE, ECE, CogSci, Neuroscience, ...)

- Core ML, theory (Charles, Gert (ECE), Lawrence, Sanjoy, Yoav)
- Cognitive science and neuroscience (Gary, Nuno (ECE), ...)
- Data mining (Charles)
- Speech and audio (Gert, Lawrence)
- Vision (David, Nuno, Serge)
- Other applications (Charles, Gert, Lawrence, Yoav, ...)
- ...

See <http://ai.ucsd.edu/> for more.

# What is UCSD known for?

Some BFDs:

- Charles: judge for the \$1M Netflix competition.
- David: founded “Photometria” (virtual cosmetics company).
- Gary: pioneering work in neural networks and cognitive modeling.
- Lawrence: co-editor-in-chief of Journal of Machine Learning Research.
- Sanjoy: co-author of popular algorithms textbook.
- Serge: founded “Digital Persona” (fingerprint tech company).
- Yoav: co-winner of Gödel and Kanellakis prizes for “AdaBoost”.

Questions?