CS 412: INTRODUCTION TO MACHINE LEARNING

Prof. Xinhua Zhang

Course Goals

Enable you to **build useful machine learning systems** for new (and old) problem settings.

- Learn about key ML problems/paradigms
  - From simple approaches to state-of-the-art
- Develop algorithms for reasoning about data
  - Probability & statistics, optimization
- Gain experience with real data

Prerequisites

Formal: CS251, STAT 381/IE 342

Less formal:

Comfort with "abstract mathematical concepts"

(discrete math, calculus, probability & statistics, matrix algebra)

Computer science/programming experience

(data structures, dynamic programming, e.g., C++, Java, Python)

We will use Python

LECTURE 1: WHAT IS MACHINE LEARNING?

Prof. Xinhua Zhang

Big Data

- Widespread use of personal computers and wireless communication leads to “big data”
- We are both producers and consumers of data
  - Customize/personalize
- Data is not random, it has structure, e.g., customer behavior
- We need “big theory” to extract that structure from data for

  (a) Understanding the process (descriptive)

  (b) Making predictions for the future (predictive)

Why "Learn" ?

- Machine learning is programming computers to optimize a performance criterion using example data or past experience.
- There is no need to "learn" to calculate payroll
- Reverse parking?
- Learning is used when:
  - Human expertise does not exist (navigating on Mars),
  - Humans are unable to explain their expertise (speech recognition)
  - Solution changes in time (routing on a computer network)
  - Solution needs to be adapted to particular cases (user biometrics)

What We Talk About When We Talk About "Learning"

- Learning general models from a data of particular examples
- Data is cheap and abundant (data warehouses, data marts); knowledge is expensive and scarce.
- Example in retail: Customer transactions to consumer behavior:

  People who bought "Blink" also bought "Outliers" (www.amazon.com)

- Build a model that is *a good and useful approximation* to the data.

Huge Implications

"If you invent a breakthrough in artificial intelligence, so machines can learn, that is worth 10 Microsofts."

Bill Gates

Data Mining

- 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
- …

What is Machine Learning?

- Optimize a performance criterion using example data or past experience.
- Role of Statistics: Inference from a sample
- Role of Computer science: Efficient algorithms to
  - Solve the optimization problem
  - Representing and evaluating the model for inference

Applications

- Association
- Supervised Learning
  - Classification
  - Regression
- Unsupervised Learning
- Reinforcement Learning

Learning Associations

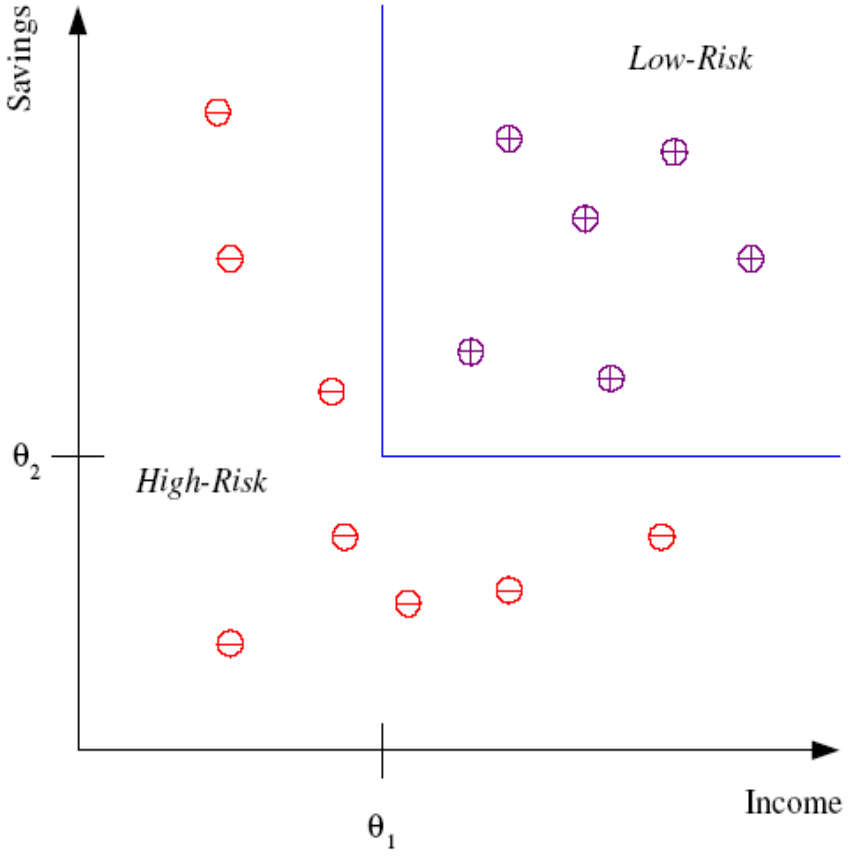
- Basket analysis:

  $P(Y \mid X)$ probability that somebody who buys X also buys Y where X and Y are products/services.

  Example: $P(\text{chips} \mid \text{beer}) = 0.7$

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
- Text Classification:

spam or not spam

Classification: Applications

- Face recognition: Pose, lighting, occlusion (glasses, beard), make-up, hair style

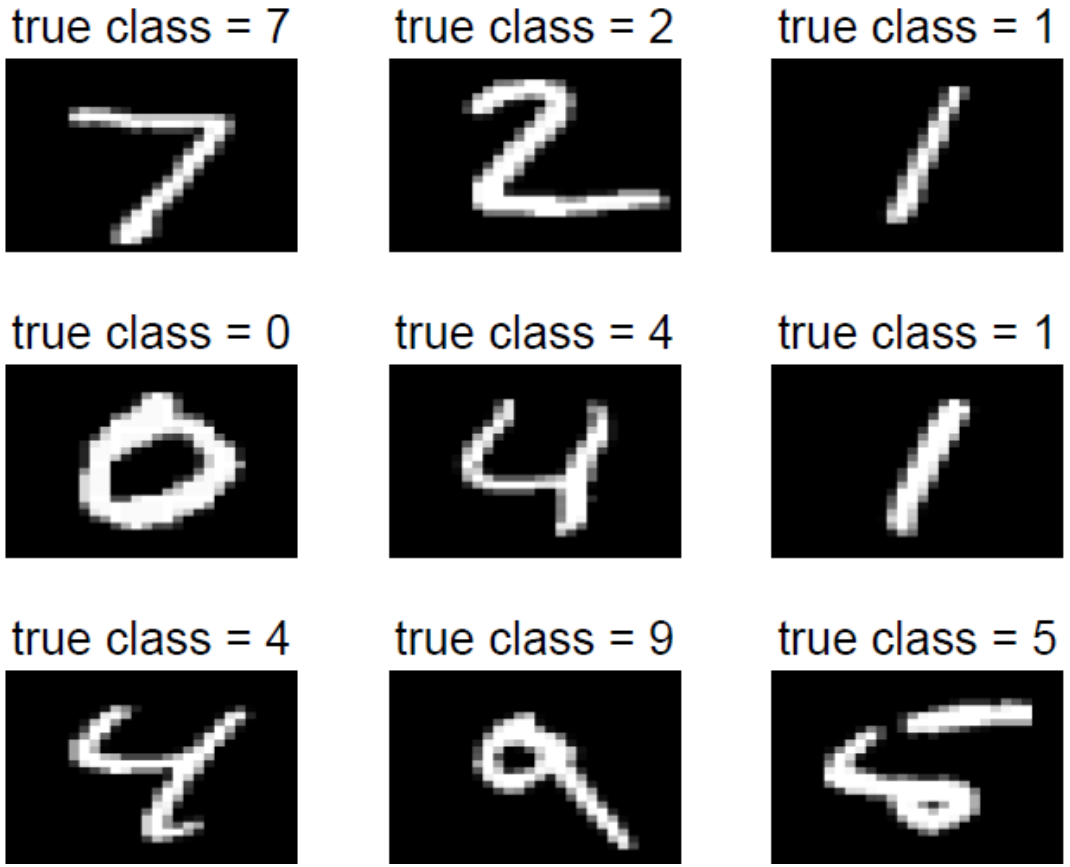
Training examples of a person

Test images

Classification: Applications

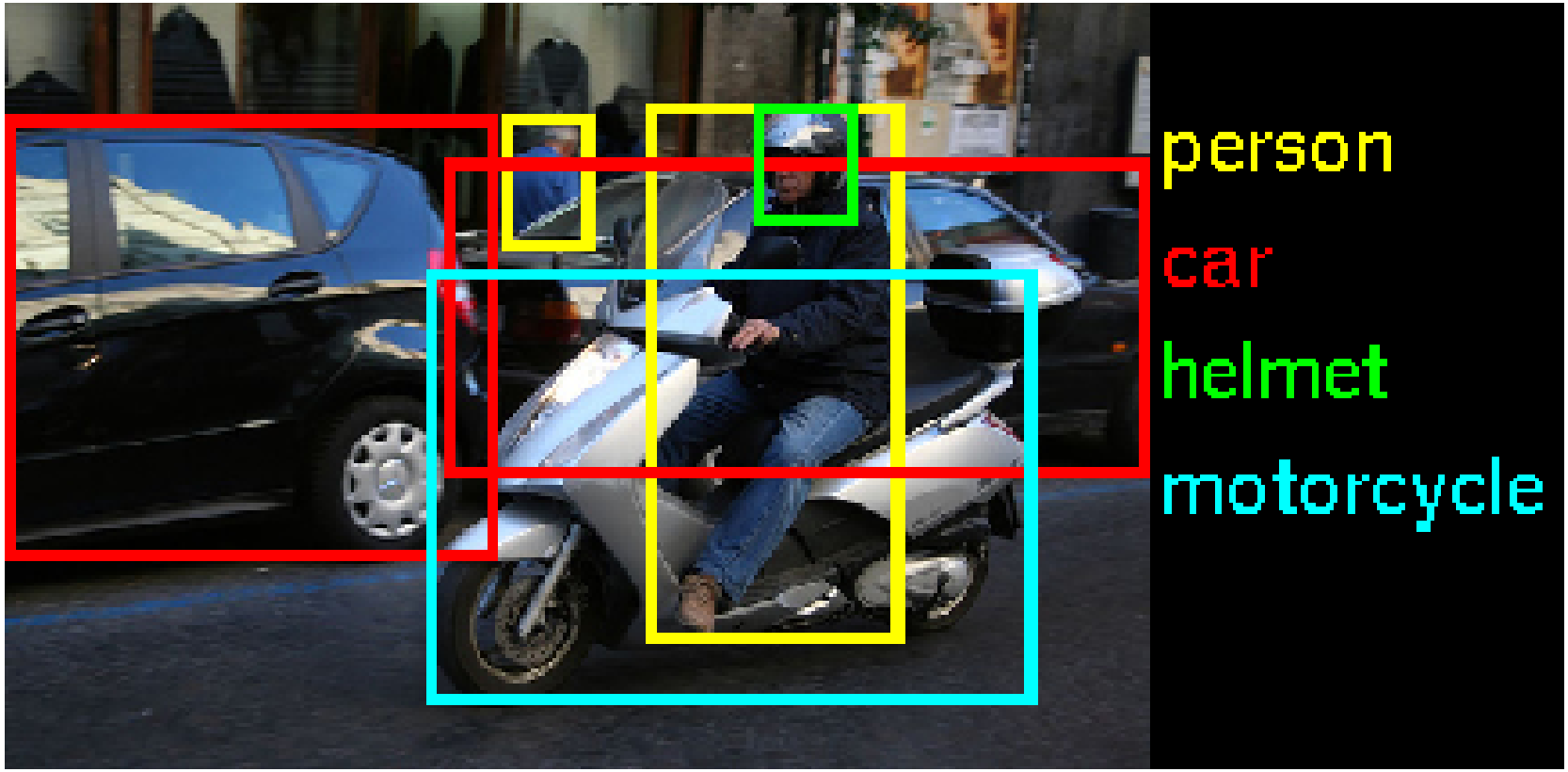
- Aka Pattern recognition
- Character recognition:

Different handwriting styles



Classification: Applications

- Aka Pattern recognition
- Object recognition:



Classification: Applications

- Aka Pattern recognition
- Recommendation systems:

What other movies?

Example: Recommendation Systems

NETFLIX

Netflix Prize

Home | Rules | Leaderboard | Register | Update | Sub

Leaderboard

Rank	Team Name	Best Score	% Improvement	Last Submit Time
1	BellKor's Pragmatic Chaos	0.8558	10.05	2009-06-26 18:42:37
Grand Prize - RMSE <= 0.8563				
2	PragmaticTheory	0.8582	9.80	2009-06-25 22:15:51
3	BellKor in BigChaos	0.8590	9.71	2009-05-13 08:14:09
4	Grand Prize Team	0.8593	9.68	2009-06-12 08:20:24
5	Dace	0.8604	9.56	2009-04-22 05:57:03
6	BigChaos	0.8613	9.47	2009-06-23 23:06:52
Progress Prize 2008 - RMSE = 0.8616 - Winning Team: BellKor in BigChaos				
7	BellKor	0.8620	9.40	2009-06-24 07:16:02
8	xlvector	0.8630	9.29	2009-06-27 14:08:39
9	Gravity	0.8634	9.25	2009-04-22 18:31:32
10	Opera Solutions	0.8638	9.21	2009-06-26 23:18:13
11	BruceDengDaoCiYiYou	0.8638	9.21	2009-06-27 00:55:55
12	pengpengzhou	0.8638	9.21	2009-06-27 01:06:43
13	Feeds2	0.8641	9.18	2009-06-26 22:51:55

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
- Outlier/novelty detection:

Regression

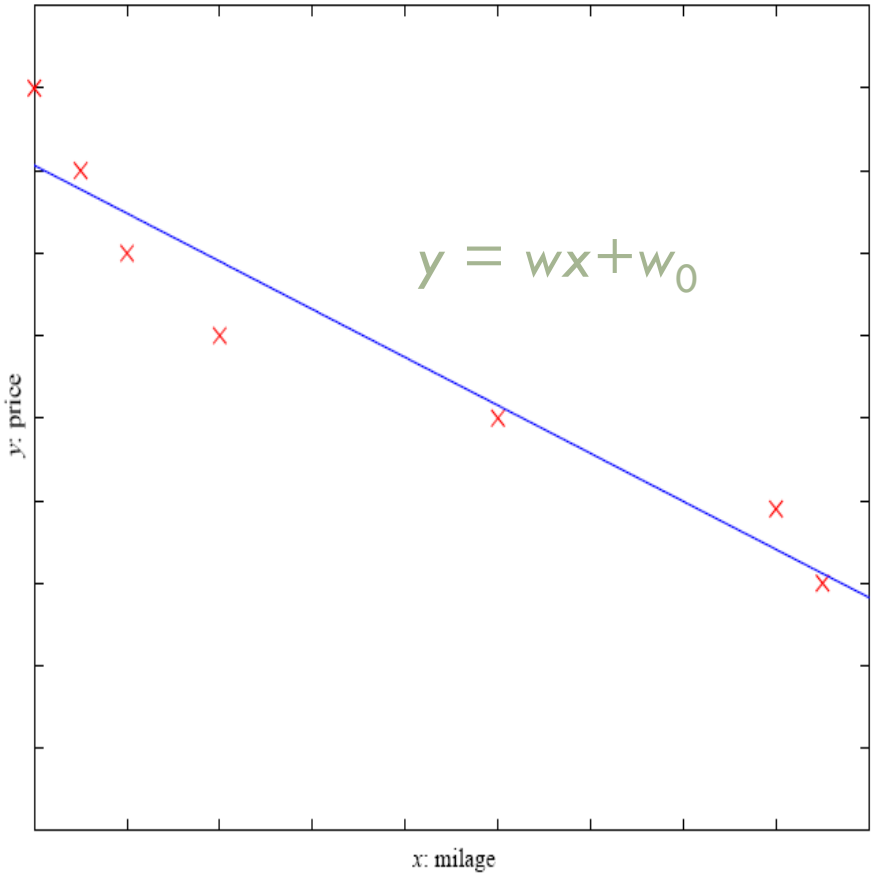
- Example: Price of a used car
- x : car attributes

  y : price

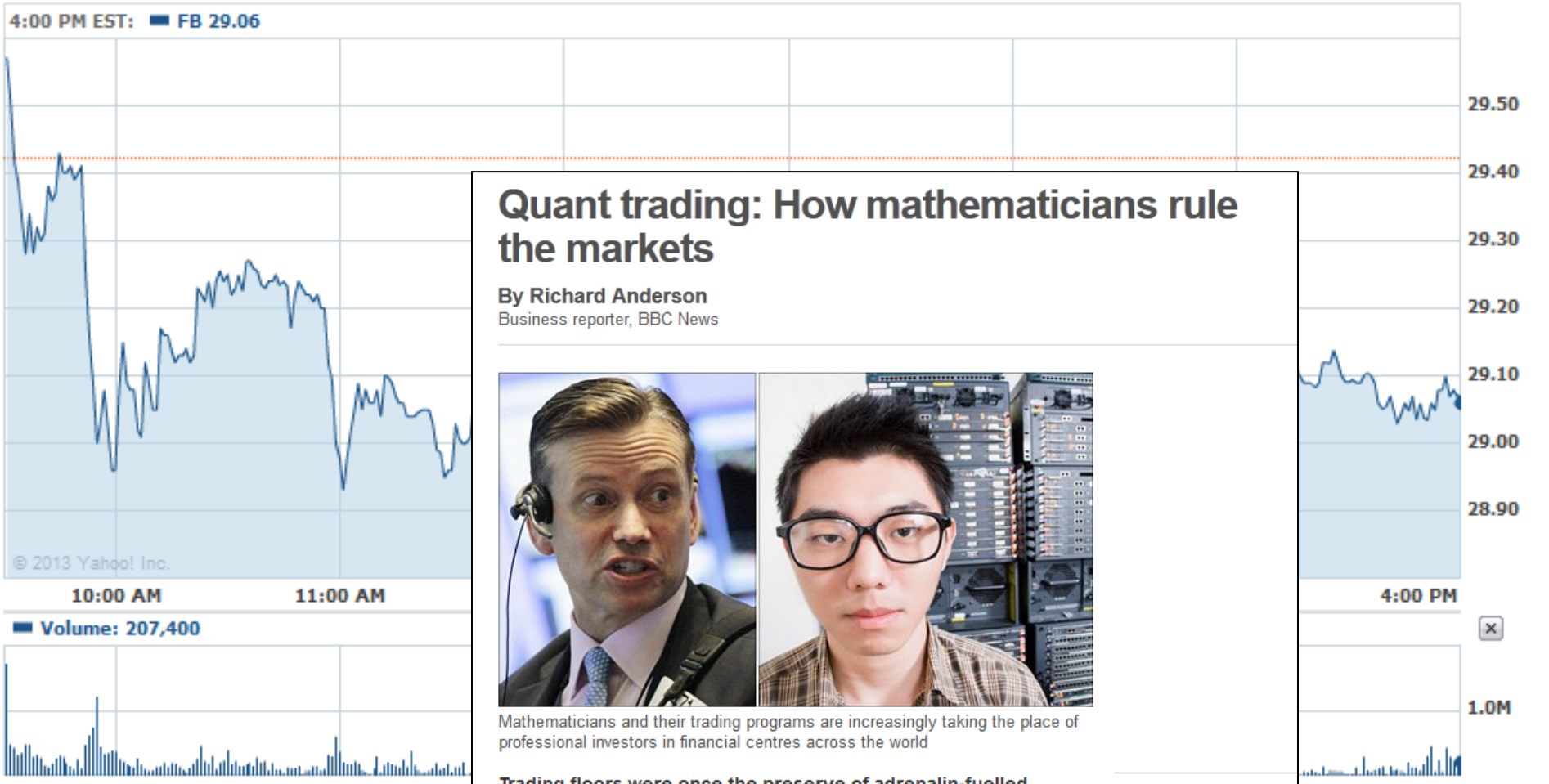
  $$y = g(x \mid \theta)$$

  $g(\ )$ model,

  θ parameters



Example: Algorithmic Trading



Quant trading: How mathematicians rule the markets

By Richard Anderson
Business reporter, BBC News

Mathematicians and their trading programs are increasingly taking the place of professional investors in financial centres across the world

Trading floors were once the preserve of adrenalin-fuelled dealers aggressively executing the orders of brokers who relied on research, experience and gut instinct to decide where best to invest.

Long ago computers made dealers redundant, yet brokers and their ilk have remained the masters of the investment universe, free to buy and

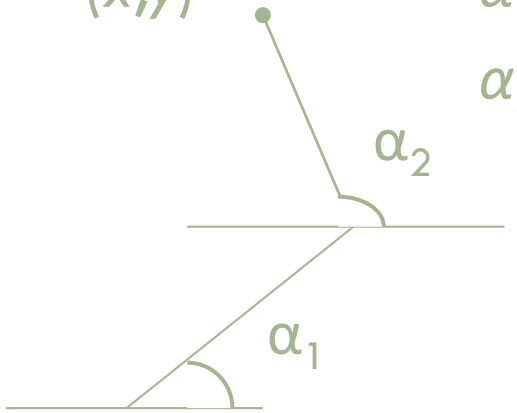
Related Stories

When algorithms control the world

Lone trader caused US share crash

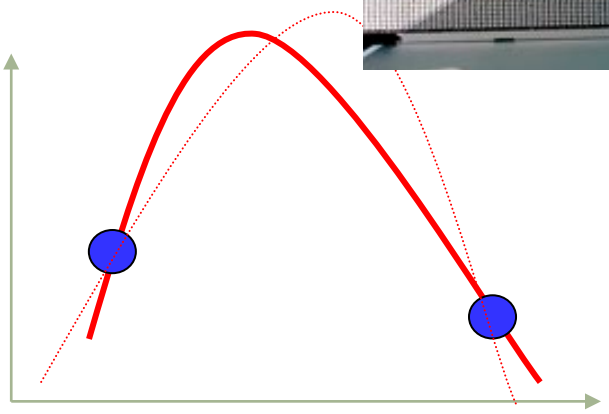
Regression Applications

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



$\alpha_1 = g_1(x,y)$

$\alpha_2 = g_2(x,y)$



- Response surface design

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: credit card application
- Compression: The rule is simpler than the data it explains: feature selection
- Outlier detection: Exceptions that are not covered by the rule, e.g., fraud

Applications

- Association
- Supervised Learning
  - Classification
  - Regression
- Unsupervised Learning
- Reinforcement Learning

Unsupervised Learning

- Learning “what normally happens”
- No output
- Clustering: Grouping similar instances
- Example applications
  - Customer segmentation in CRM
  - Image compression: Color quantization
  - Bioinformatics: Learning motifs

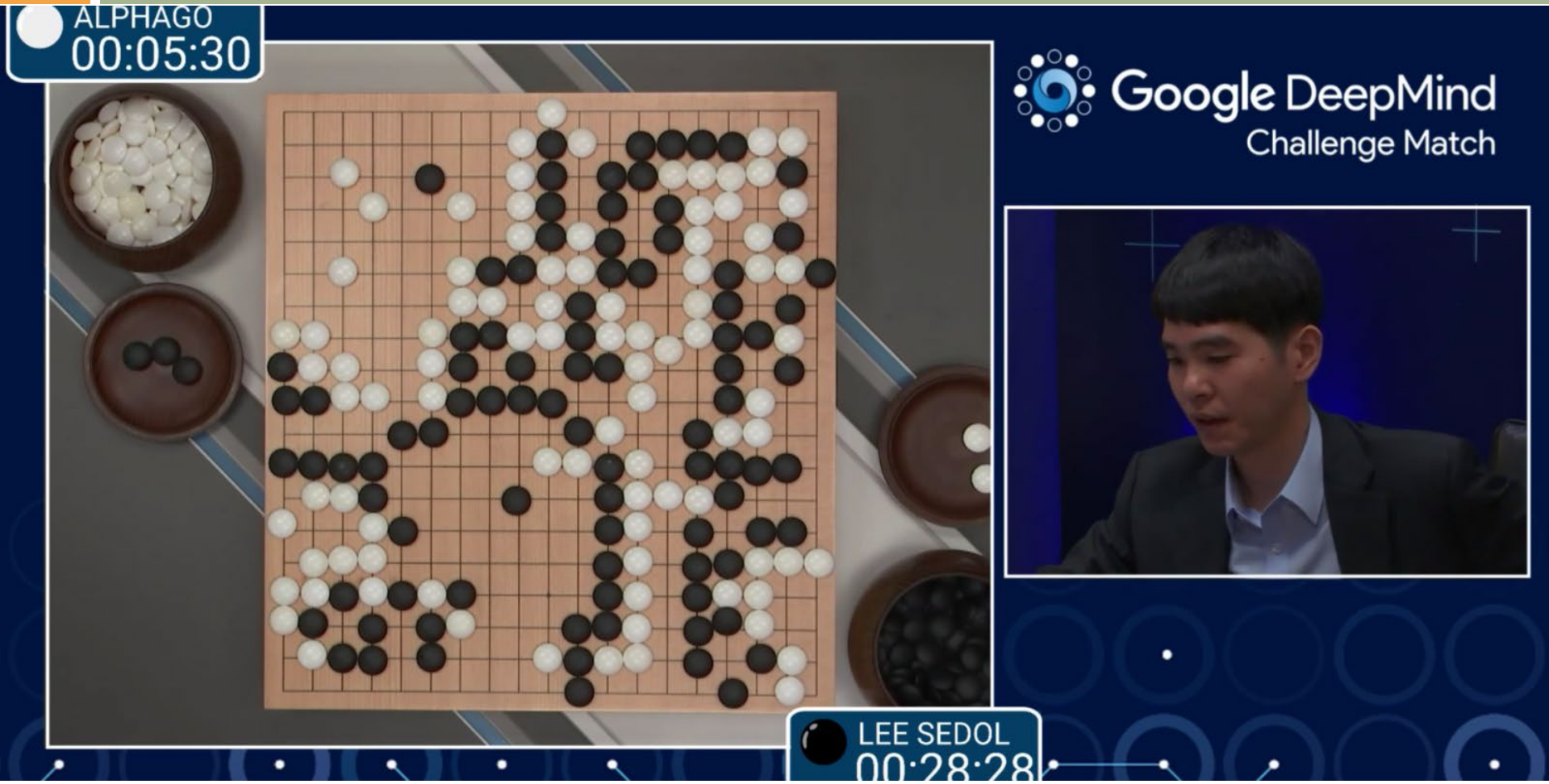
Unsupervised Learning

- Generative models

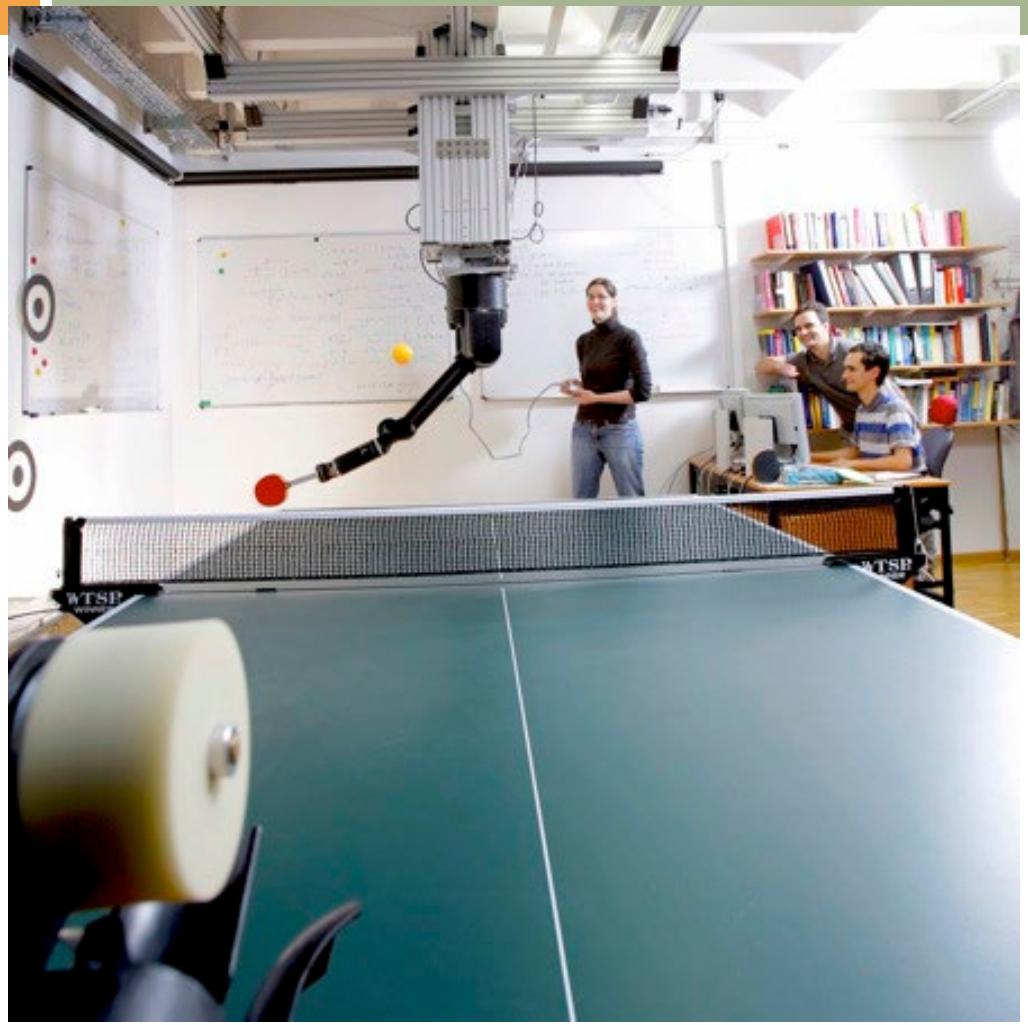
Reinforcement Learning

- Learning a policy: A sequence of outputs
- No supervised output but delayed reward
- Credit assignment problem
- Game playing
- Robot in a maze
- Multiple agents, partial observability, ...

Example: Learning to Act



Example: Learning to Act



Robotic Ping Pong

Resources: Datasets

- UCI Repository: http://www.ics.uci.edu/~mlearn/MLRepository.html
- Statlib: http://lib.stat.cmu.edu/

Resources: Journals

- **Journal of Machine Learning Research www.jmlr.org**
- **IEEE Trans on Pattern Analysis and Machine Intelligence**
- **Machine Learning**
- Neural Computation
- Neural Networks
- IEEE Trans on Neural Networks and Learning Systems
- Journals on Statistics/Data Mining/Signal Processing/Natural Language Processing/Bioinformatics/…

Resources: Conferences

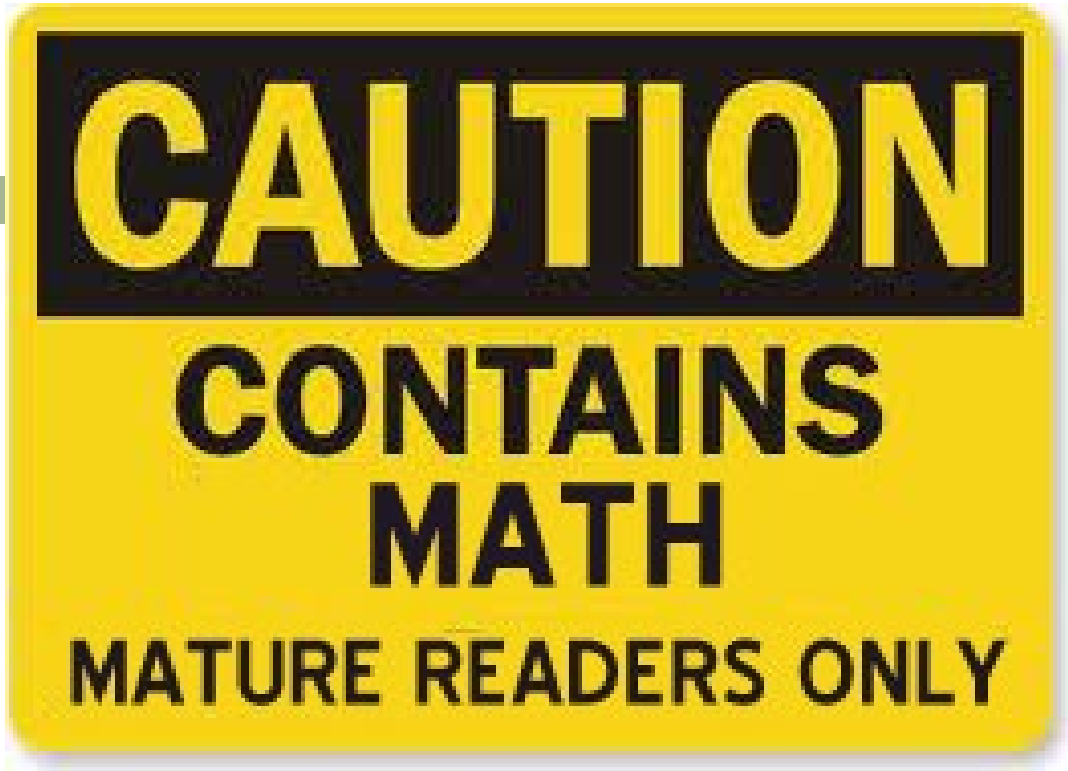
- **International Conference on Machine Learning (ICML)**
- **Neural Information Processing Systems (NeurIPS)**
- **Uncertainty in Artificial Intelligence (UAI)**
- **Computational Learning Theory (COLT)**
- International Conference on AI & Statistics (AISTATS)
- European Conference on Machine Learning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

No Free Lunch

"All models are wrong, but some are useful."

George Box, Statistician

- Modeling assumptions that work well for one problem, may not work well for another
- No "universal learner" exists
  - We need many different techniques for different domains and task characteristics


CAUTION
CONTAINS
MATH
MATURE READERS ONLY