CS 412 Introduction to Machine Learning

Introduction to CS 412

Instructor: Wei Tang

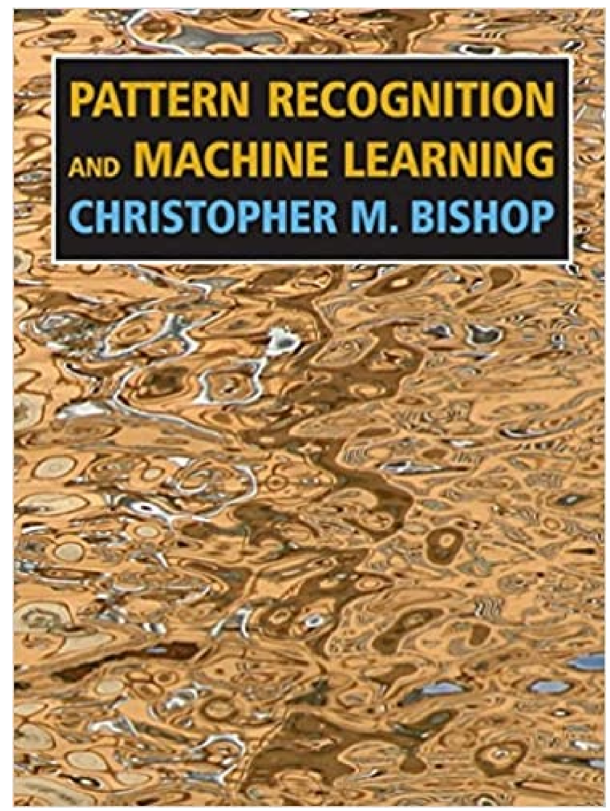
Department of Computer Science
University of Illinois at Chicago
Chicago IL 60607

https://tangw.people.uic.edu
tangw@uic.edu

Slides credit: Sargur Srihari, Eric Eaton

Course information

- Textbook
  - "Pattern Recognition and Machine Learning“, Chris Bishop, Springer 2006
- Instructor office hours
  - Fri 1:30 PM - 2:30 PM
  - Blackboard Collaborate

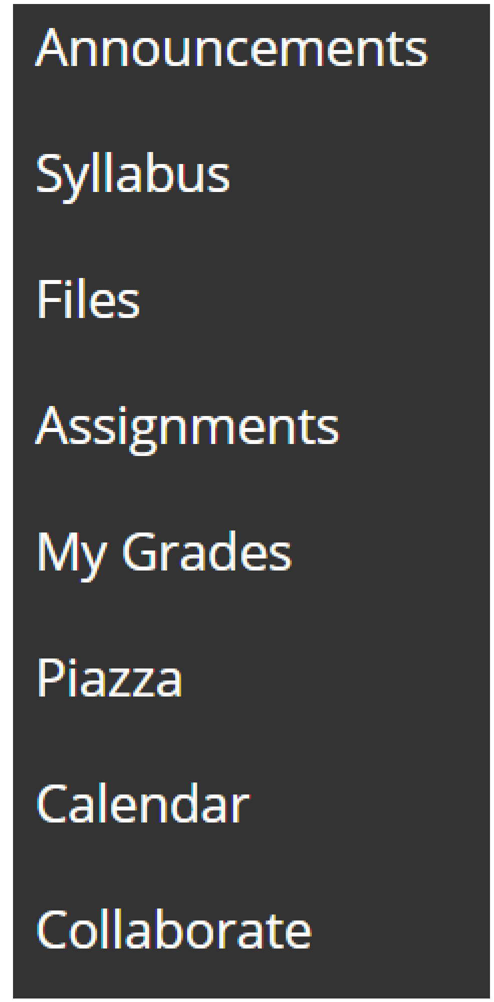


Teaching assistant (TA)

- Peng Zou
  - CS PhD candidate
  - pzou3@uic.edu
- Homework grading
- TA office hours
  - Wes 3 PM - 4 PM
  - Blackboard Collaborate

Course resources

- Blackboard
  - Syllabus
  - Announcements
  - Discussion forum (Piazza)
  - Slides
  - Assignments
  - Project
  - Grades



Grade

- Grades
  - 4 machine problems: 25%
    - Late submission: 1 pt off per hour (2 days max)
  - Final project presentation: 10%
  - Final project report: 15%
  - Final exam: 50%
- Bonus
  - Optional part of machine problems
  - Participation in Piazza (up to 3 pts added to final grade): answers endorsed by TA or Instructor

Final exam is on-campus

- The current plan is to have an on-campus final exam during the UIC Final Exam week.
  - between 12/6 and 12/10.
- The exact date will be announced later in the semester.

Slides

- Slides will be available on Blackboard
  - Files/Notes
  - Check a few minutes before each lecture

Final projects

- Apply what you have learned to solve a problem interests you
- Projects should be done individually or in groups of up to 3 students.

Final projects

1. Find a machine learning application that interests you.
2. Find an appropriate dataset for this application.
   – Check UCI Machine Learning Repository (https://archive.ics.uci.edu/ml/datasets.php)
3. Identify the learning problem: is it classification, clustering or regression?
4. Apply 2 or 3 techniques you learned in the course to solve the problem.
   – You may also try something new.
   – You are free to use code from any machine learning library.
5. Evaluate each method and compare them.
   – In addition to the accuracy, you may also compare their training time, inference speeds and memory consumption

Final projects

- <u>Each group</u> should present a 5 minutes talk
- <u>Each student</u> should submit an individual report
  - Collaboration on the report is prohibited
- More details in Blackboard

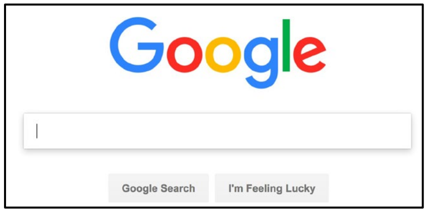
Course policy

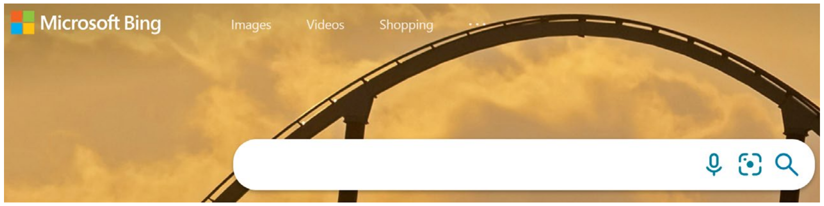
- Late submission penalty: 1 point off per hour (2 days max).
- No late submission for the project report is allowed.
- Collaboration on the project report is prohibited.
- Zero tolerance on plagiarism.

Any questions on the course?

Today AI is ubiquitous

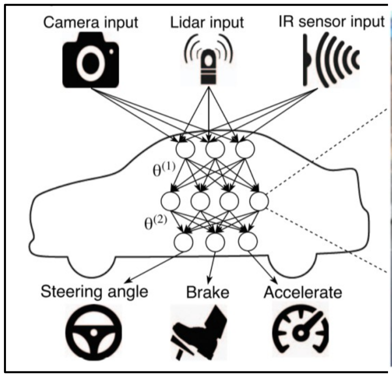
- Automate routine labor
  - Search

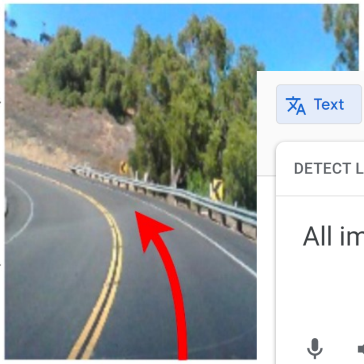




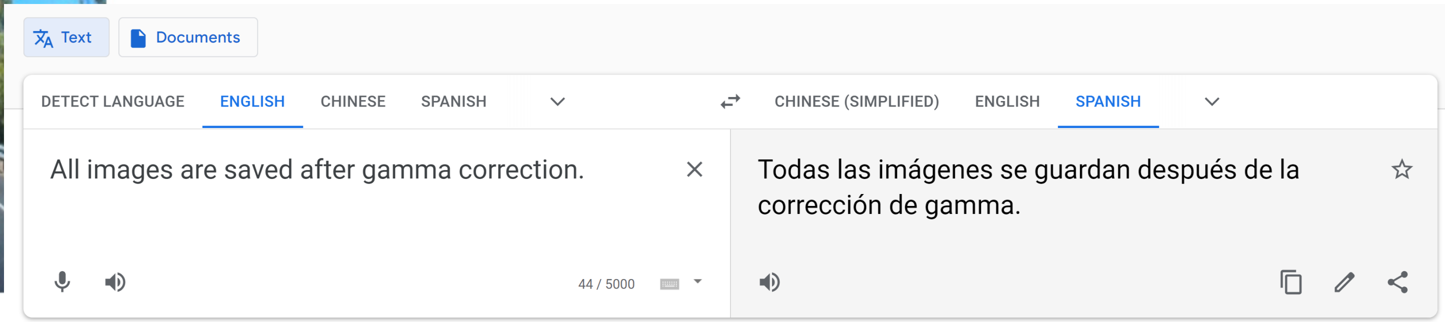
- Understand speech
  - SIRI, Alexa

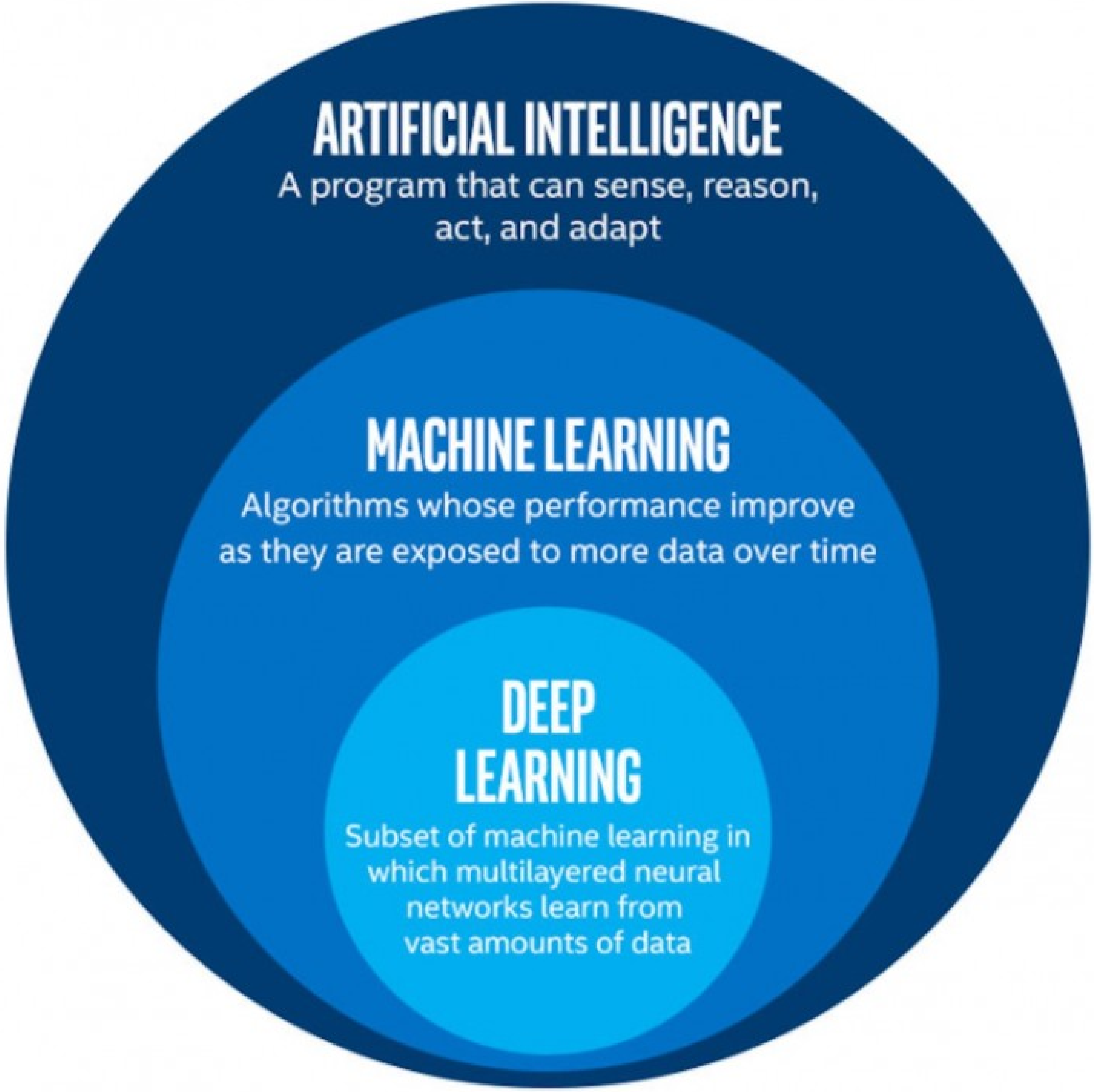
- Autonomous Vehicles



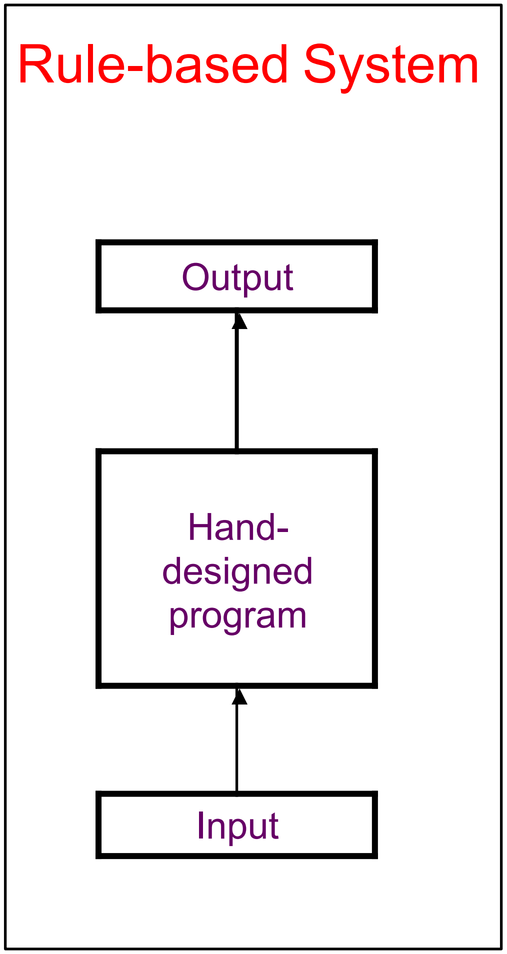


- Machine Translation




ARTIFICIAL INTELLIGENCE
A program that can sense, reason, act, and adapt
MACHINE LEARNING
Algorithms whose performance improve as they are exposed to more data over time
DEEP LEARNING
Subset of machine learning in which multilayered neural networks learn from vast amounts of data

Knowledge-Based AI



Disadvantage:

- Time of human experts
- People struggle to formalize rules with enough complexity to describe the world

The Machine Learning approach

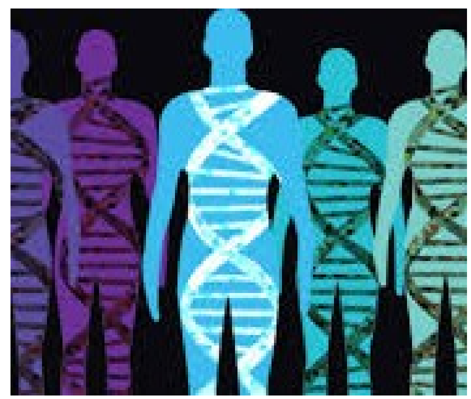
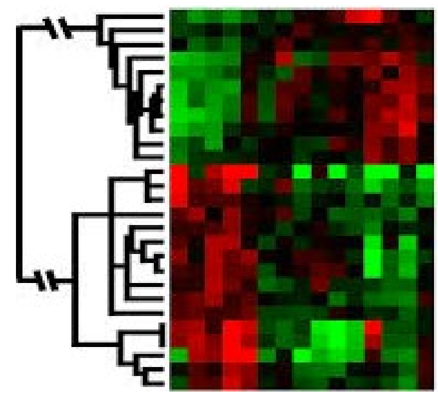
- Difficulties of hard-coded approach suggests:
  - Allow computers to learn from experience
- First determine what *features* to use
  - E.g., zip code, school scores for housing price prediction
- Learn to map the features to outputs

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 are based on huge amounts of data (genomics)



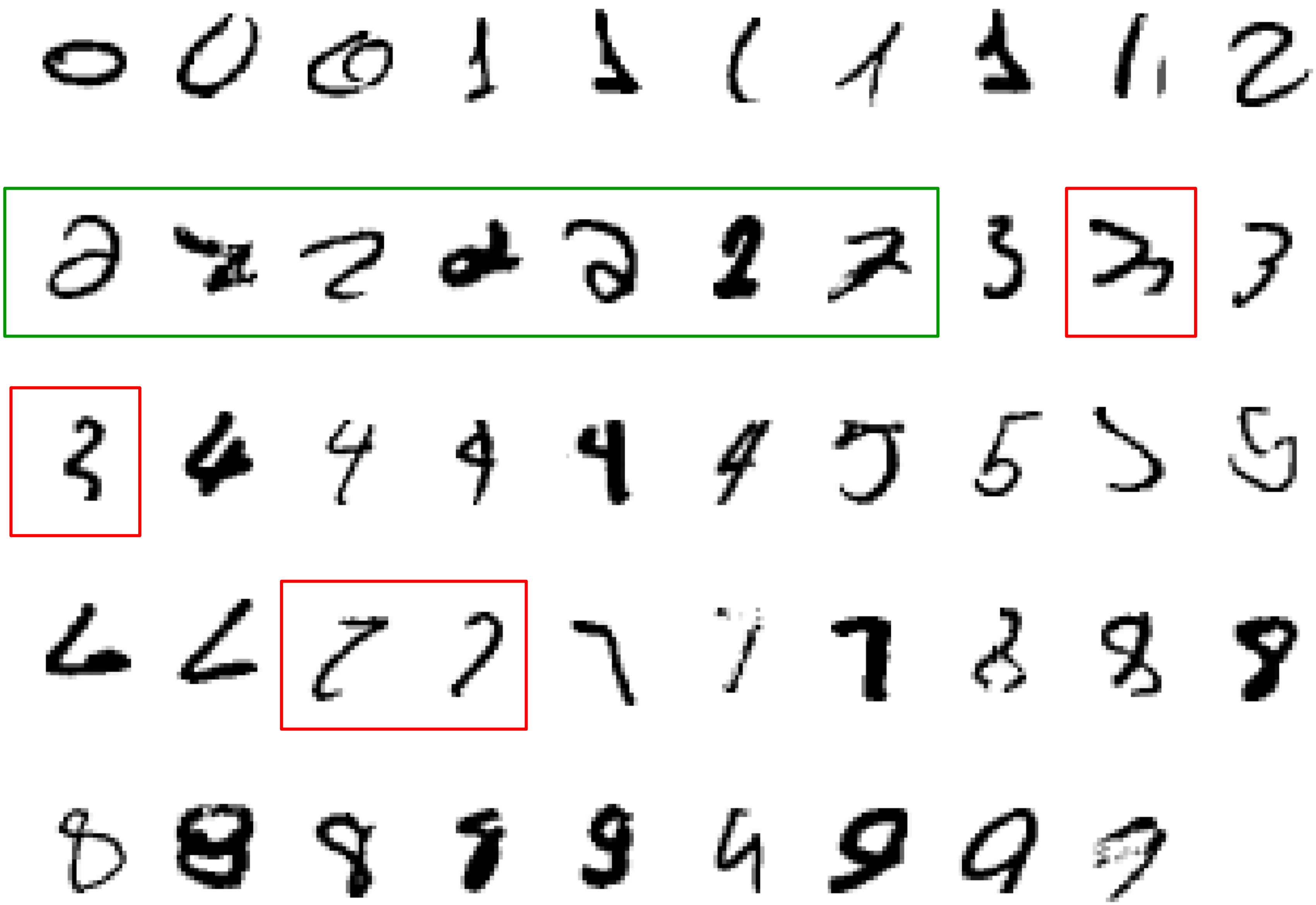



Learning isn't always useful:

- 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

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 sentences
- Recognizing anomalies:
  - Unusual credit card transactions
  - Unusual patterns of sensor readings in a nuclear power plant
- Prediction:
  - Future stock prices or currency exchange rates

Slide credit: Geoffrey Hinton

Sample Applications

- Web search
- Computational biology
- Finance
- E-commerce
- Robotics
- Information extraction
- Social networks
- Debugging software
- [Your favorite area]

Slide credit: Pedro Domingos

What is machine learning?

"Machine Learning: Field of study that gives computers the ability to learn without being explicitly programmed."   -Arthur Samuel (1959)

Defining the Learning Task

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

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

T: Predict stock price from price history.
P: Difference between the predicted and true prices.
E: Database of stock price histories

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

Slide credit: Ray Mooney

State of the Art Applications of Machine Learning

Deep Learning in the Headlines

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

This is Freescale make it

Bloomberg Businessweek 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.

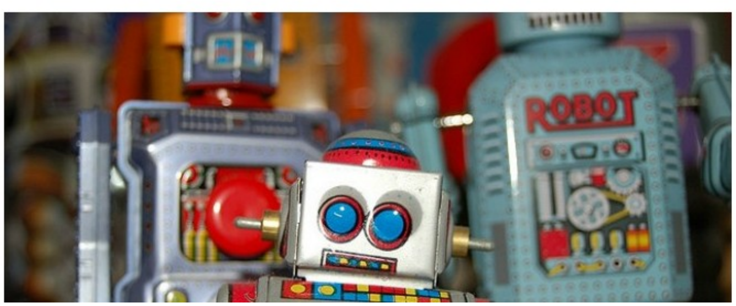
According to Lee, Microsoft, Facebook (FB), and Google find themselves in a battle for deep learning talent. Microsoft has gone from four full-time deep learning experts to 70 in the past three years. "We would have more if the talent was there to

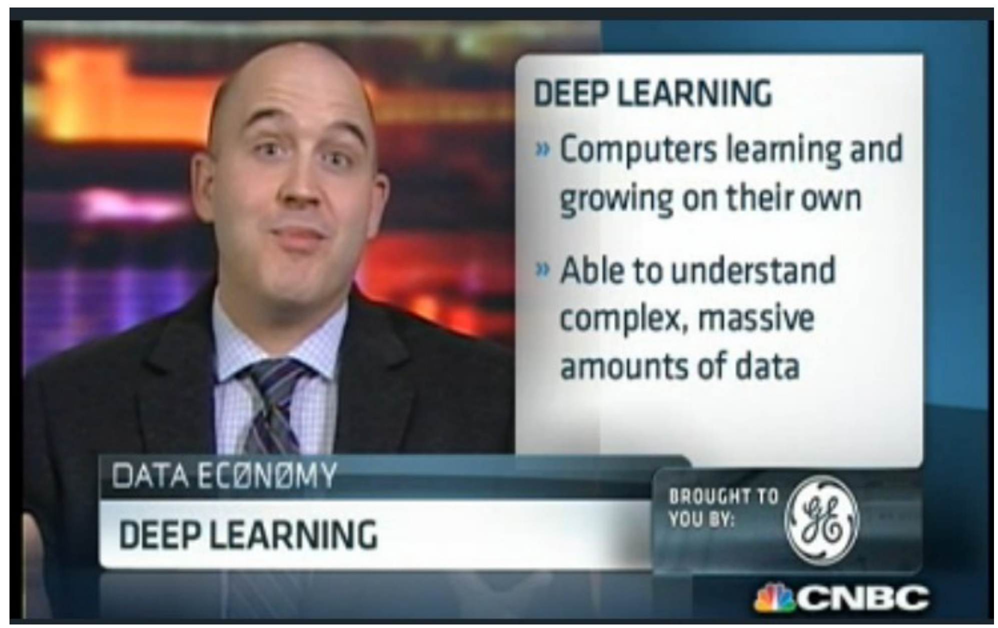
WIRED GEAR SCIENCE ENTERTAINMENT BUSINESS SECURITY DESIG

INNOVATION INSIGHTS community content featured

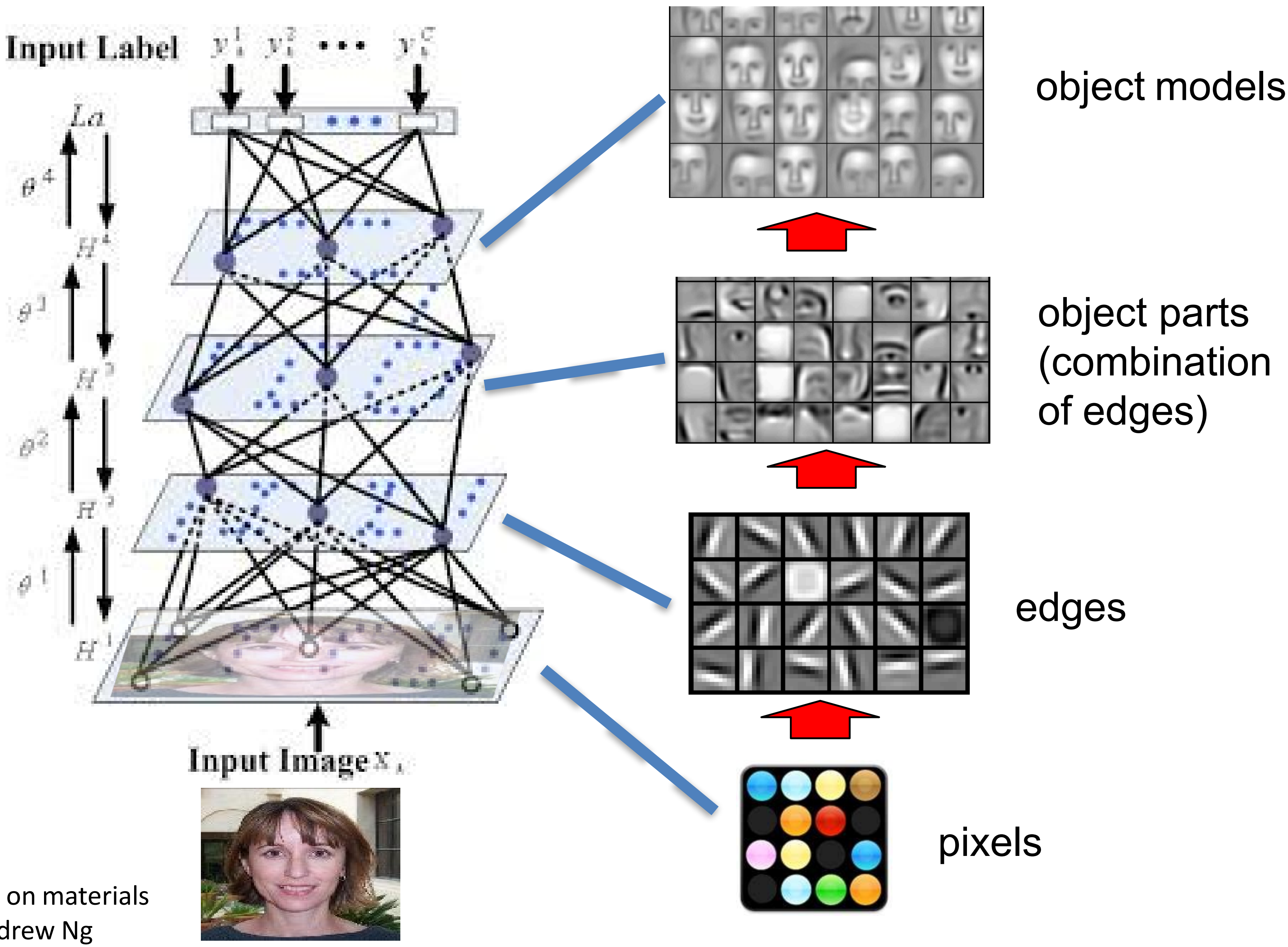
Deep Learning's Role in the Age of Robots

BY JULIAN GREEN, JETPAC 05.02.14 2:56 PM



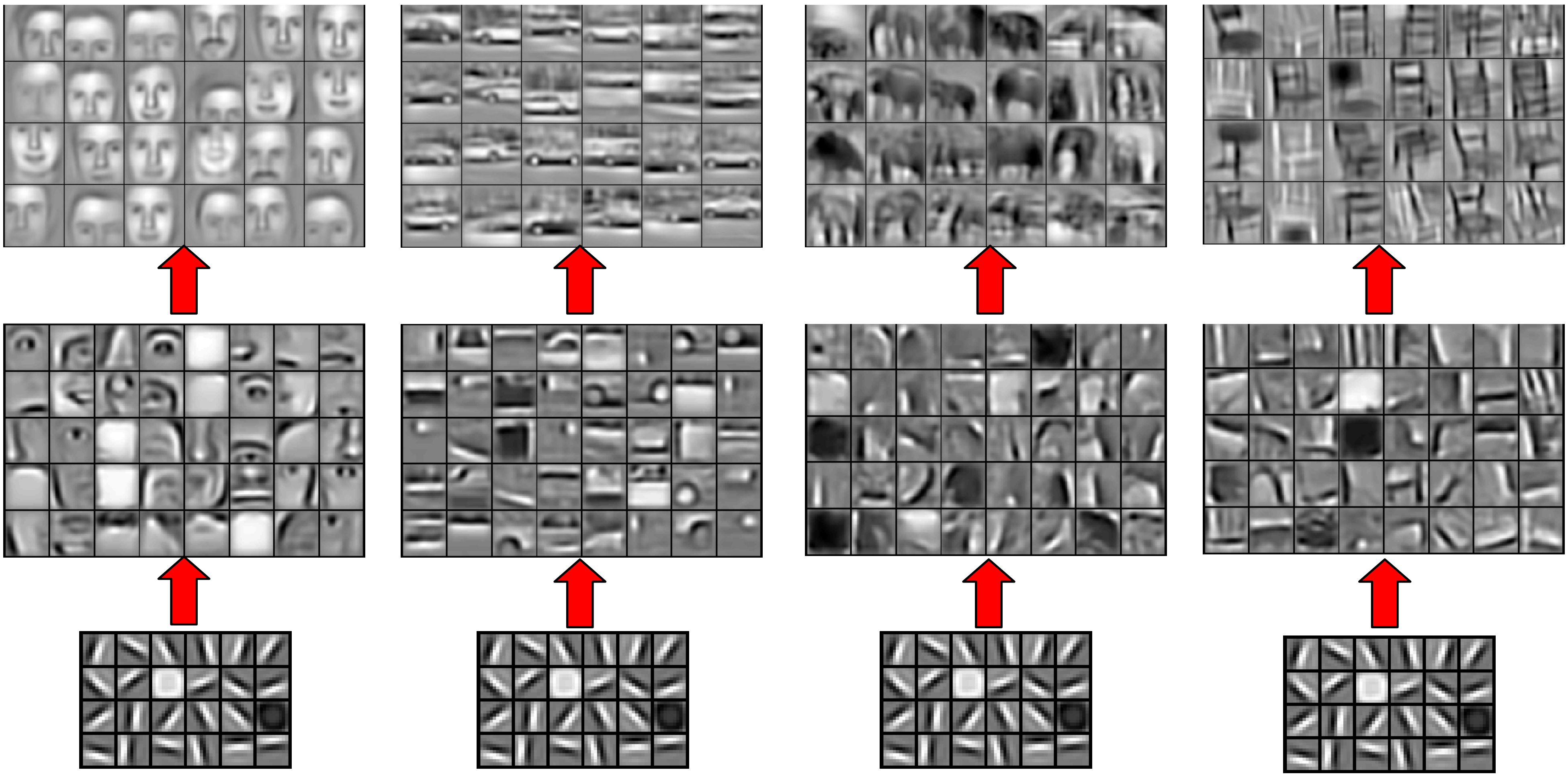


Deep Neural Networks on Face Images



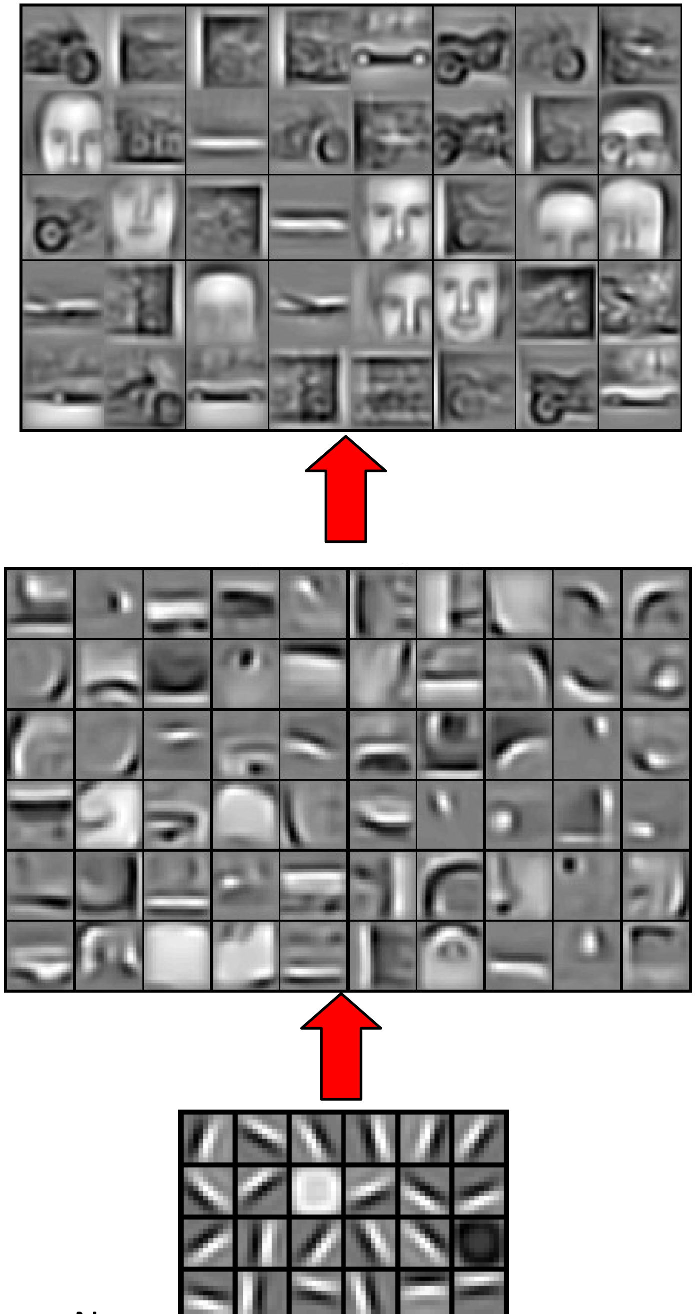
Based on materials
by Andrew Ng

Learning of Object Parts



Slide credit: Andrew Ng

Training on Multiple Objects



Trained on 4 classes (cars, faces, motorbikes, airplanes).

Second layer: Shared-features and object-specific features.

Third layer: More specific features.

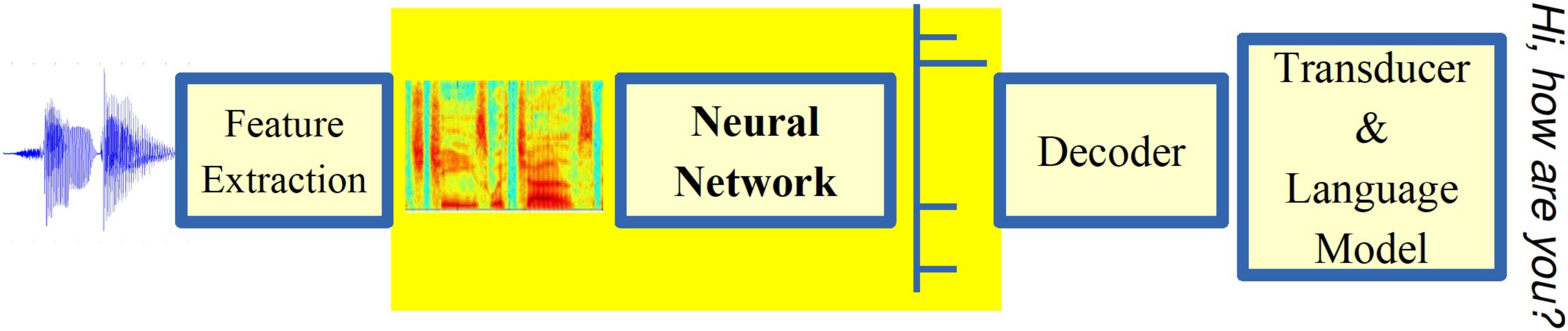
Slide credit: Andrew Ng

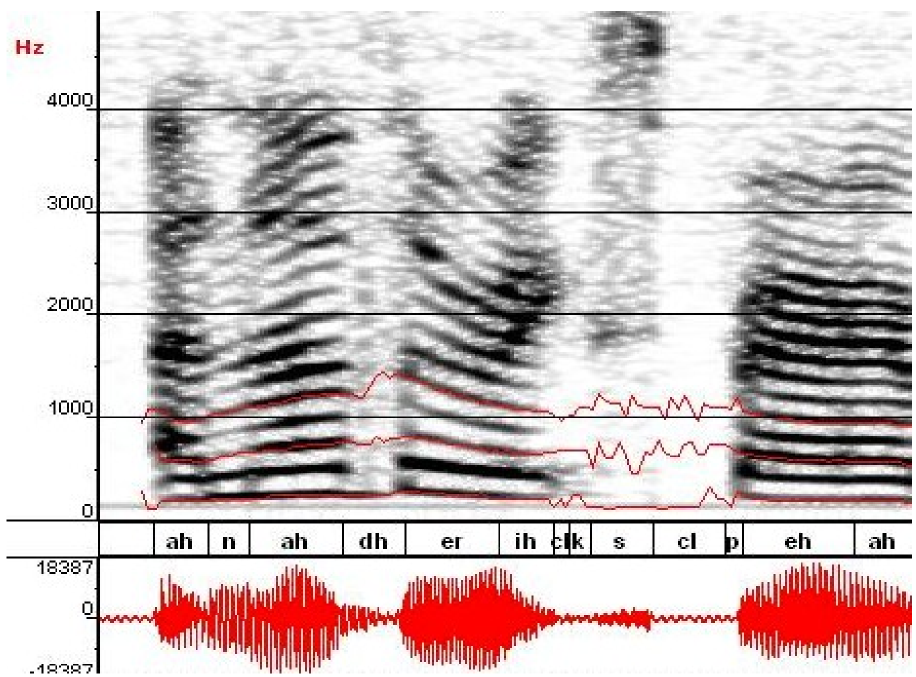
Scene Labeling via Deep Learning

[Farabet et al. ICML 2012, PAMI 2013]

Machine Learning in Automatic Speech Recognition

A Typical Speech Recognition System





Deep learning has state-of-the-art results

# Hidden Layers	1	2	4	8	10	12
Word Error Rate %	16.0	12.8	11.4	10.9	11.0	11.1

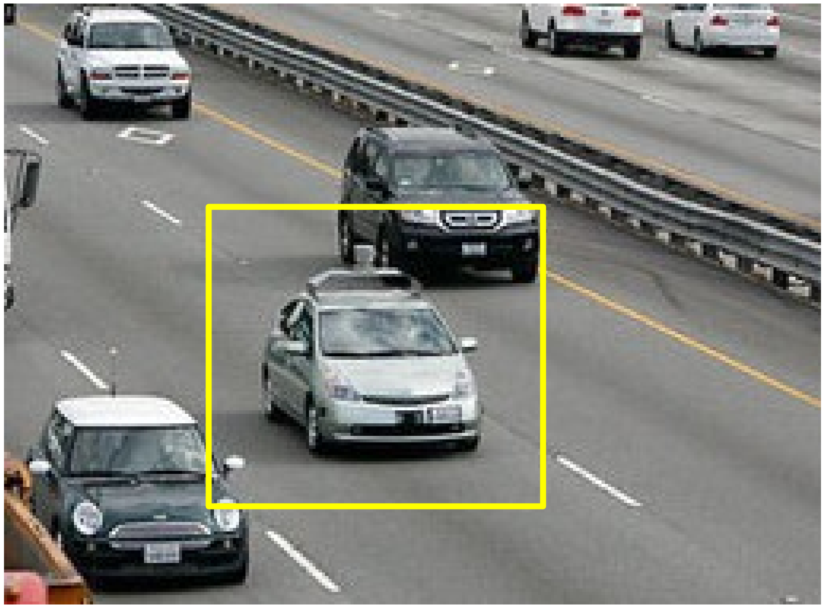
Baseline GMM performance = 15.4%

[Zeiler et al. “On rectified linear units for speech recognition” ICASSP 2013]

Impact of Deep Learning in Speech Technology

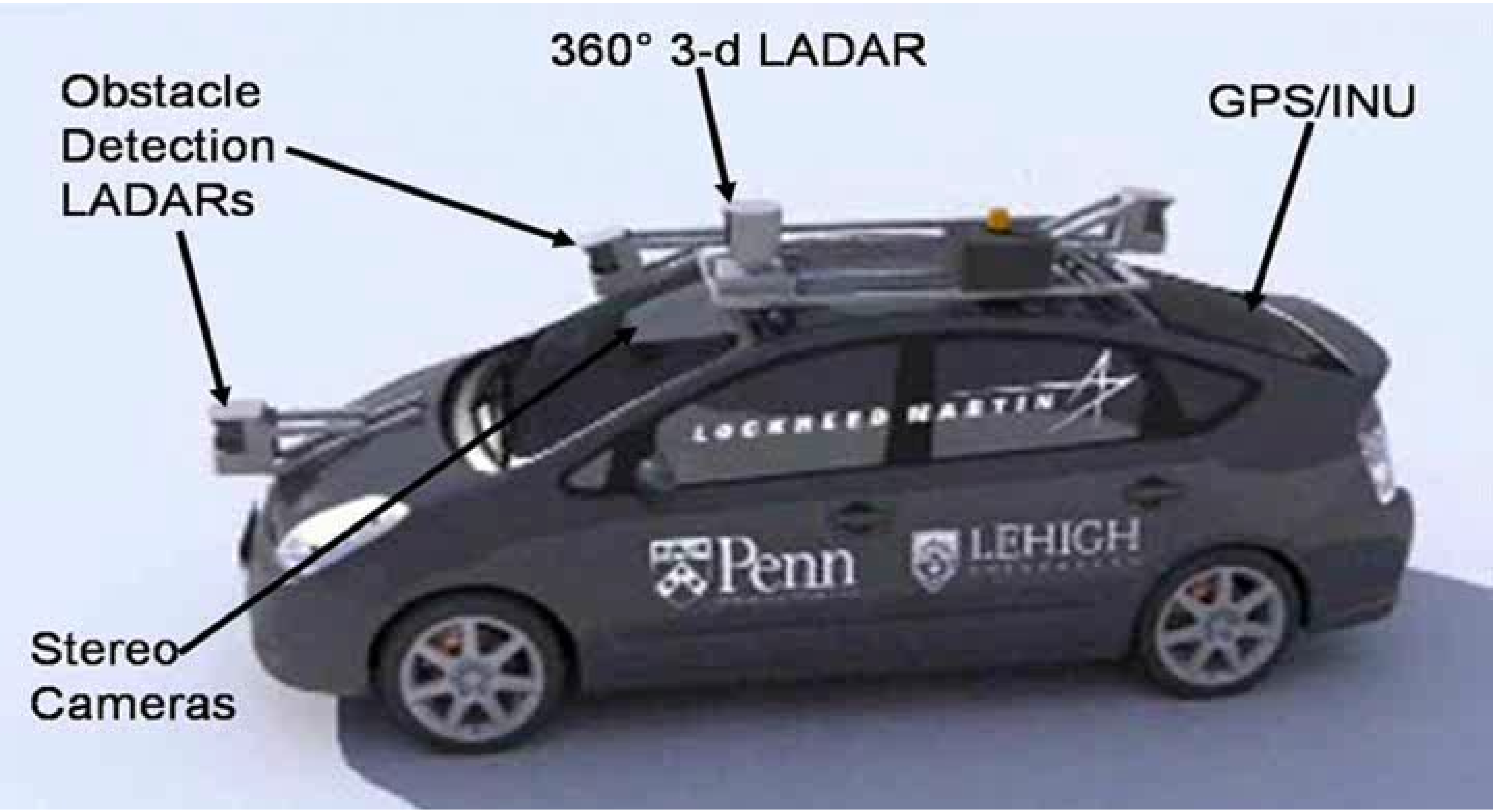
Slide credit: Li Deng, MS Research

Autonomous Cars

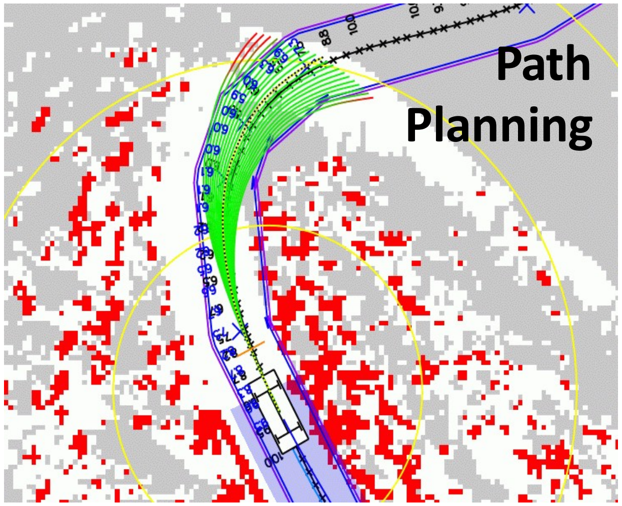


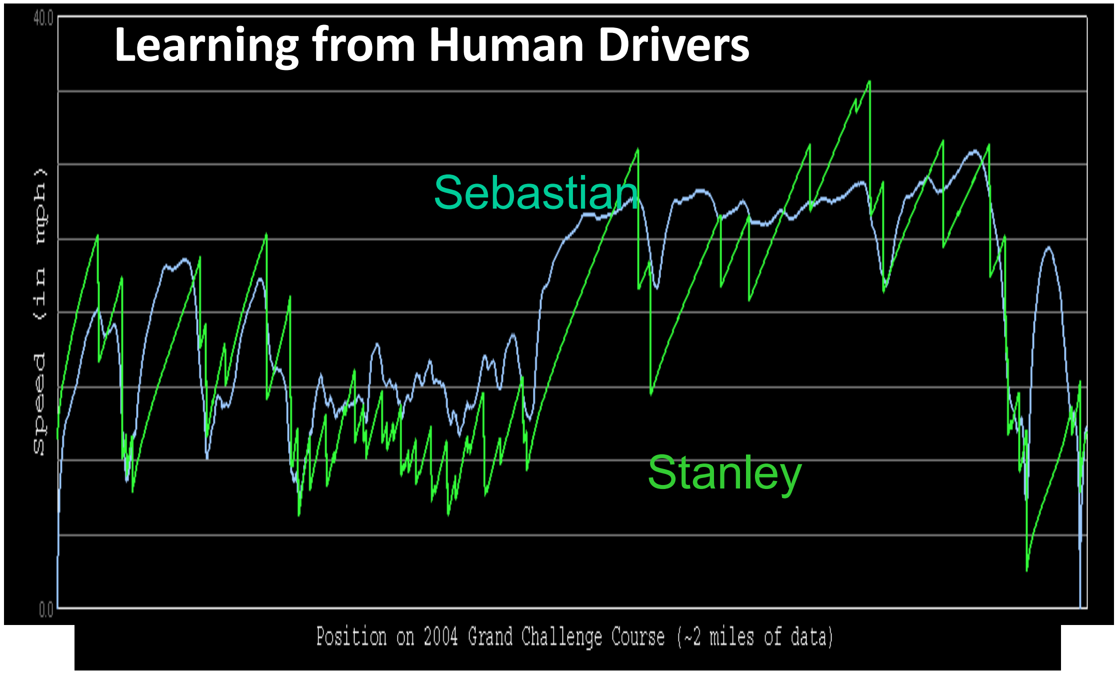
- Nevada made it legal for autonomous cars to drive on roads in June 2011
- As of 2013, four states (Nevada, Florida, California, and Michigan) have legalized autonomous cars

Autonomous Car Sensors



Autonomous Car Technology





The ML Approach

Data Collection

Model Selection

Parameter Estimation (The Learning Process)

Training

Inference
Find responses to queries

Inference
or
Testing

Learning Problem Definition

- Improving some measure of performance P when executing some task T through some type of training experience E
- Example: Learning to detect credit card fraud

- **Task** T
  – Assign label of fraud or not fraud to credit card transaction
- **Performance measure** P
  – Accuracy of fraud classifier

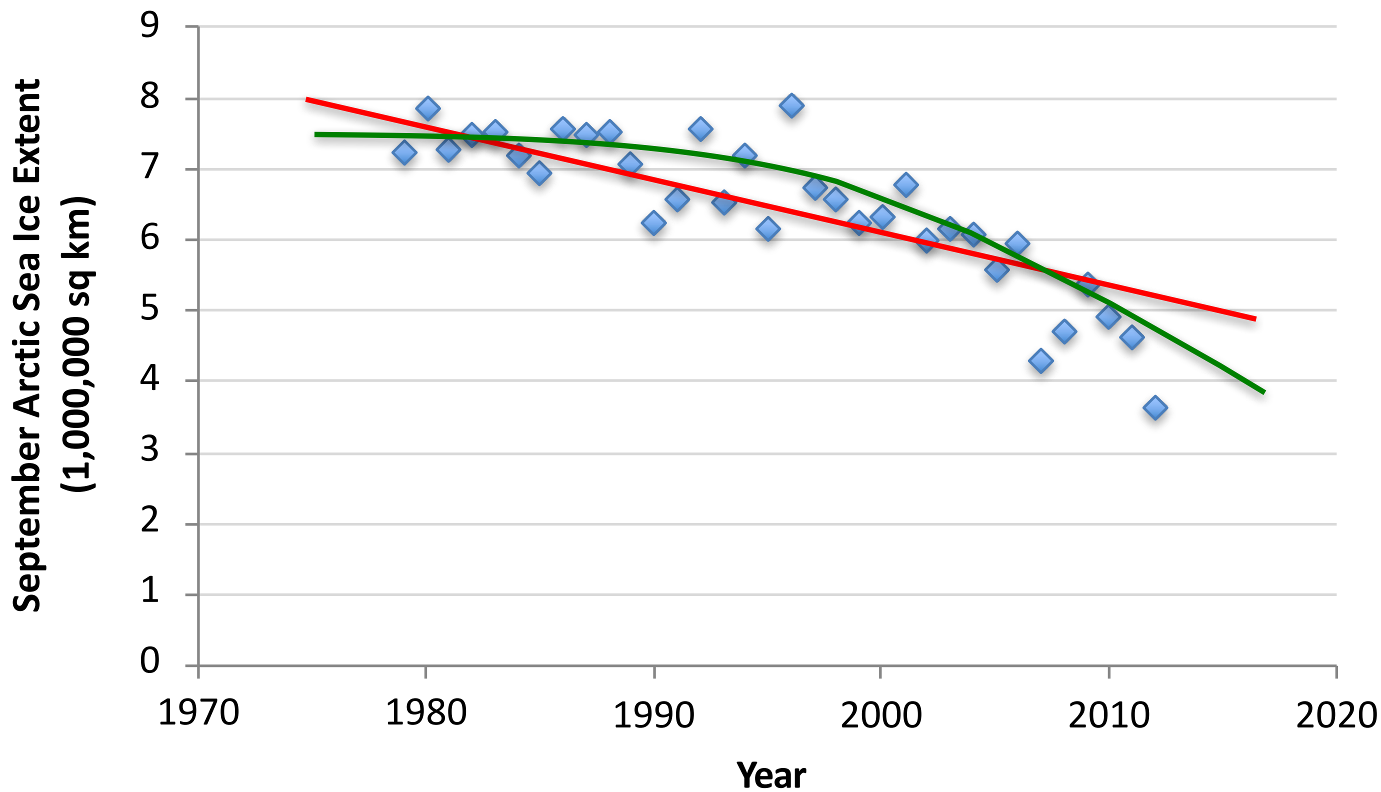
  With higher penalty when fraud is labeled as not fraud
- **Training experience** E
  – Historical credit card transactions labeled as fraud or not

Types of Learning

- **Supervised 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

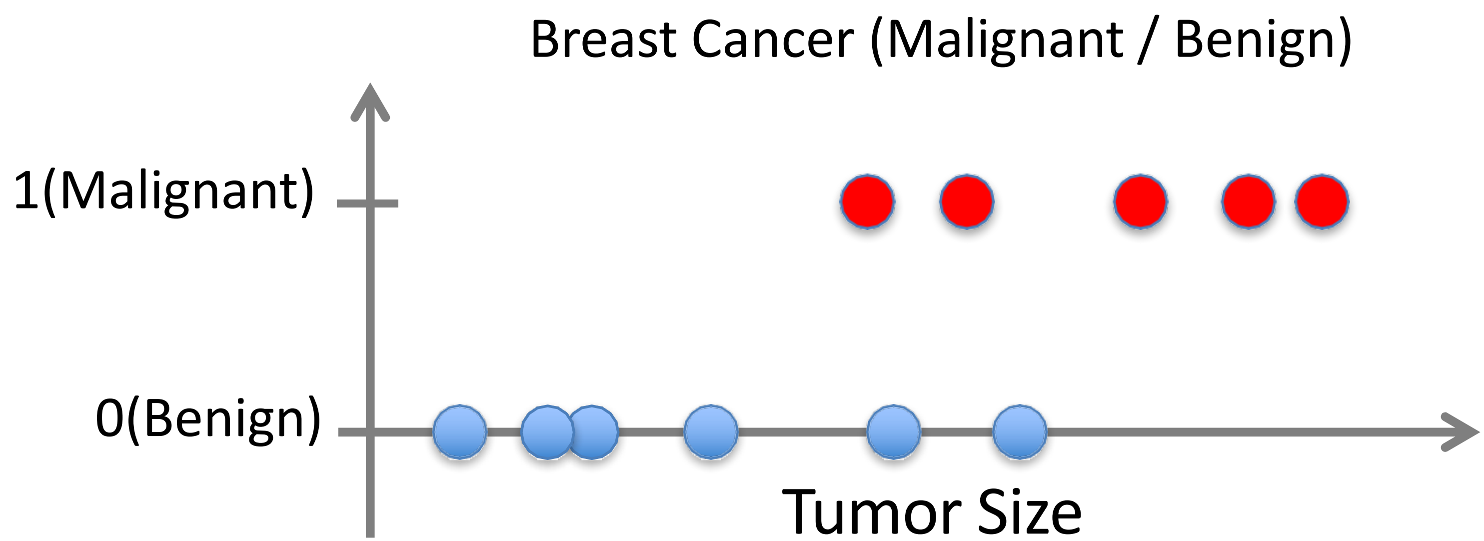
- Given $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$
- Learn a function $f(x)$ to predict y given x
  - y is real-valued == regression



Data from G. Witt. Journal of Statistics Education, Volume 21, Number 1 (2013)

Supervised Learning: Classification

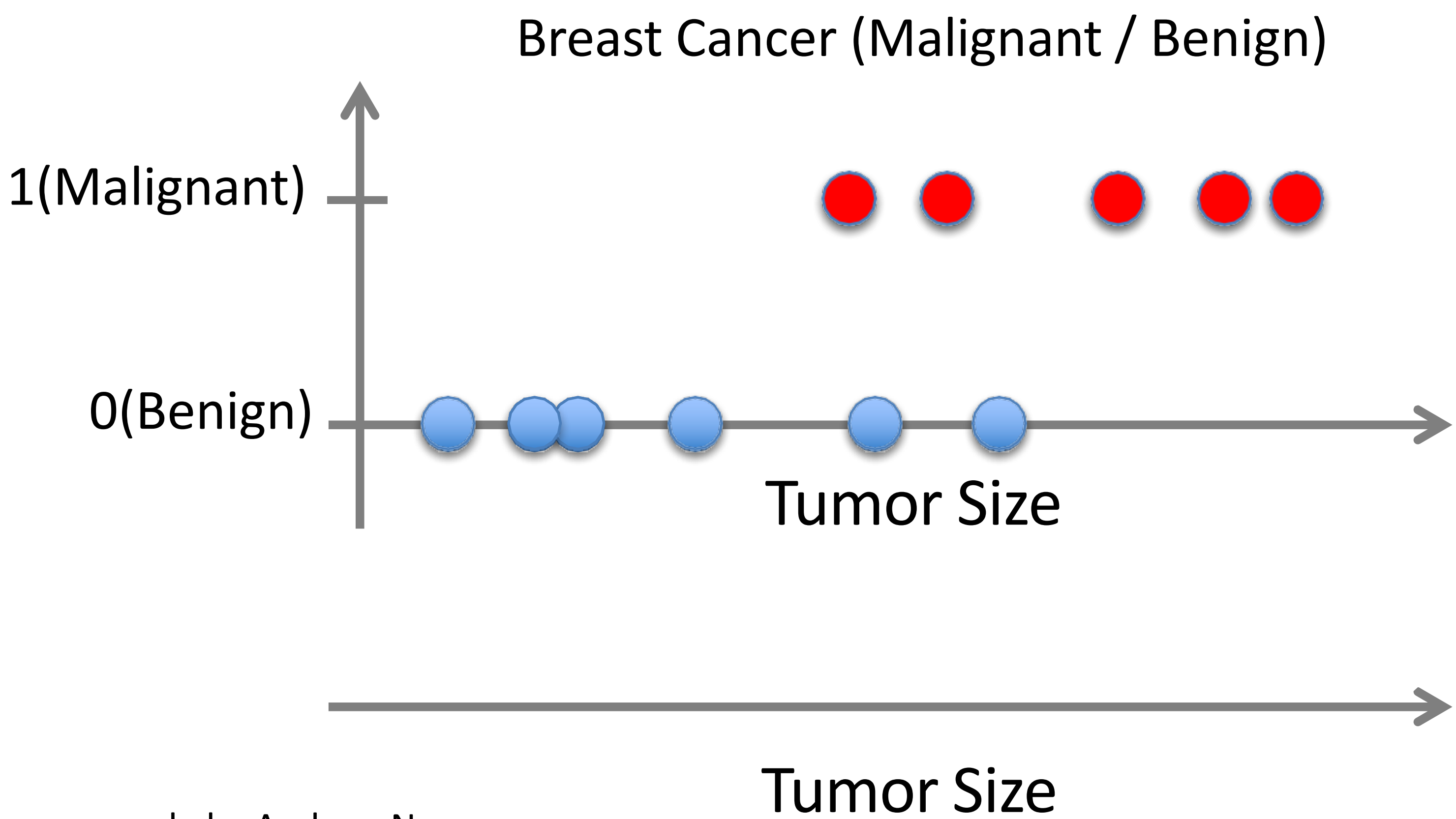
- Given $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$
- Learn a function $f(x)$ to predict y given x
  - y is categorical == classification



Based on example by Andrew Ng

Supervised Learning: Classification

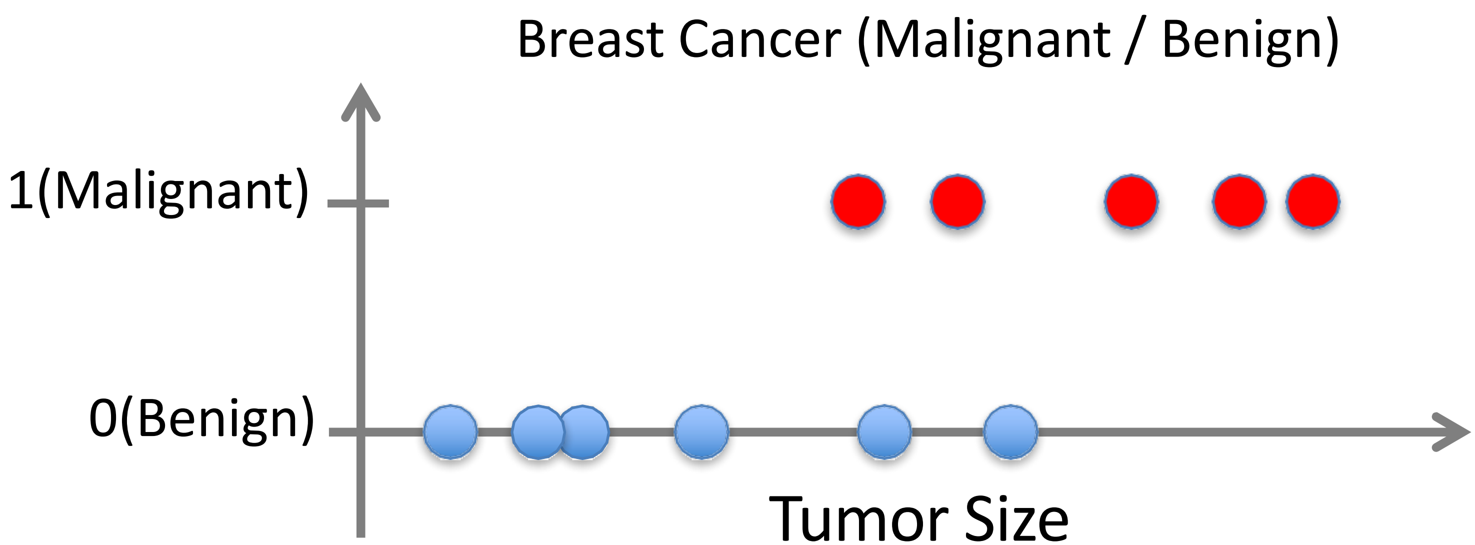
- Given $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$
- Learn a function $f(x)$ to predict y given x
  - y is categorical == classification

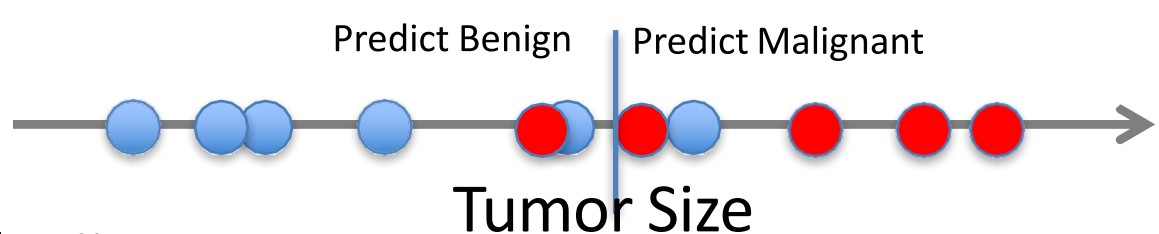


Based on example by Andrew Ng

Supervised Learning: Classification

- 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 categorical == classification

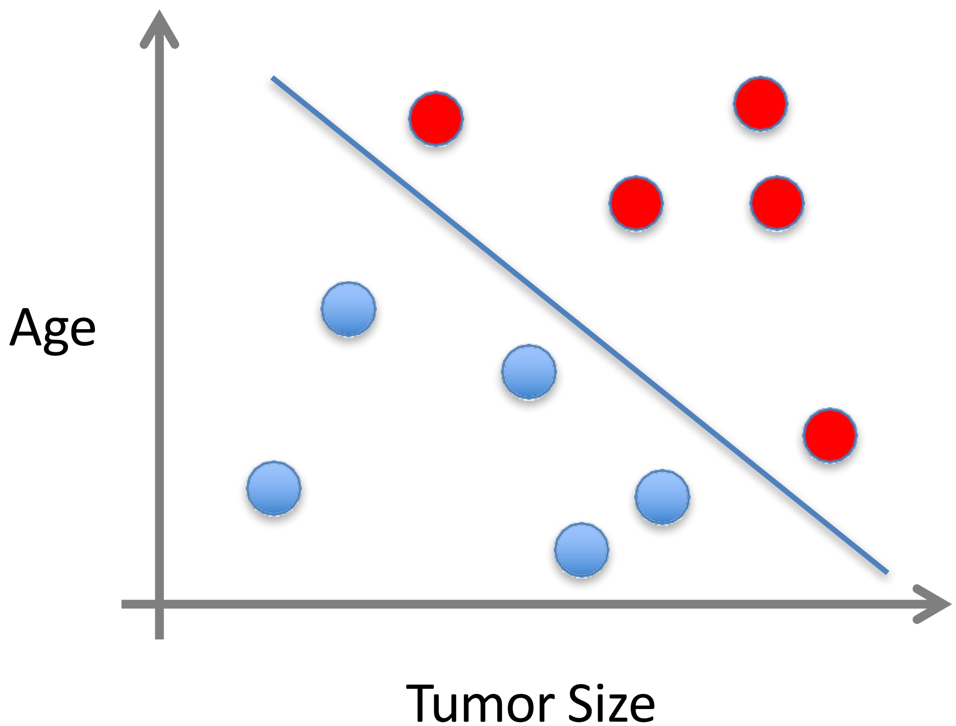




Based on example by Andrew Ng

Supervised Learning

- x can be multi-dimensional
  - Each dimension corresponds to an attribute



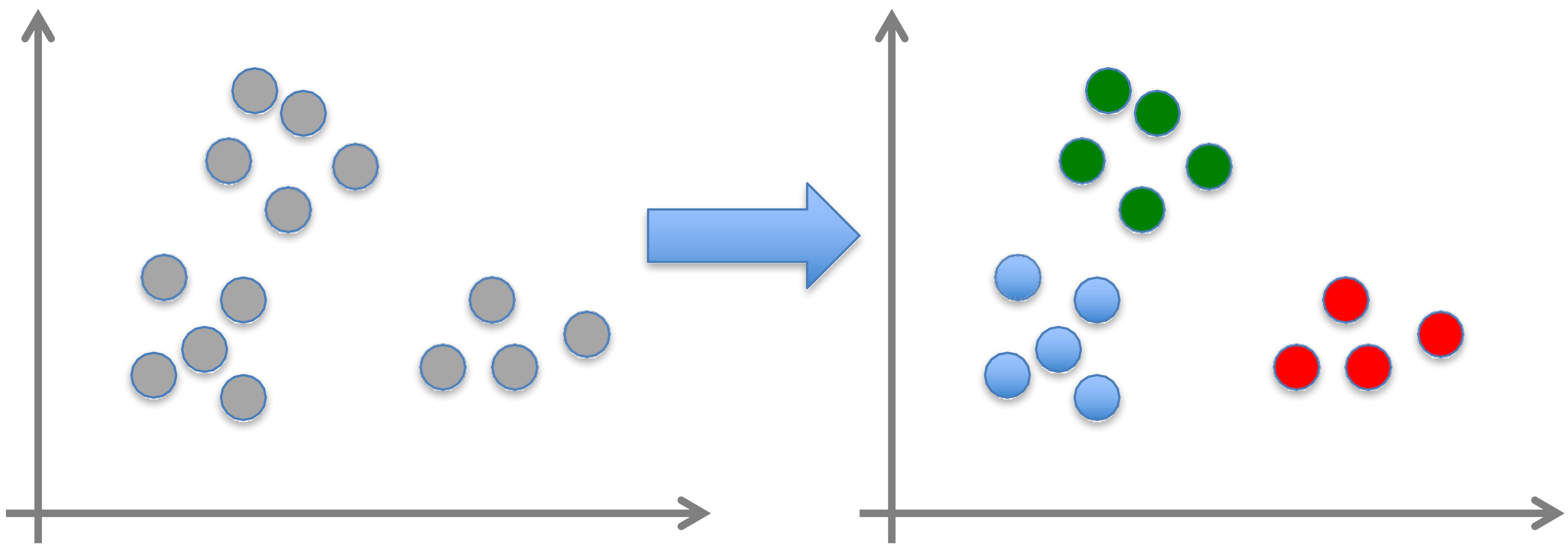
- Clump Thickness
- Uniformity of Cell Size
- Uniformity of Cell Shape

…

Based on example by Andrew Ng

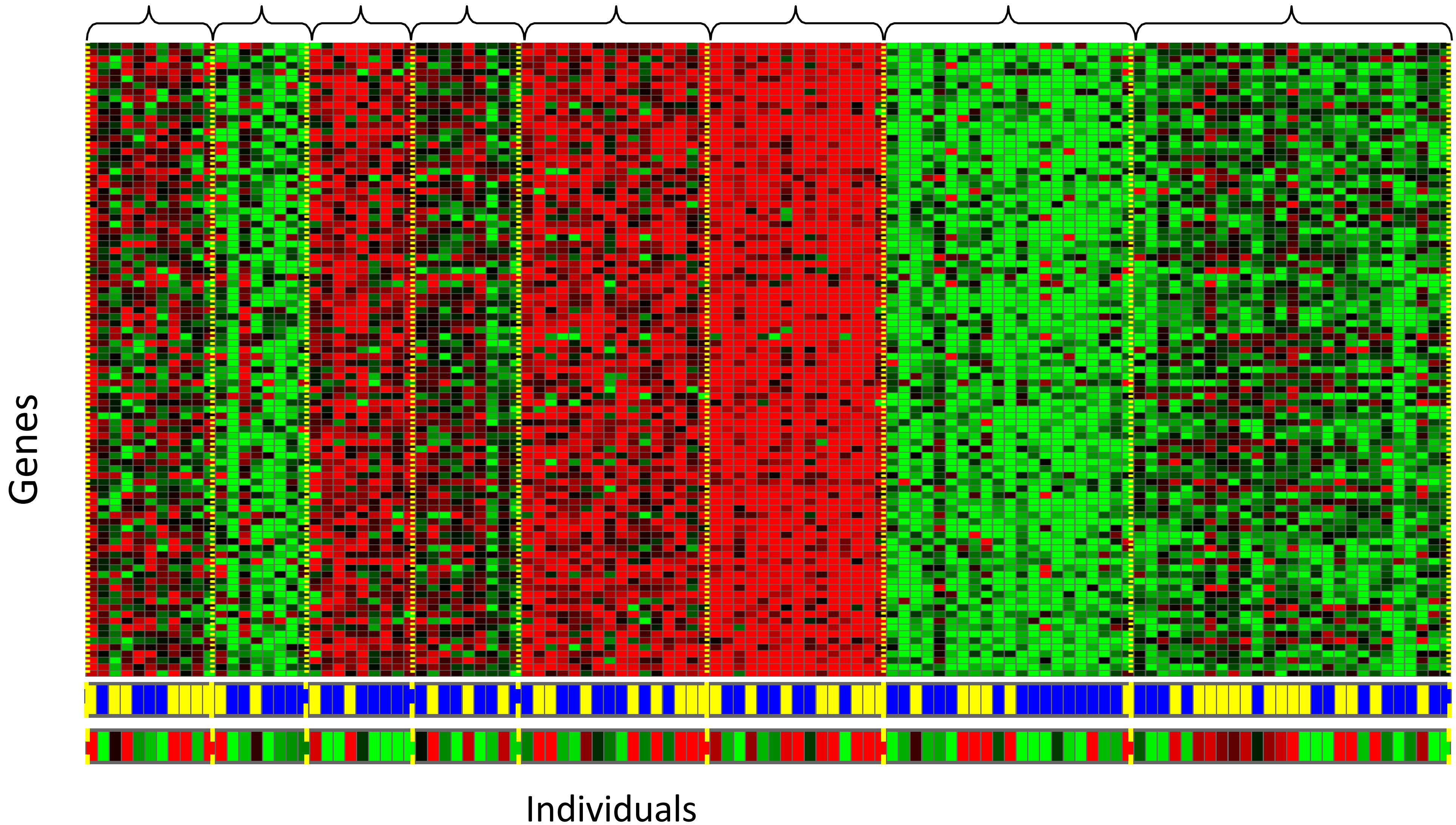
Unsupervised Learning

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



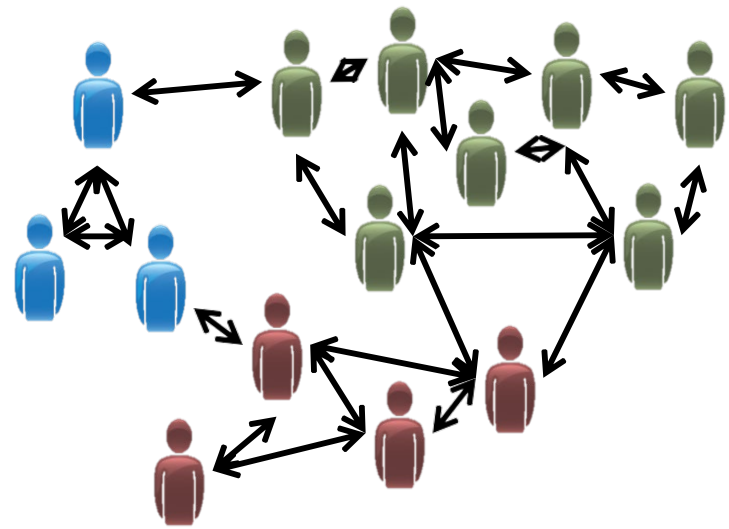
Unsupervised Learning

Genomics application: group individuals by genetic similarity



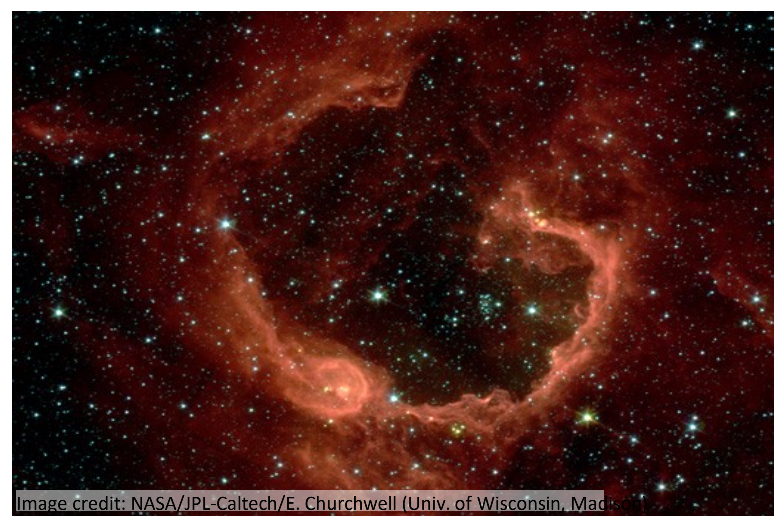
[Source: Daphne Koller]

Unsupervised Learning



Social network analysis

Market segmentation

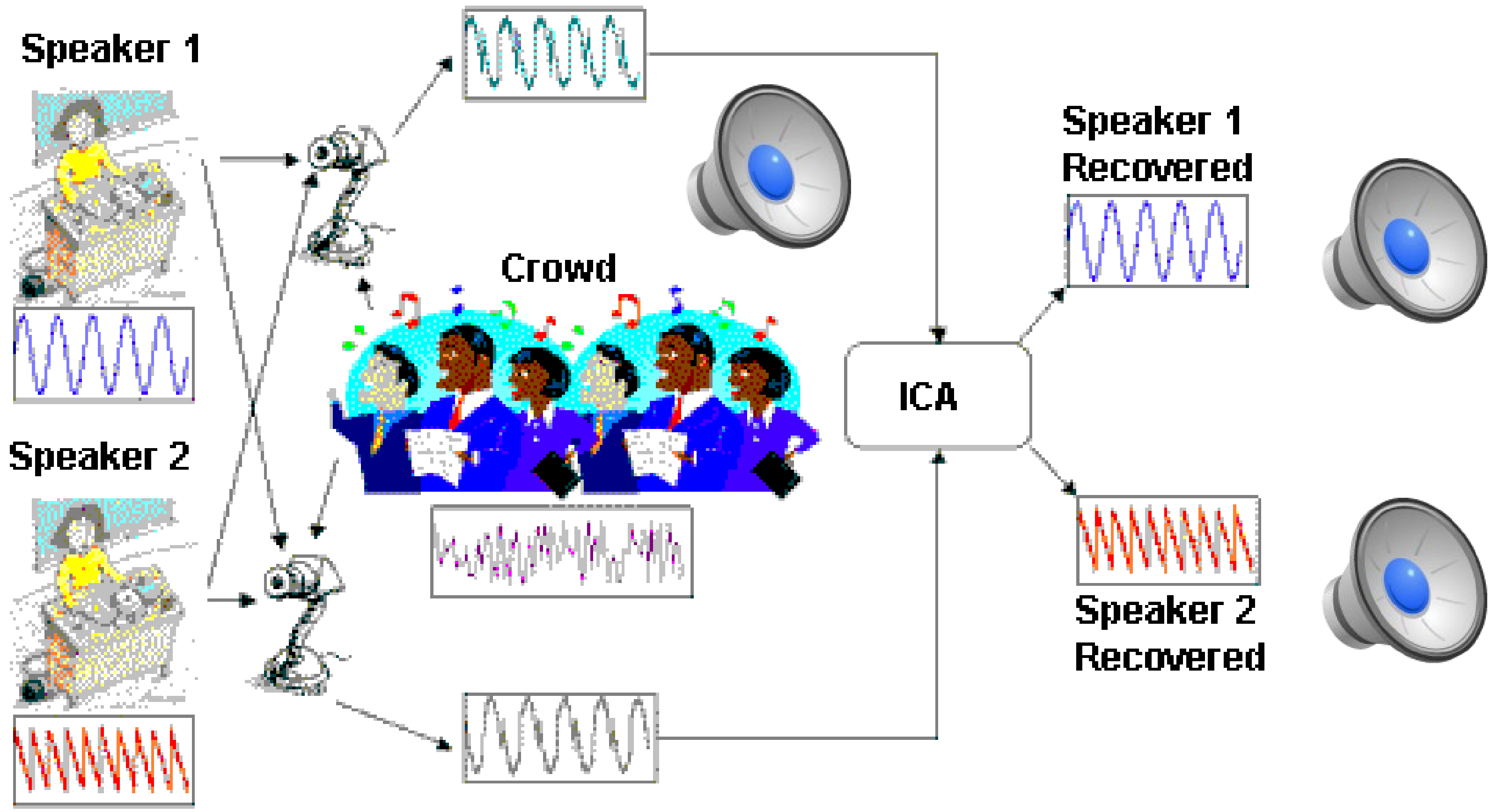


Astronomical data analysis

Slide credit: Andrew Ng

Unsupervised Learning

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



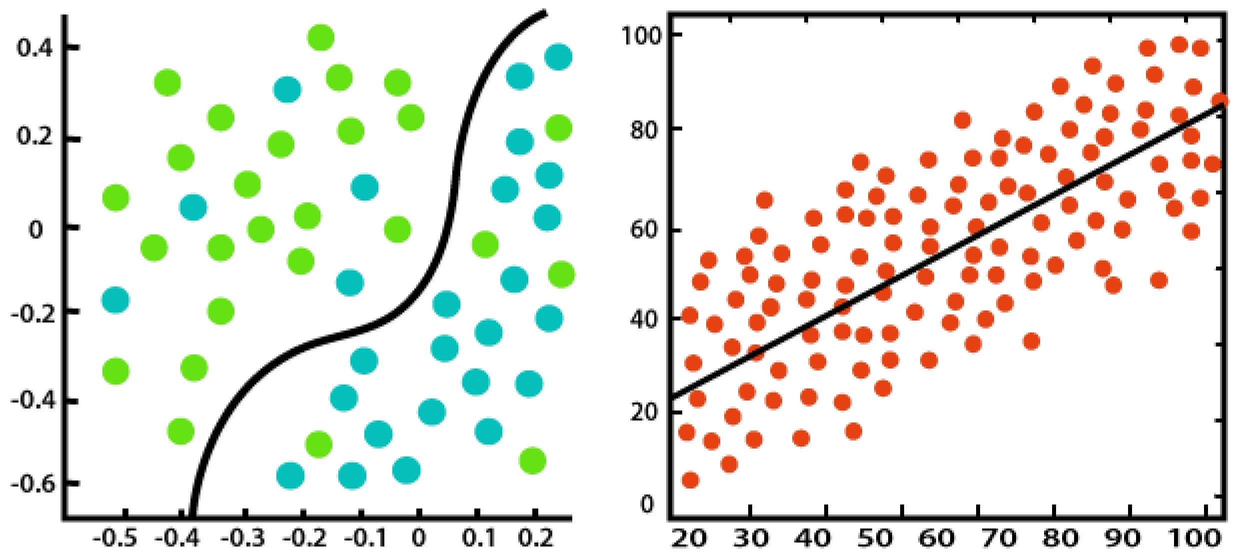
Reinforcement Learning

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

Reinforcement Learning

https://www.youtube.com/watch?v=4cgWya-wjgY

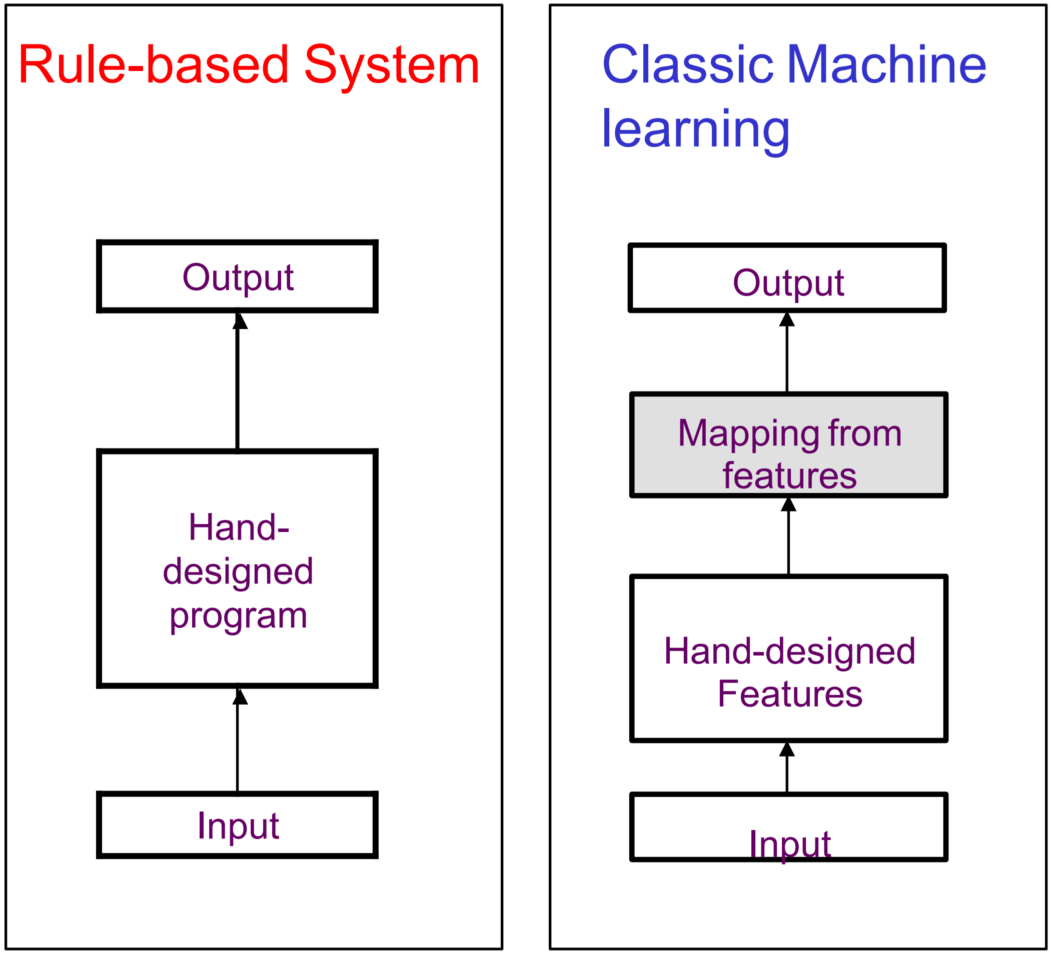
Classification vs Regression



Classification

Regression

Two paradigms in AI

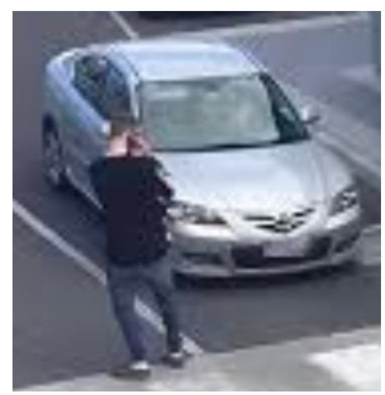


Shaded boxes indicate components that can learn from data

Designing right set of features

- Simple Machine Learning depends heavily on *representation* of given data
- For detecting a car in photographs
  - Tire shape difficult in terms of pixel values
    - Shadows, occlusion, similar background



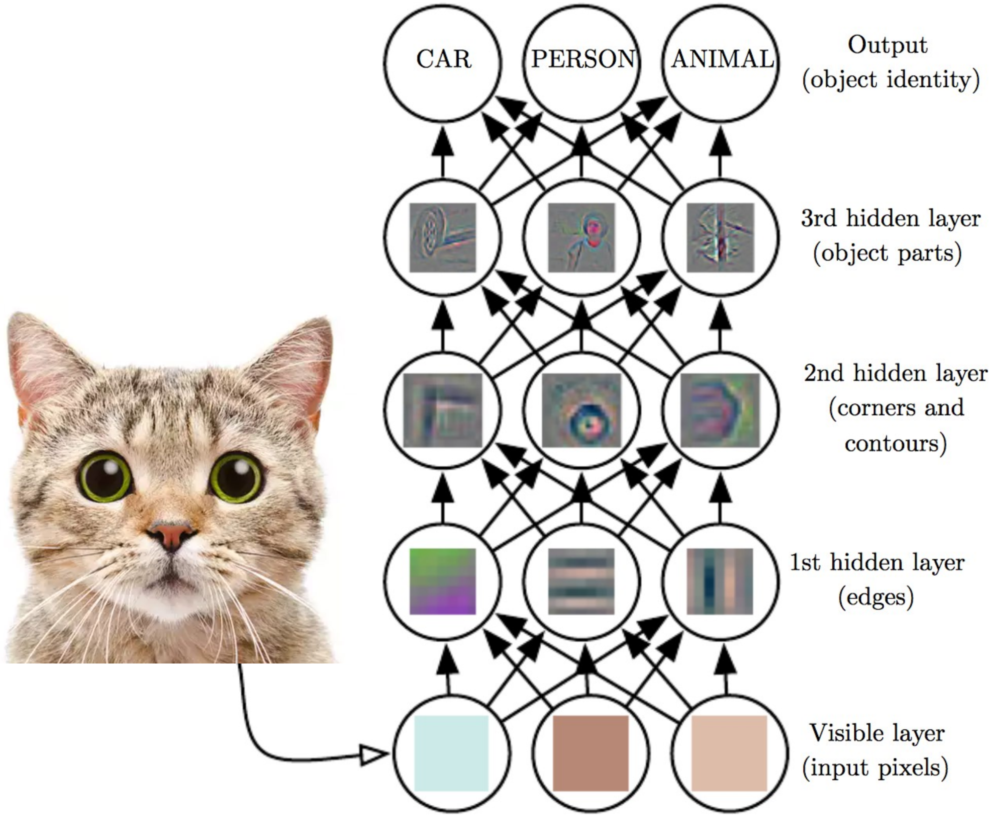


Representation Learning

- Solution: use ML to not only learn mapping from representation to output but representation itself
  - Better results than hand-coded representations
- Allows AI systems to rapidly adapt to new tasks
  - Designing features can take great human effort
    - Can take decades for a community of researchers
- Does not need programmer to have deep knowledge of the problem domain

Feature Learning for Classification

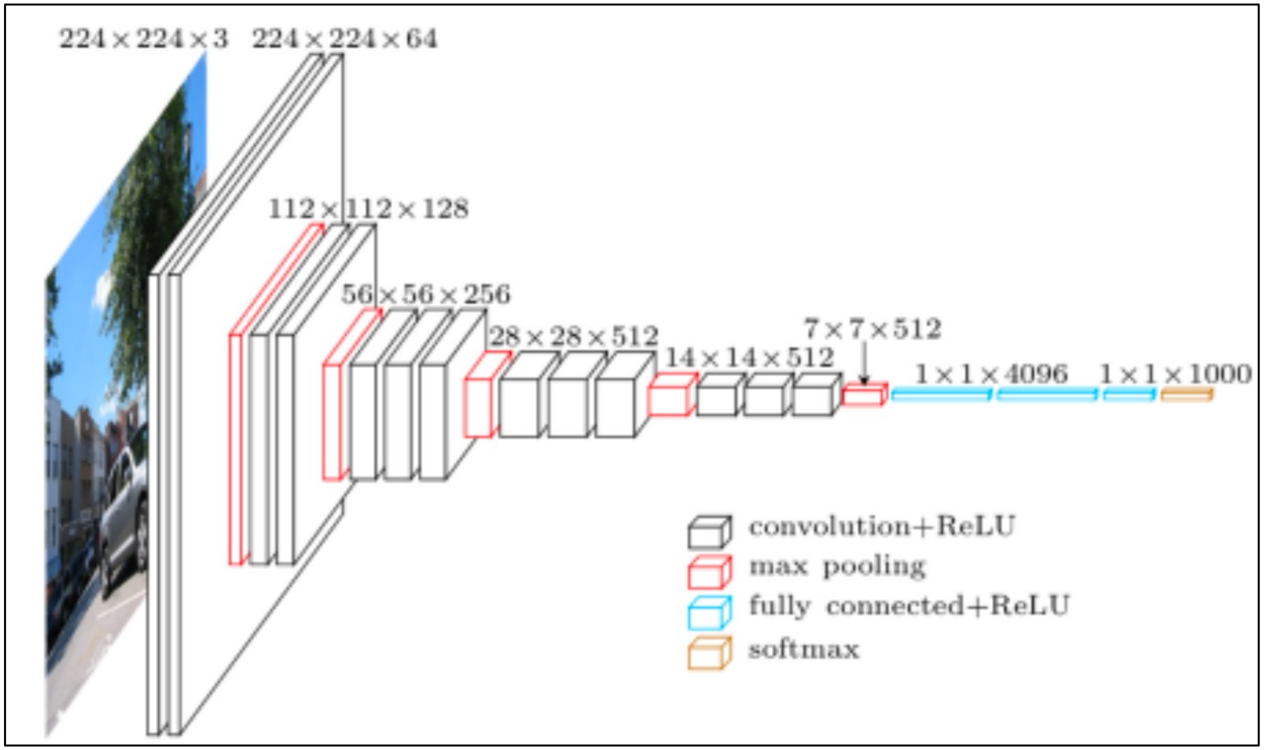
- Function to map pixels to object identity is complicated
- Series of hidden layers extract increasingly abstract features
- Final decision made by a simple classifier

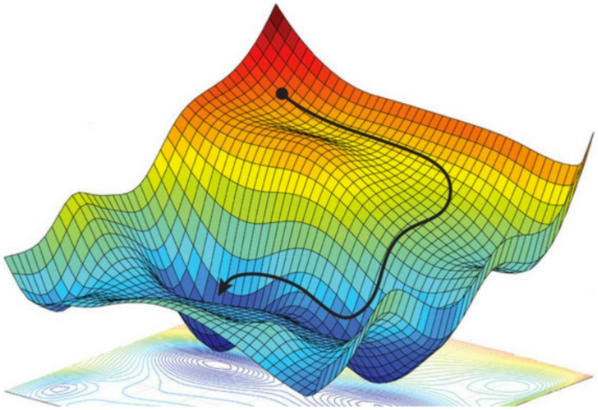


Deep Learning

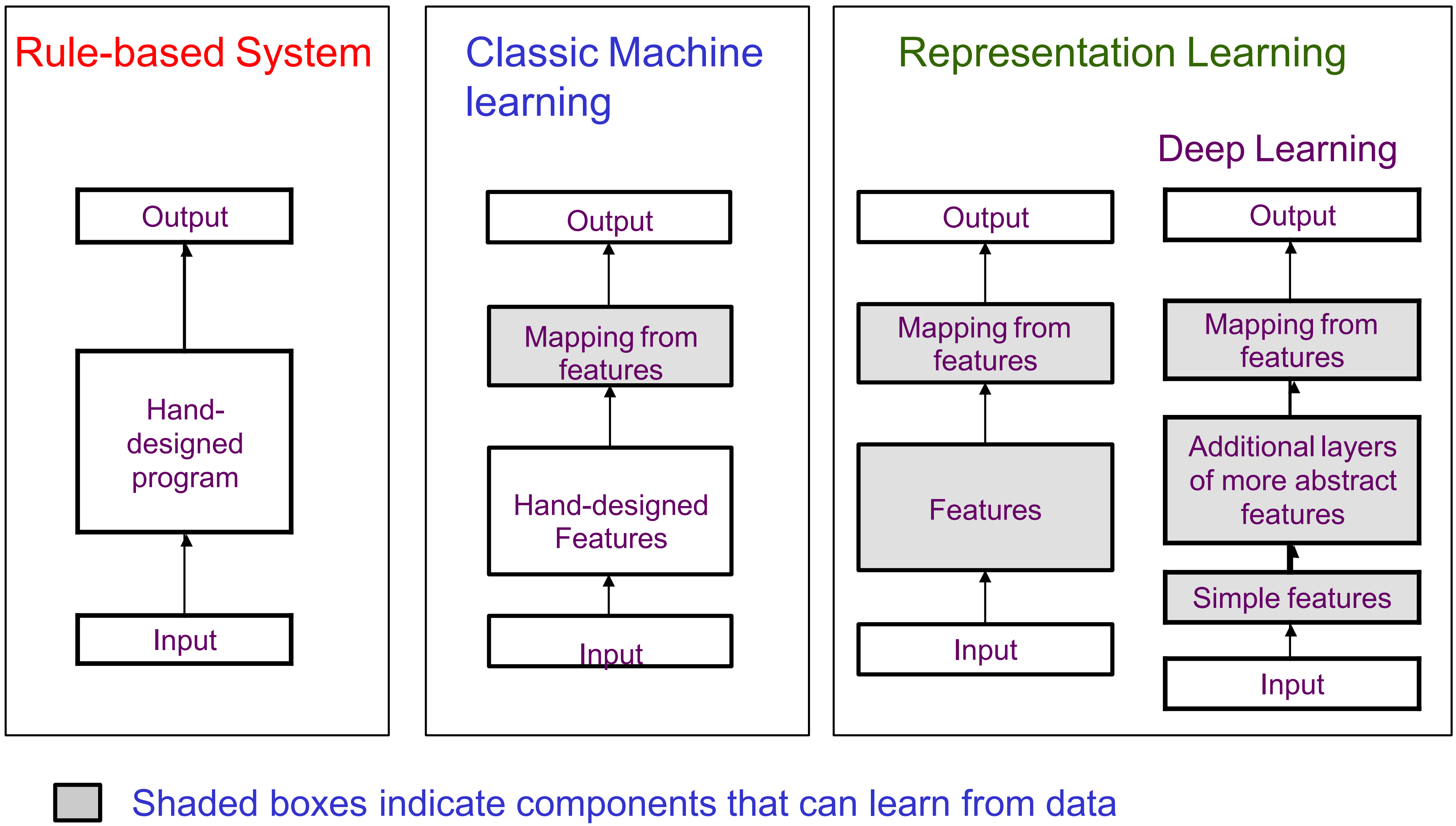
- Understand the world as hierarchy of concepts
  - How these concepts are built on top of each other is deep, with many layers
  - Weights learnt by gradient descent

$$x^{t+1} = x^t - \varepsilon \nabla_{x^t} f(x^t)$$





Summary of AI Models



AI Paradigm Shift

- Physics paradigm shift
  - Newtonian Physics
    - Cannot explain black-body radiation
  - Quantum Mechanics
- AI paradigm shift
  - Knowledge-based systems
    - Cannot perform simple recognition tasks
  - Simple machine learning methods
    - Cannot perform complex recognition tasks
  - Deep Learning methods

Summary

- Machine Learning as an AI approach
  - Overcomes limitations of knowledge-based systems
- Types of AI tasks
  - Data: Supervised, Unsupervised, Semi-supervised, Reinforcement
  - Output: Classification, Regression, Ranking