IF5180 – Pembelajaran Mesin Lanjut

Semester Ganjil 2019 - 2020

Dwi H. Widyantoro
Program Studi Informatika
Sekolah Teknik Elektro dan Informatika - ITB

Institut Teknologi Bandung

About IF5032

- Instructor:
  - Dwi H. Widyantoro
  - E-mail: `dwi@stei.itb.ac.id`
- Classroom Meeting: Thursday, 08.00 – 10.00
  Friday, 09.00 – 10.00
- Office hour: by appointment.
- Course web page: to be decided
- Textbook:
  - Tom M. Mitchell, Machine Learning, Mc. Graw Hill, 1997
  - Chris Bishop, Pattern Recognition and Machine Learning book, Springer-Verlag New York Inc., 2016

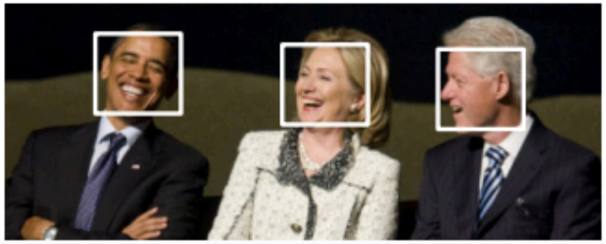
Course Overview

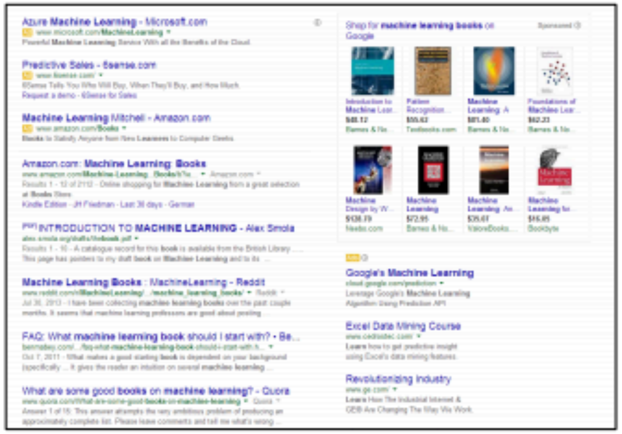
- Decision Tree Learning
- Bayesian Learning
- Logistics Regression
- Graphical Model
- Computational Learning Theory
- Hidden Markov Model
- Learning Representation
- Support Vector Machine
- Boosting
- Neural Network
- Deep Learning

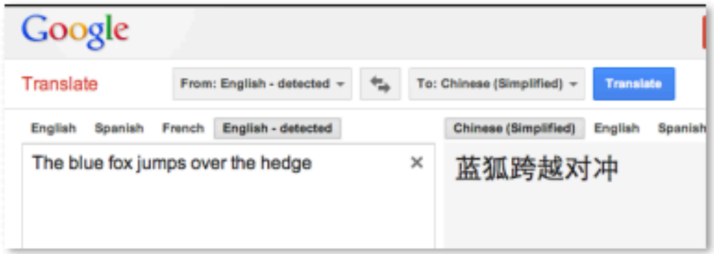
Grading Component

- Class Participation
- Assignment
- Mid-term Exam
- Final-term Exam
- Quiz

Machine Learning is Everywhere







What is Machine Learning

"Learning is any process by which a system improves performance from experience."

- Herbert Simon

Definition by Tom Mitchell (1998):

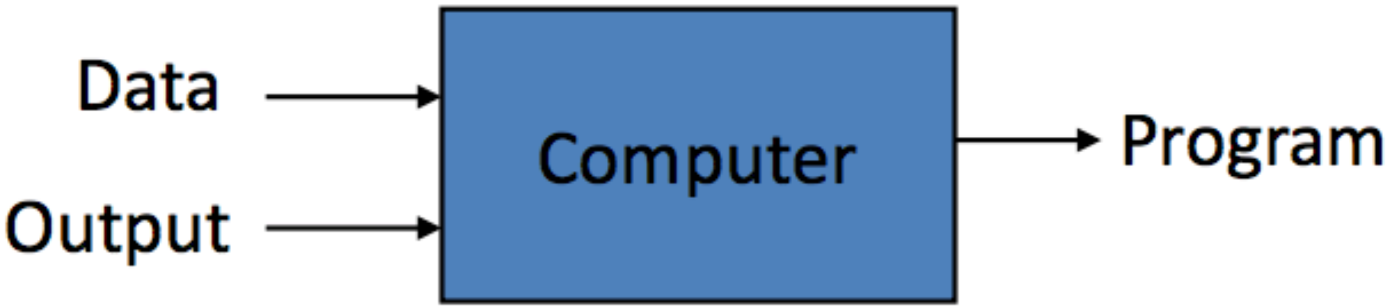
Machine Learning is the study of algorithms that

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

A well-defined learning task is given by <P, T, E>.

Traditional Programming

Machine Learning



Slide credit: Pedro Domingos

Why Study Machine Learning

"A breakthrough in machine learning would be worth ten Microsofts"
-Bill Gates, Chairman, Microsoft

"Machine learning is the next Internet"
-Tony Tether, Director, DARPA

Machine learning is the hot new thing"
-John Hennessy, President, Stanford

"Web rankings today are mostly a matter of machine learning"
-Prabhakar Raghavan, Dir. Research, Yahoo

"Machine learning is going to result in a real revolution"
-Greg Papadopoulos, CTO, Sun

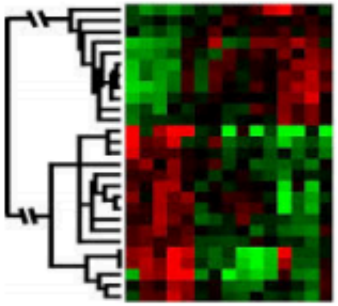
"Machine learning is today's discontinuity"
-Jerry Yang, CEO, Yahoo

When Do We Use Machine Learning

ML is used when:

- Human expertise does not exist (navigating on Mars)
- Humans can't explain their expertise (speech recognition)
- Models must be customized (personalized medicine)
- Models are based on huge amounts of data (genomics)



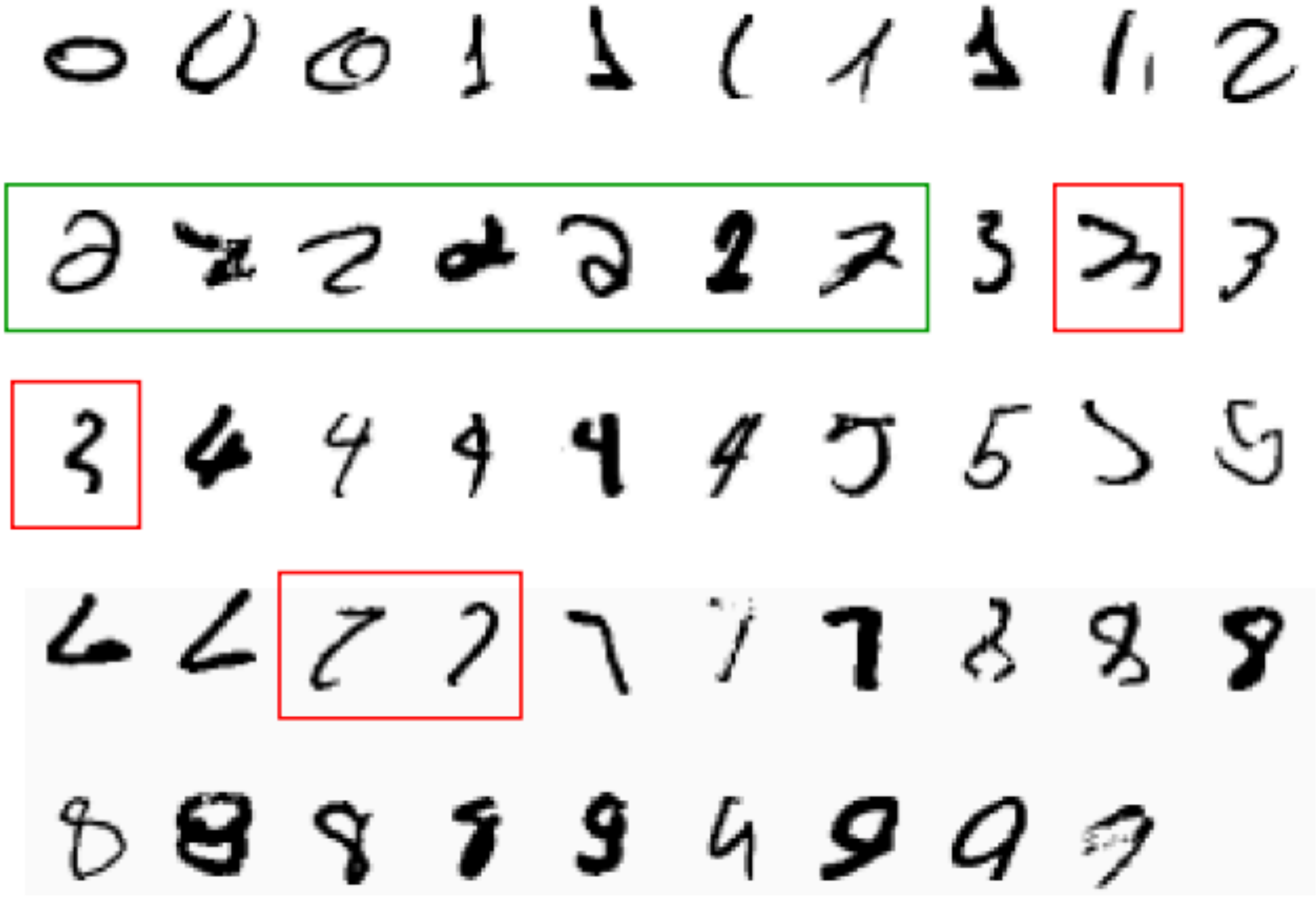



Learning isn't always useful:

- There is no need to "learn" to calculate payroll

Based on slide by E. Alpaydin

A classic example of a task that requires machine learning: It is very hard to say what makes a 2



Slide credit: Geoffrey Hinton

Problems Too Difficult to Program by Hand

Learning = Generalization

H. Simon -

"Learning denotes changes in the system that are adaptive in the sense that they enable the system to do the task or tasks drawn from the same population more efficiently and more effectively the next time."

The ability to perform a task in a situation which has never been encountered before

Some more examples of tasks that are best solved by using a learning algorithm

- Recognizing patterns:
  - Facial identities or facial expressions
  - Handwritten or spoken words
  - Medical images
- Generating patterns:
  - Generating images or motion sequences
- Recognizing anomalies:
  - Unusual credit card transactions
  - Unusual patterns of sensor readings in a nuclear power plant
- Prediction:
  - Future stock prices or currency exchange rates

Other Applications

- Retail: Market basket analysis, Customer relationship management (CRM)
- Finance: Credit scoring, fraud detection
- Manufacturing: Control, robotics, troubleshooting
- Medicine: Medical diagnosis
- Telecommunications: Spam filters, intrusion detection
- Bioinformatics: Motifs, alignment
- Web mining: Search engines
- ...

Defining the Learning Task

Improve on task T, with respect to performance metric P, based on experience E

T: Playing checkers
P: Percentage of games won against an arbitrary opponent
E: Playing practice games against itself

T: Recognizing hand-written words
P: Percentage of words correctly classified
E: Database of human-labeled images of handwritten words

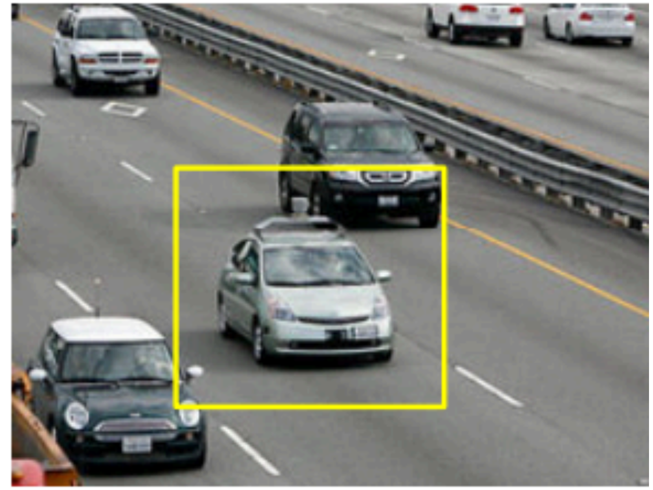
T: Driving on four-lane highways using vision sensors
P: Average distance traveled before a human-judged error
E: A sequence of images and steering commands recorded while observing a human driver.

T: Categorize email messages as spam or legitimate.
P: Percentage of email messages correctly classified.
E: Database of emails, some with human-given labels

Slide credit: Ray Mooney

State of the Art Applications of Machine Learning

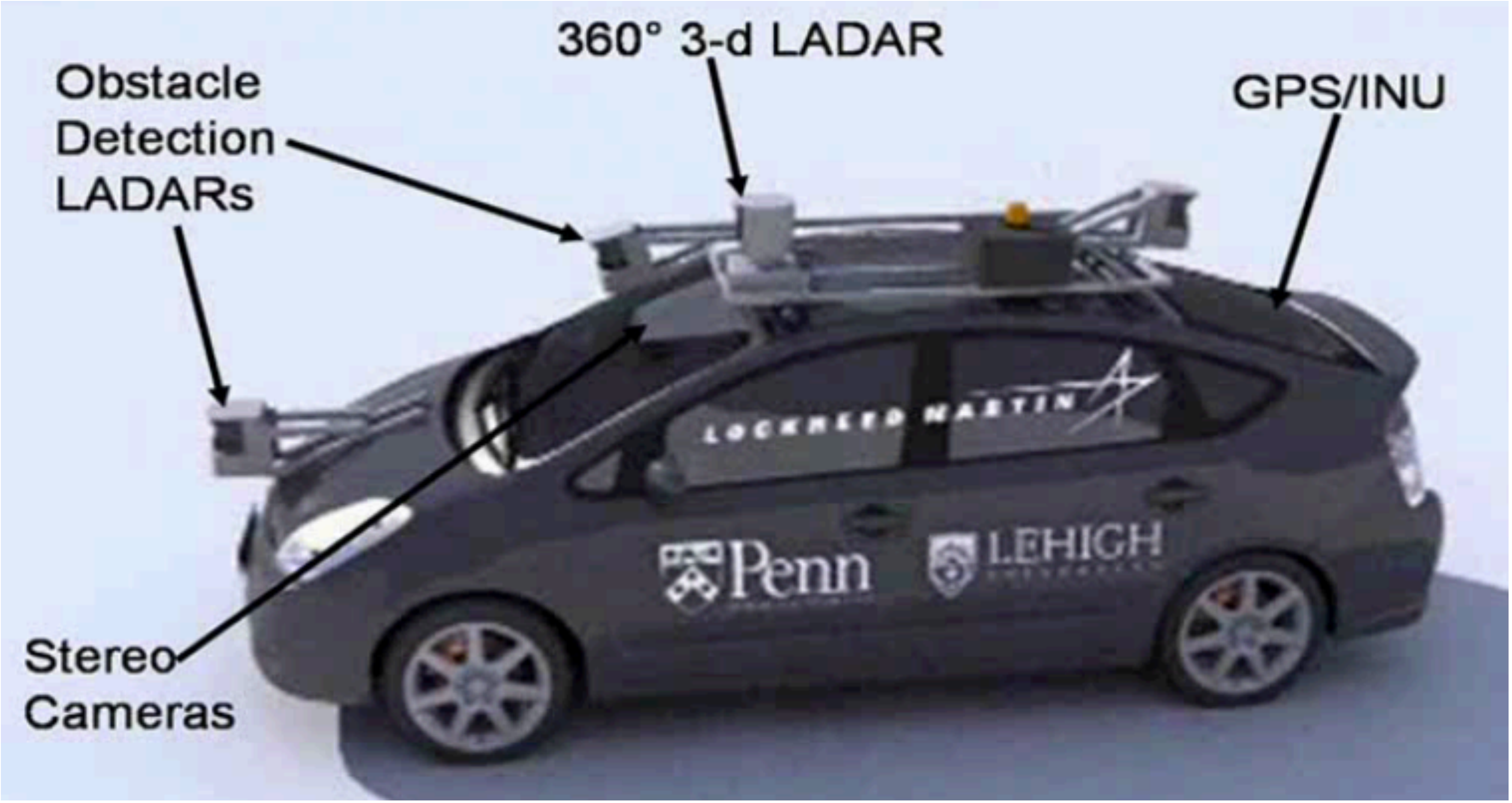
Autonomous Cars



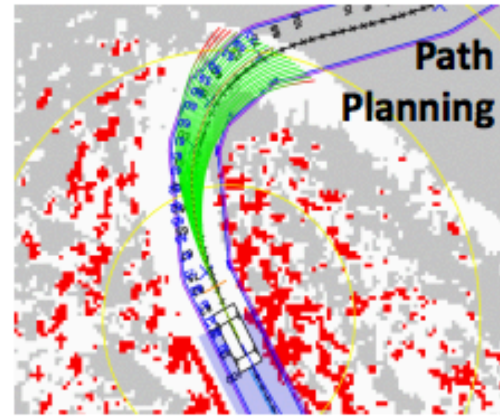
- Nevada made it legal for autonomous cars to drive on roads in June 2011
- As of 2017, 29 states have enacted legislation regarding autonomous cars

Penn's Autonomous Car →
(Ben Franklin Racing Team)

Autonomous Car Sensors

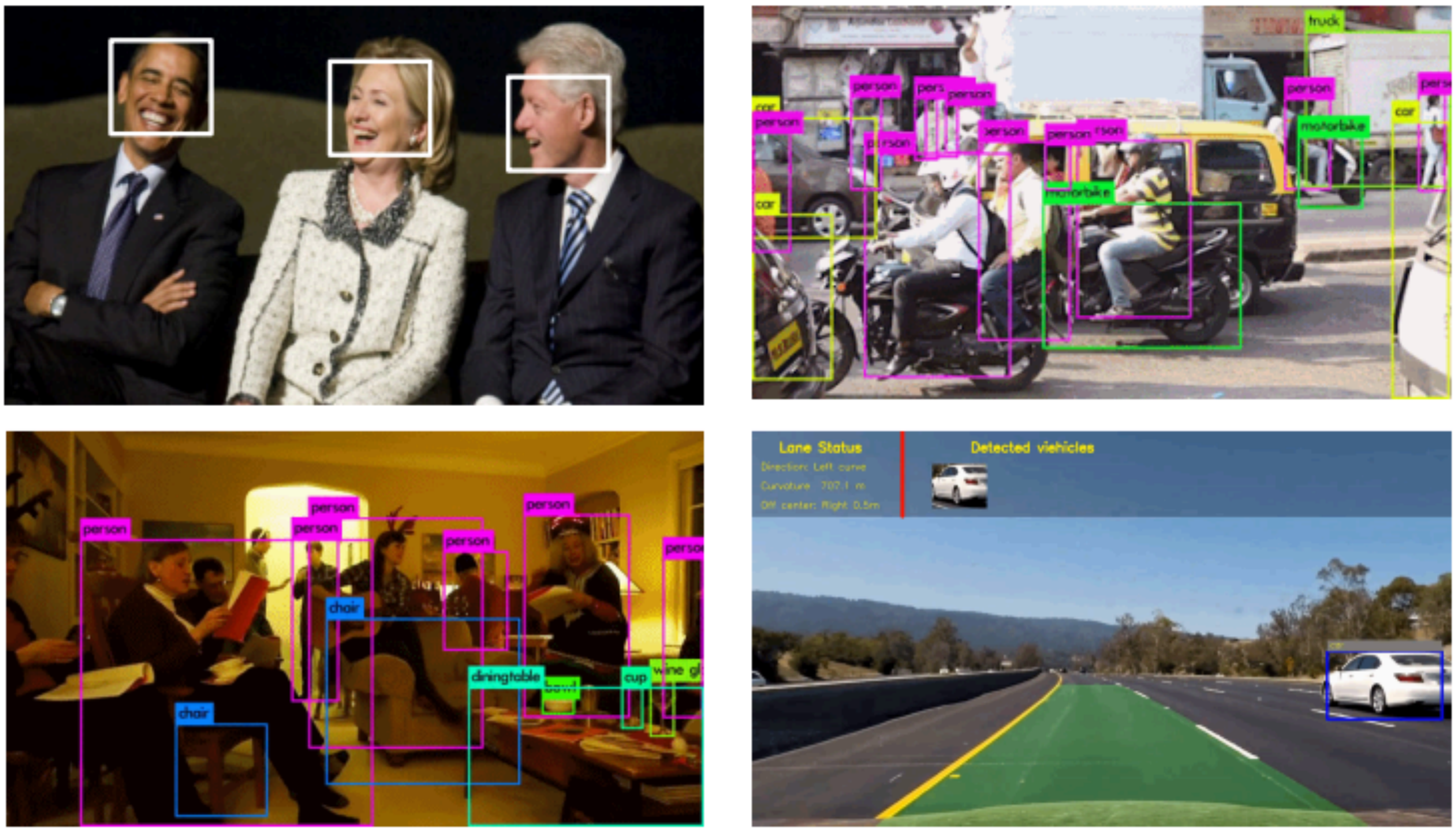


Autonomous Car Technologies



Images and movies taken from Sebastian Thrun's multimedia website.

Applications: Object Detection



Deep Learning in the Headline


BUSINESS NEWS

MIT Technology Review

Is Google Cornering the Market on Deep Learning?

A cutting-edge corner of science is being wooed by Silicon Valley, to the dismay of some academics.

By Antonio Regalado on January 29, 2014

How much are a dozen deep-learning researchers worth? Apparently, more than $400 million.

This week, Google reportedly paid that much to acquire DeepMind Technologies, a startup based in


BloombergBusinessweek Technology

Acquisitions

The Race to Buy the Human Brains Behind Deep Learning Machines

By Ashlee Vance | January 27, 2014

intelligence projects. "DeepMind is bona fide in terms of its research capabilities and depth," says Peter Lee, who heads Microsoft Research.

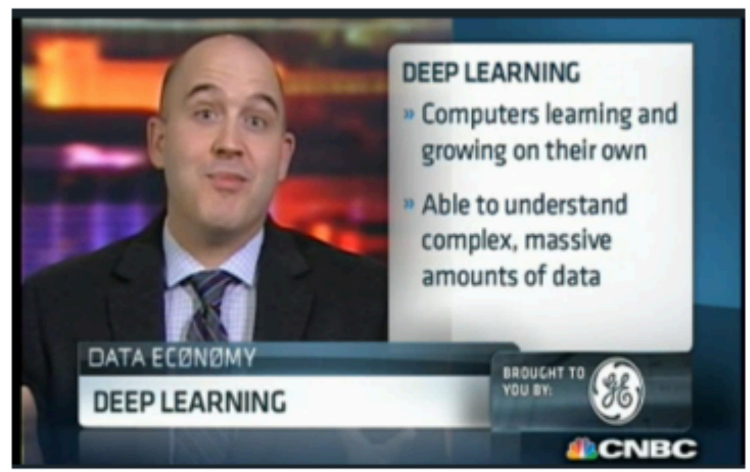
According to Lee, Microsoft, Facebook (FB), and Google find themselves in a battle for deep learning talent. Microsoft has gone from four full-time deep learning experts to 70 in the past three years. "We would have more if the talent was there to


WIRED GEAR SCIENCE ENTERTAINMENT BUSINESS SECURITY

INNOVATION INSIGHTS community content featured

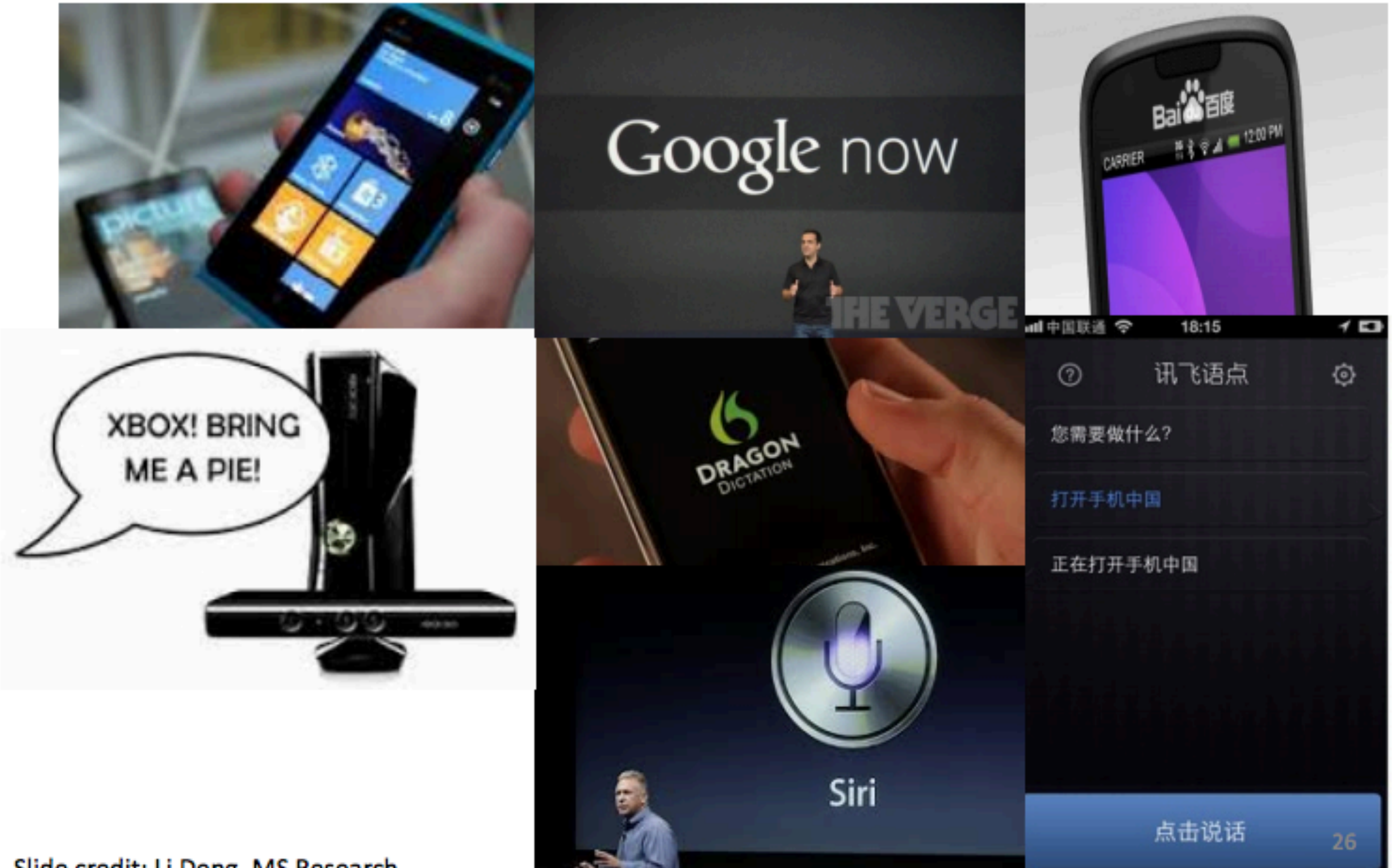
Deep Learning's Role in the Age of Robots

BY JULIAN GREEN, JETPAC 05.02.14 2:56 PM



Scene Labeling via Deep Learning

Impact of Deep Learning in Speech Technology



Slide credit: Li Deng, MS Research

Typical Data mining Task

Data:

Patient103 time=1	→	*Patient103* time=2	... →	*Patient103* time=n
Age: 23		Age: 23		Age: 23
FirstPregnancy: no		FirstPregnancy: no		FirstPregnancy: no
Anemia: no		Anemia: no		Anemia: no
Diabetes: no		Diabetes: YES		Diabetes: no
PreviousPrematureBirth: no		PreviousPrematureBirth: no		PreviousPrematureBirth: no
Ultrasound: ?		Ultrasound: abnormal		Ultrasound: ?
Elective C−Section: ?		Elective C−Section: no		Elective C−Section: no
Emergency C−Section: ?		Emergency C−Section: ?		**Emergency C−Section: Yes**
...		...		...

Typical Data mining Task

- Given:
  - 9714 patient records, each describing a pregnancy and birth
  - Each patient record contains 215 features

- Learn to predict:
  - Classes of future patients at high risk for Emergency Cesarean section

Data mining Result

One of 18 learned rules:

```
If No previous vaginal delivery, and
  Abnormal 2nd Trimester Ultrasound, and
  Malpresentation at admission
Then Prob of Emergency C-Section is 0.6
  Over training data: 26/41 = .63
  Over test data: 12/20 = .60
```

Credit Risk Analysis

Customer103: (time=t0)

Years of credit: 9
Loan balance: $2,400
Income: $52k
Own House: Yes
Other delinquent accts: 2
Max billing cycles late: 3
Profitable customer?: ?
...

Customer103: (time=t1)

Years of credit: 9
Loan balance: $3,250
Income: ?
Own House: Yes
Other delinquent accts: 2
Max billing cycles late: 4
Profitable customer?: ?
...

...

Customer103: (time=tn)

Years of credit: 9
Loan balance: $4,500
Income: ?
Own House: Yes
Other delinquent accts: 3
Max billing cycles late: 6
Profitable customer?: No
...

Credit Risk Analysis

Rules learned from synthesized data:

If Other-Delinquent-Account > 2, and
  Number-Delinquent-Billing-Cycles > 1
Then Profitable-Customer = No
  [Deny Credit Application]

If Other-Delinquent-Accounts = 0, and
  (Income > $30k) or (Years-of-Credit > 3)
Then Profitable-Customer = Yes
  [Accept Credit Card Application]

Customer Purchase Behavior

Customer103: (time=t0)	*Customer103:* (time=t1)	...	*Customer103:* (time=tn)
Sex: M	Sex: M		Sex: M
Age: 53	Age: 53		Age: 53
Income: $50k	Income: $50k		Income: $50k
Own House: Yes	Own House: Yes		Own House: Yes
MS Products: Word	MS Products: Word		MS Products: Word
Computer: 386 PC	Computer: Pentium		Computer: Pentium
Purchase Excel?: ?	Purchase Excel?: ?		**Purchase Excel?: Yes**
...	...		...

Customer Retention

Customer103: (time=t0)

Sex: M
Age: 53
Income: $50k
Own House: Yes
Checking: $5k
Savings: $15k
Current-customer?: yes

Customer103: (time=t1)

Sex: M
Age: 53
Income: $50k
Own House: Yes
Checking: $20k
Savings: $0
Current-customer?: yes

...

Customer103: (time=tn)

Sex: M
Age: 53
Income: $50k
Own House: Yes
Checking: $0
Savings: $0
Current-customer?: No

Relevant Disciplines

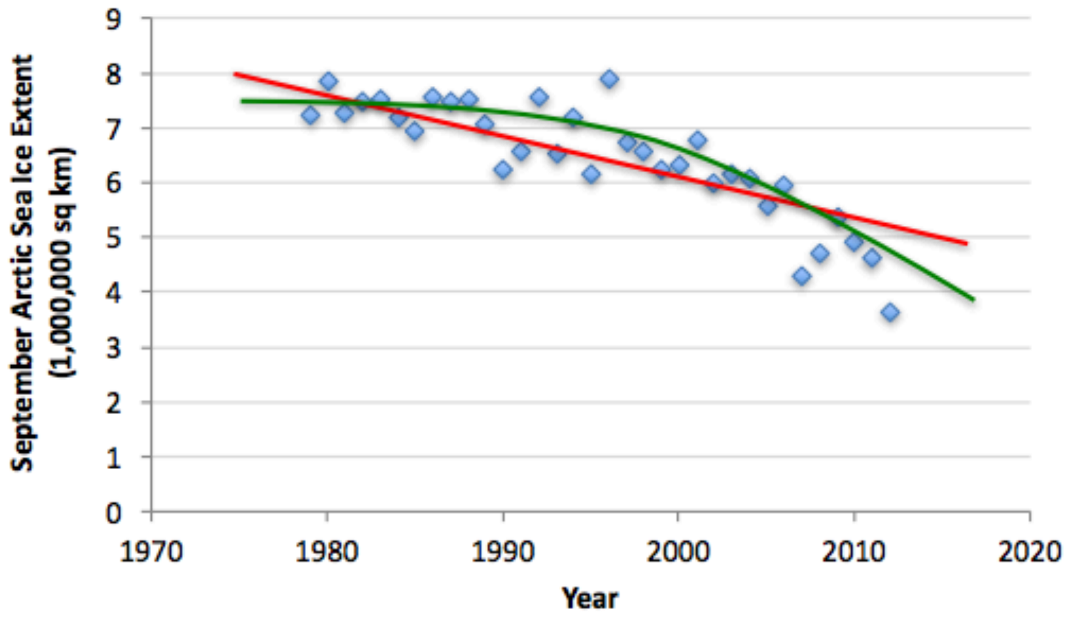
- Artificial intelligence
- Bayesian methods
- Computational complexity theory
- Control theory
- Information theory
- Philosophy
- Psychology and neurobiology
- Statistics, …

Learning Approaches

- **Supervised (inductive) learning**
  – Given: training data + desired outputs (labels)
- **Unsupervised learning**
  – Given: training data (without desired outputs)
- **Semi-supervised learning**
  – Given: training data + a few desired outputs
- **Reinforcement learning**
  – Rewards from sequence of actions

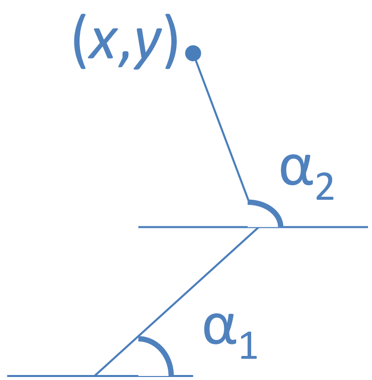
Supervised Learning: Regression

- Given $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$
- Learn a function $f(x)$ to predict y given x
  - y is real-valued == regression



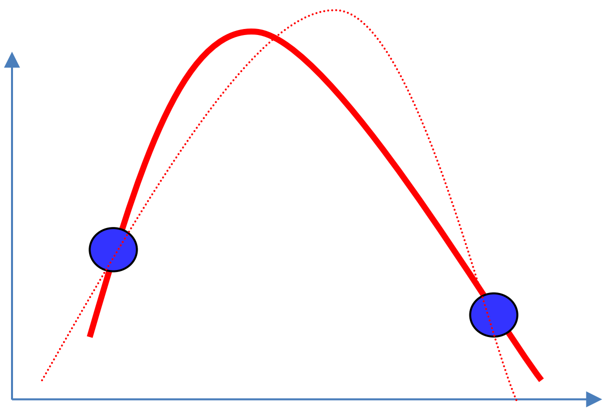
Regression Applications

- Navigating a car: Angle of the steering
- Prediction of Stock markets
- Kinematics of a robot arm



$\alpha_1 = g_1(x,y)$

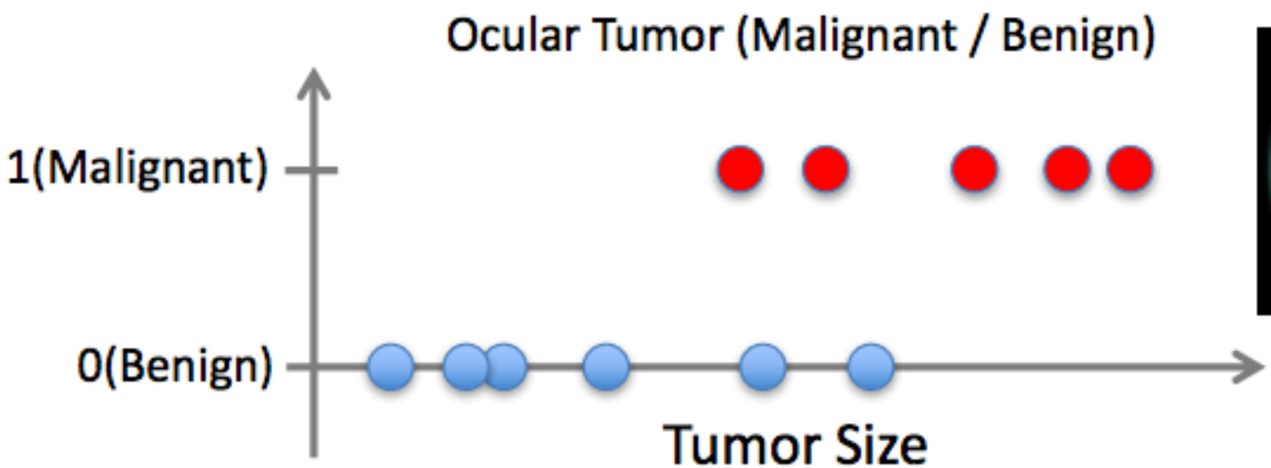
$\alpha_2 = g_2(x,y)$

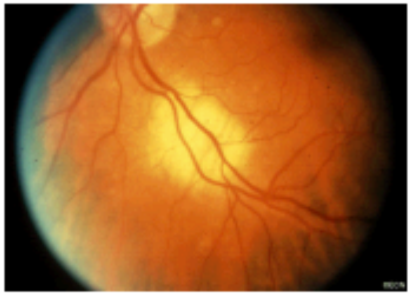


- Response surface design

Supervised Learning: Classification

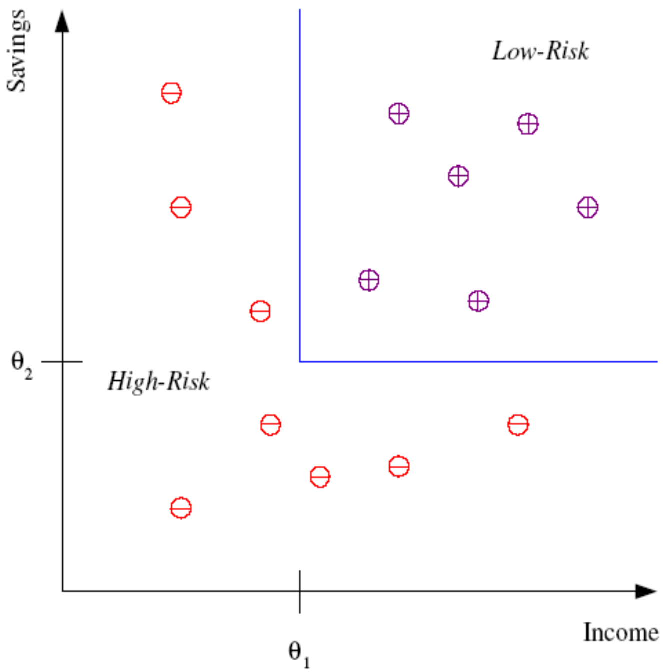
- Given $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$
- Learn a function $f(x)$ to predict y given x
  - y is categorical == classification





Classification

- Example: Credit scoring
- Differentiating between low-risk and high-risk customers from their *income* and *savings*



Discriminant: IF *income* > θ_1 AND *savings* > θ_2
THEN low-risk ELSE high-risk

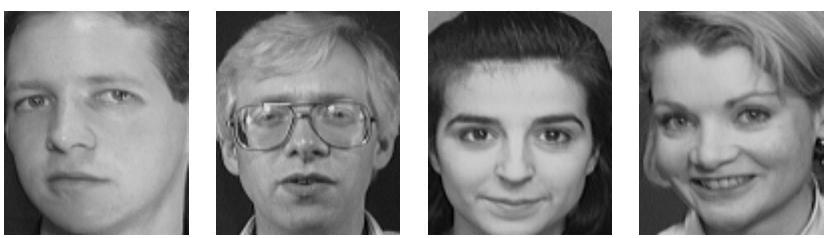
Classification: Applications

- Aka Pattern recognition
- Face recognition: Pose, lighting, occlusion (glasses, beard), make-up, hair style
- Character recognition: Different handwriting styles.
- Speech recognition: Temporal dependency.
- Medical diagnosis: From symptoms to illnesses
- Biometrics: Recognition/authentication using physical and/or behavioral characteristics: Face, iris, signature, etc
- ...

Face Recognition

Training examples of a person

Test images



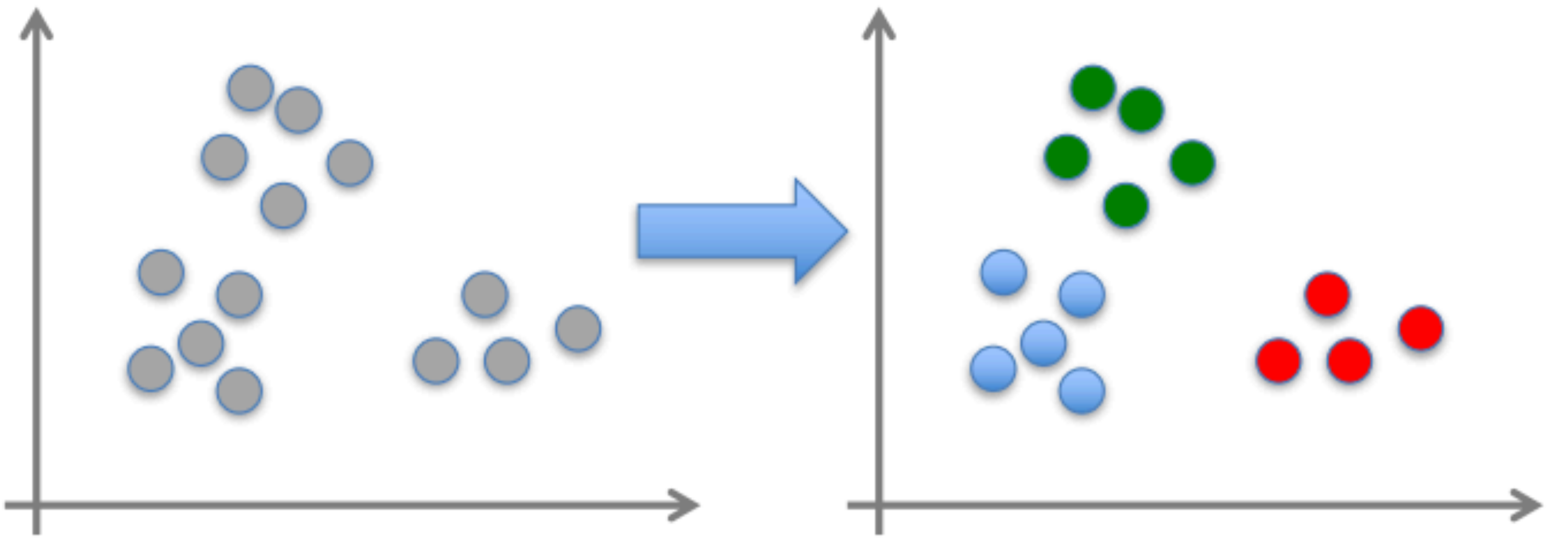
ORL dataset,
AT&T Laboratories, Cambridge UK

Supervised Learning: Uses

- Prediction of future cases: Use the rule to predict the output for future inputs
- Knowledge extraction: The rule is easy to understand
- Compression: The rule is simpler than the data it explains
- Outlier detection: Exceptions that are not covered by the rule, e.g., fraud

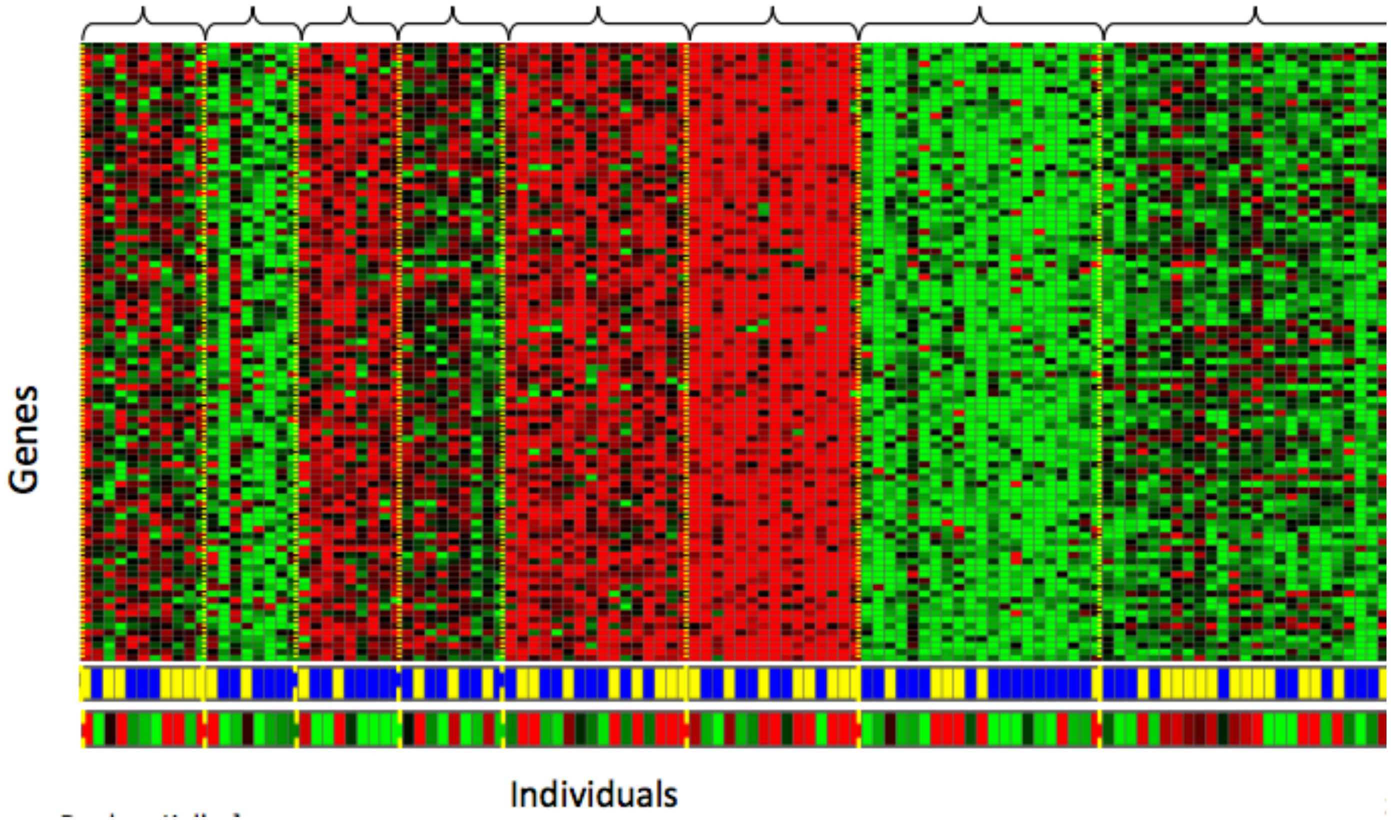
Unsupervised Learning

- Given x_1, x_2, ..., x_n (without labels)
- Output hidden structure behind the x's
  - E.g., clustering



Unsupervised Learning

Genomics application: group individuals by genetic similarity



Unsupervised Learning

- Independent component analysis – separate a combined signal into its original sources

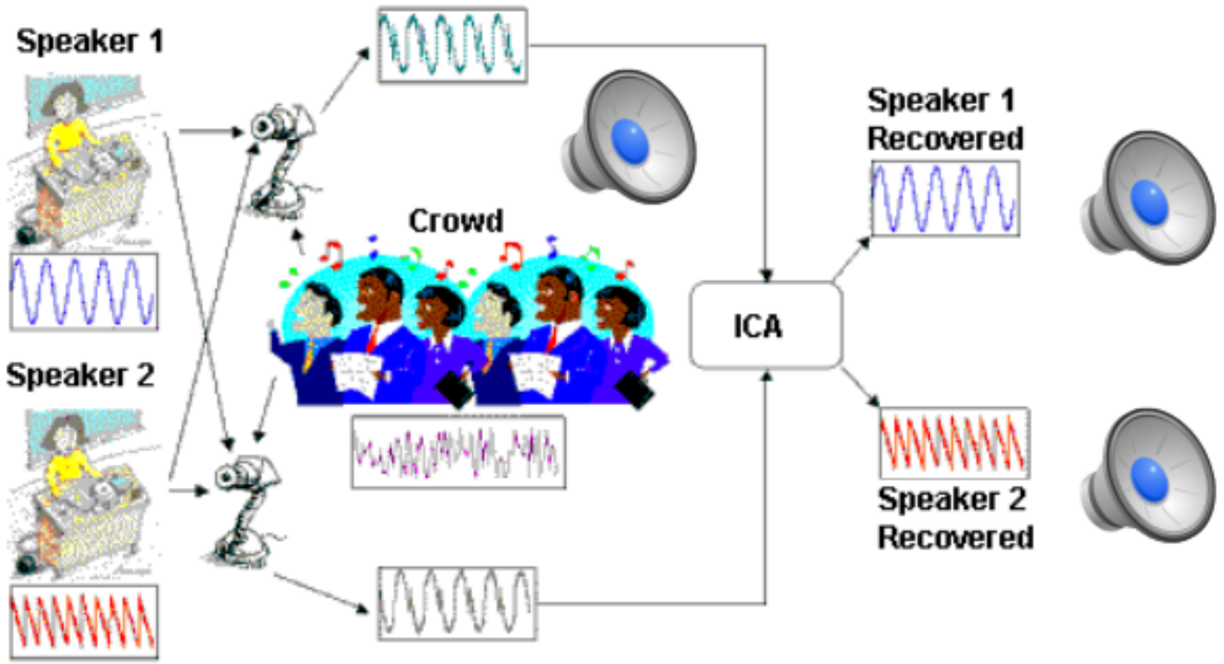
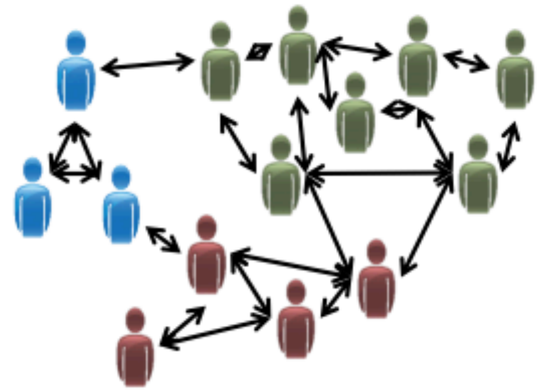


Image credit: statsoft.com Audio from http://www.ism.ac.jp/~shiro/research/blindsep.html

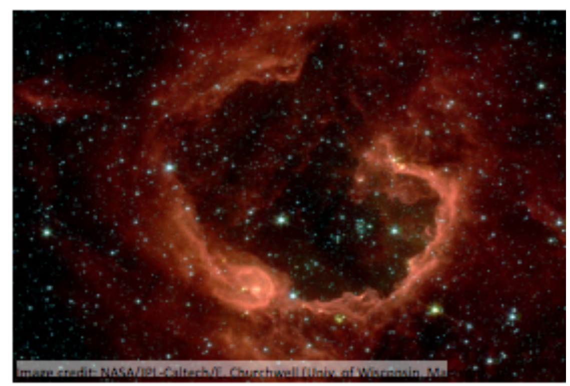
Unsupervised Learning

Organize computing clusters



Social network analysis

Market segmentation



Astronomical data analysis

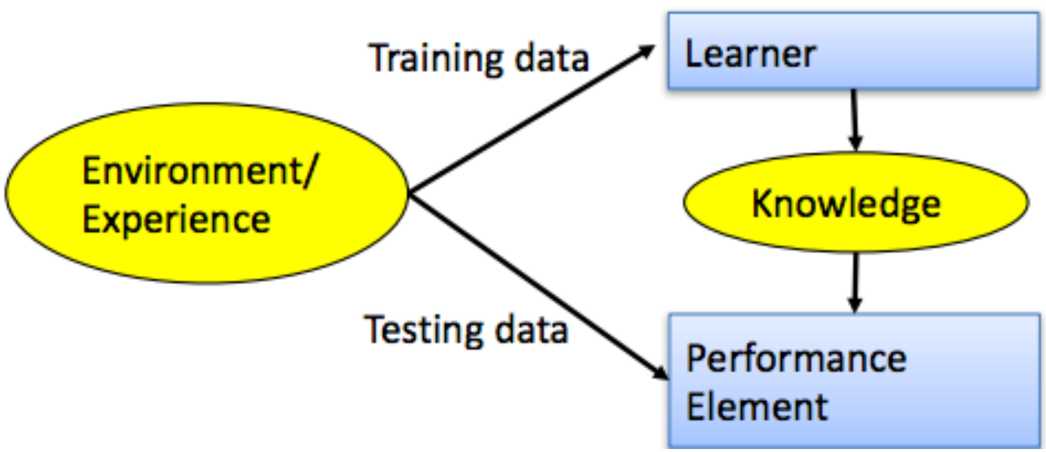
Reinforcement Learning

- Given a sequence of states and actions with (delayed) rewards, output a policy
  - Policy is a mapping from states → actions that tells you what to do in a given state
- Examples:
  - Credit assignment problem
  - Game playing
  - Robot in a maze
  - Balance a pole on your hand

Framing a Learning Problem

Design a Learning System

- Choose the training experience
- Choose exactly what is to be learned
  - i.e. the ***target function***
- Choose how to represent the target function
- Choose a learning algorithm to infer the target function from the experience



Training vs Test Distribution

- We generally assume that the training and test examples are independently drawn from the same overall distribution of data
  - We call this "i.i.d" which stands for "independent and identically distributed"

- If examples are not independent, requires ***collective classification***

- If test distribution is different, requires ***transfer learning***

ML in a Nutshell

- Tens of thousands of machine learning algorithms
  - Hundreds new every year
- Every ML algorithm has three components:
  - **Representation**
  - **Optimization**
  - **Evaluation**

Various Function Representations

- Numerical functions
  - Linear regression
  - Neural networks
  - Support vector machines
- Symbolic functions
  - Decision trees
  - Rules in propositional logic
  - Rules in first-order predicate logic
- Instance-based functions
  - Nearest-neighbor
  - Case-based
- Probabilistic Graphical Models
  - Naïve Bayes
  - Bayesian networks
  - Hidden-Markov Models (HMMs)
  - Probabilistic Context Free Grammars (PCFGs)
  - Markov networks

Various Search/Optimization Algorithms

- Gradient descent
  - Perceptron
  - Backpropagation
- Dynamic Programming
  - HMM Learning
  - PCFG Learning
- Divide and Conquer
  - Decision tree induction
  - Rule learning
- Evolutionary Computation
  - Genetic Algorithms (GAs)
  - Genetic Programming (GP)
  - Neuro-evolution

Evaluation

- Accuracy
- Precision and recall
- Squared error
- Likelihood
- Posterior probability
- Cost / Utility
- Margin
- Entropy
- K-L divergence
- etc.

ML in Practice

Loop

- Understand domain, prior knowledge, and goals
- Data integration, selection, cleaning, pre-processing, etc.
- Learn models
- Interpret results
- Consolidate and deploy discovered knowledge

Lessons Learned about Learning

- Learning can be viewed as using direct or indirect experience to approximate a chosen target function.

- Function approximation can be viewed as a search through a space of hypotheses (representations of functions) for one that best fits a set of training data.

- Different learning methods assume different hypothesis spaces (representation languages) and/or employ different search techniques.

A Brief History about Machine Learning

History of Machine Learning

- 1950s
  - Samuel's checker player
  - Selfridge's Pandemonium
- 1960s:
  - Neural networks: Perceptron
  - Pattern recognition
  - Learning in the limit theory
  - Minsky and Papert prove limitations of Perceptron
- 1970s:
  - Symbolic concept induction
  - Winston's arch learner
  - Expert systems and the knowledge acquisition bottleneck
  - Quinlan's ID3
  - Michalski's AQ and soybean diagnosis
  - Scientific discovery with BACON
  - Mathematical discovery with AM

History of Machine Learning (cont.)

- 1980s:
  - Advanced decision tree and rule learning
  - Explanation-based Learning (EBL)
  - Learning and planning and problem solving
  - Utility problem
  - Analogy
  - Cognitive architectures
  - Resurgence of neural networks (connectionism, backpropagation)
  - Valiant's PAC Learning Theory
  - Focus on experimental methodology
- 1990s
  - Data mining
  - Adaptive software agents and web applications
  - Text learning
  - Reinforcement learning (RL)
  - Inductive Logic Programming (ILP)
  - Ensembles: Bagging, Boosting, and Stacking
  - Bayes Net learning

History of Machine Learning (cont.)

- 2000s
  - Support vector machines & kernel methods
  - Graphical models
  - Statistical relational learning
  - Transfer learning
  - Sequence labeling
  - Collective classification and structured outputs
  - Computer Systems Applications (Compilers, Debugging, Graphics, Security)
  - E-mail management
  - Personalized assistants that learn
  - Learning in robotics and vision
- 2010s
  - Deep learning systems
  - Learning for big data
  - Bayesian methods
  - Multi-task & lifelong learning
  - Applications to vision, speech, social networks, learning to read, etc.
  - ???

Some Issues in Machine Learning

- What algorithms can approximate functions well (and when)?
- How does number of training examples influence accuracy?
- How does complexity of hypothesis representation impact it?
- How does noisy data influence accuracy?
- What are the theoretical limits of learnability?
- How can prior knowledge of learner help?
- What clues can we get from biological systems?
- How can systems alter their own representations?

Resources: Datasets

- UCI Repository: http://www.ics.uci.edu/~mlearn/MLRepository.html
- UCI KDD Archive: http://kdd.ics.uci.edu/summary.data.application.html
- Statlib: http://lib.stat.cmu.edu/
- Delve: http://www.cs.utoronto.ca/~delve/

Resources: Journals

- Journal of Machine Learning Research www.jmlr.org
- Machine Learning
- Neural Computation
- Neural Networks
- IEEE Transactions on Neural Networks
- IEEE Transactions on Pattern Analysis and Machine Intelligence
- Annals of Statistics
- Journal of the American Statistical Association
- ...

Resources: Conferences

- International Conference on Machine Learning (ICML)
- European Conference on Machine Learning (ECML)
- Neural Information Processing Systems (NIPS)
- Uncertainty in Artificial Intelligence (UAI)
- Computational Learning Theory (COLT)
- International Conference on Artificial Neural Networks (ICANN)
- International Conference on AI & Statistics (AISTATS)
- International Conference on Pattern Recognition (ICPR)
- ...