

# Machine Learning

## -Introduction-

Sang Hyun Park

Robotics Engineering, DGIST

shpark13135@dgist.ac.kr

# This course....

- Introduction to machine learning
- Cover the most commonly used machine learning algorithms
  - Supervised learning
  - Unsupervised learning
  - Sequential learning
  - Reinforcement learning

# Brief overview

Instructor: Sang Hyun Park

Email: [shpark13135@dgist.ac.kr](mailto:shpark13135@dgist.ac.kr)

Phone: 053-785-6222

Room: E5, 409

# Brief overview

2017.02- , DGIST



Department of Robotics Engineering  
Medical Image & Signal Processing LAB

2008.03-2014.02, Ph.D., Seoul National University



Computer Vision

2014.03-2016.02, PostDoc., University of North Carolina



Medical Image Analysis

2016.03-2017.02, PostDoc., SRI International



Analysis of  
Big Neuroimage Data

# Brief overview

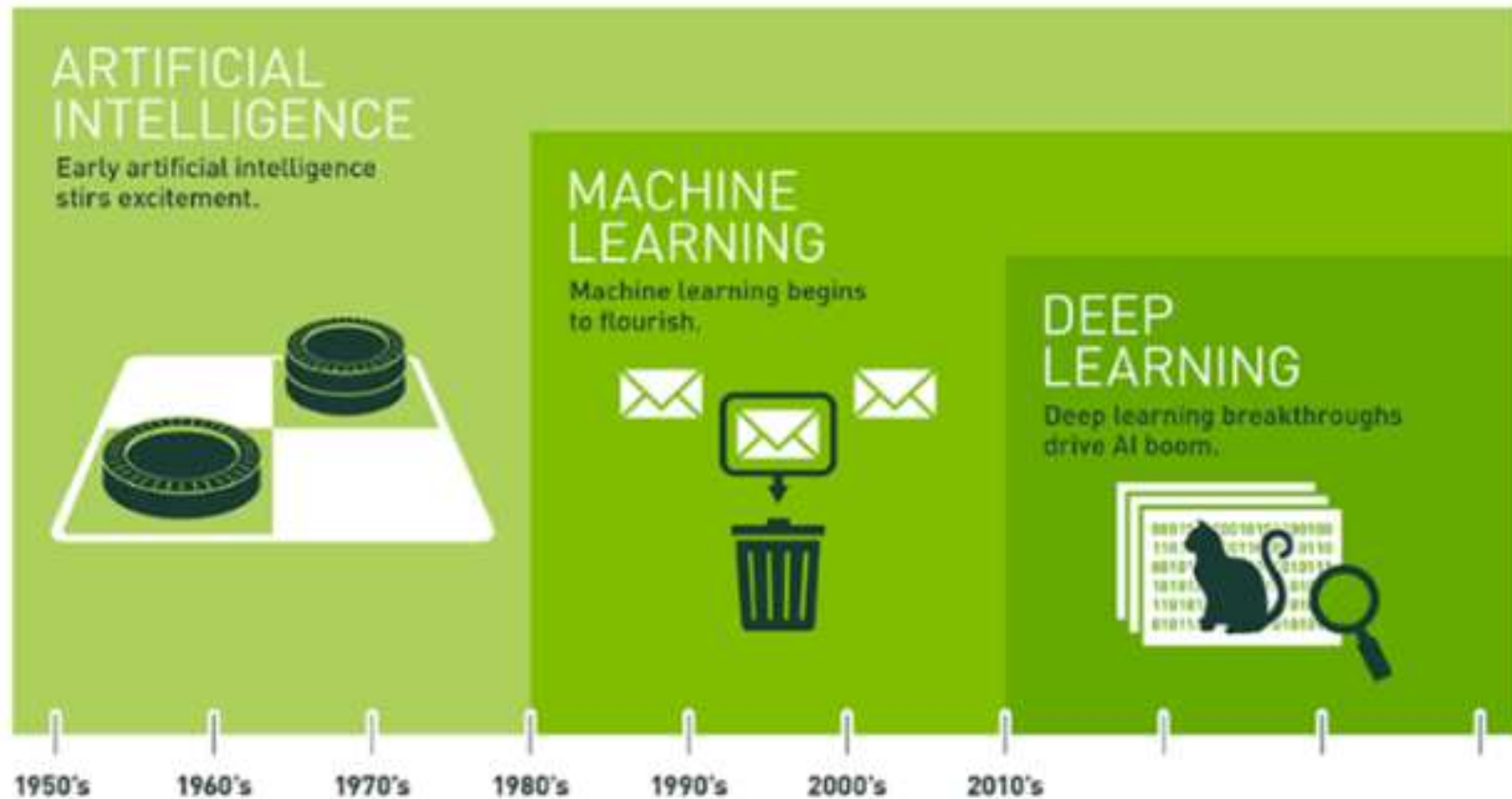
Assignments (40%)

Final exam (40%)

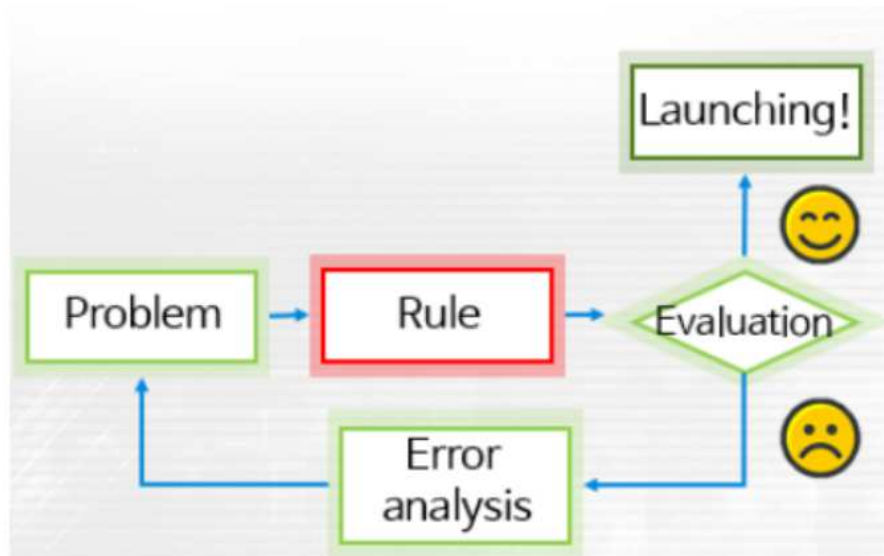
Attendance (20%)

# What is Machine Learning?

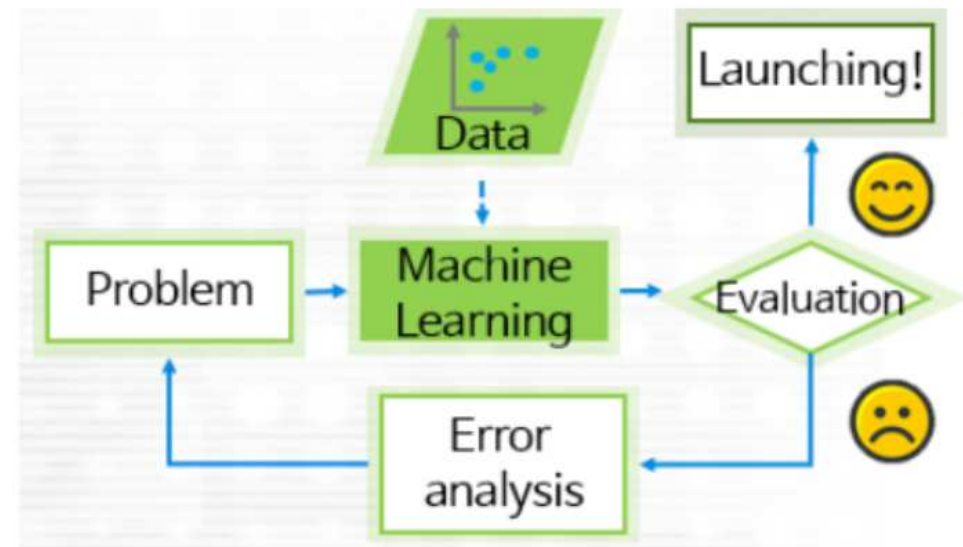
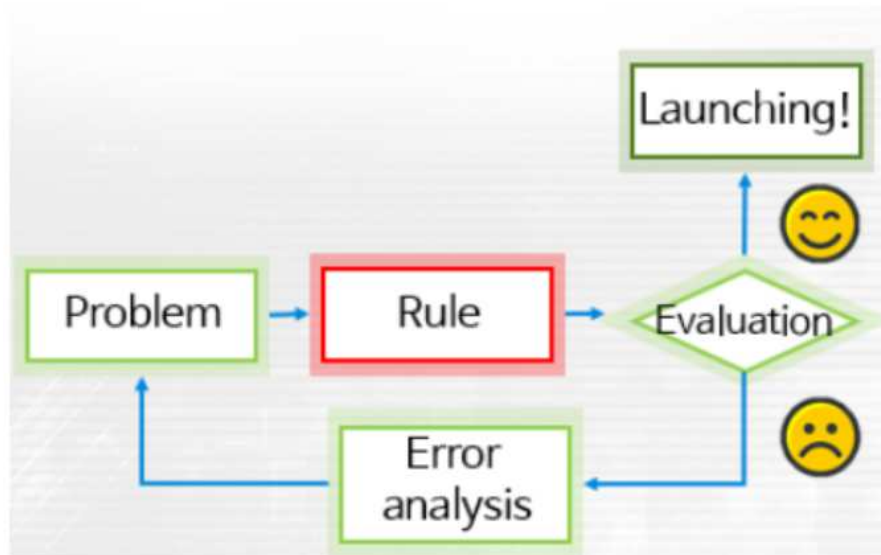
Artificial Intelligent    vs    Machine Learning



# What is Machine Learning?



# What is Machine Learning?





# Contents in this course

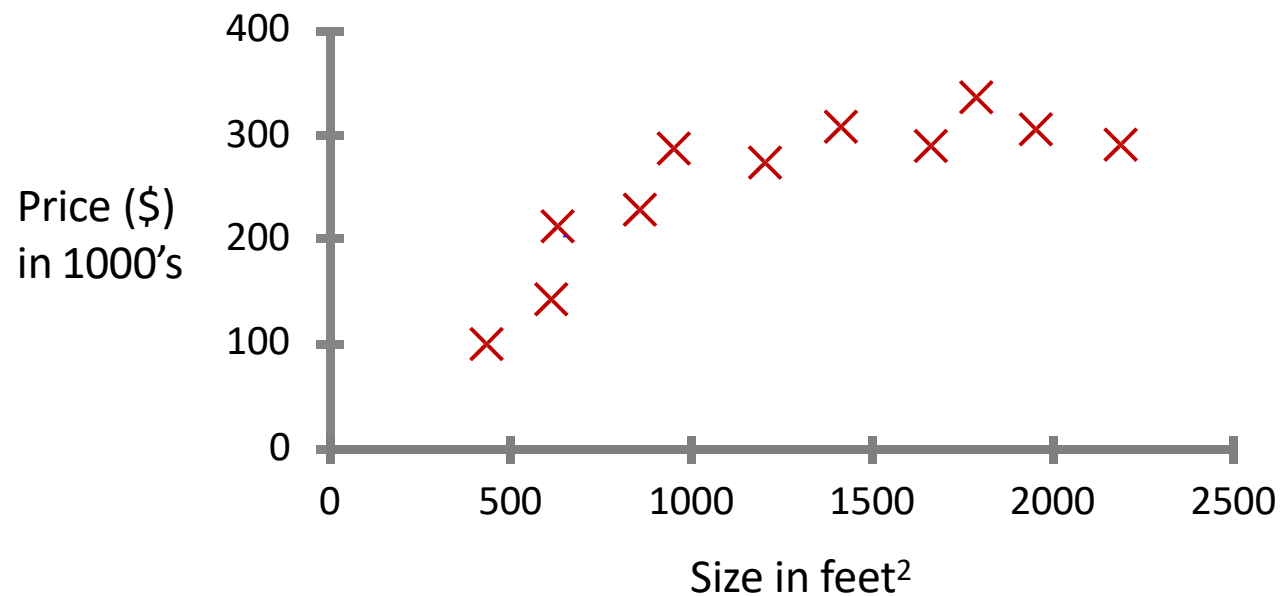
- Supervised Learning
- Unsupervised Learning
- Sequential learning - Hidden Markov Model
- Reinforcement Learning

# Contents in this course

- Supervised Learning
  - Linear regression
  - Logistic regression
  - Neural network
  - Support vector machine
  - Ensemble learning, Adaboost
  - Decision Tree, Random Forest
- Unsupervised Learning
  - Clustering
  - EM algorithm
  - Auto-encoder
  - Principal component analysis
  - Collaborate filtering
  - Semi-supervised learning
- Sequential learning - Hidden Markov Model
- Reinforcement Learning

# Regression

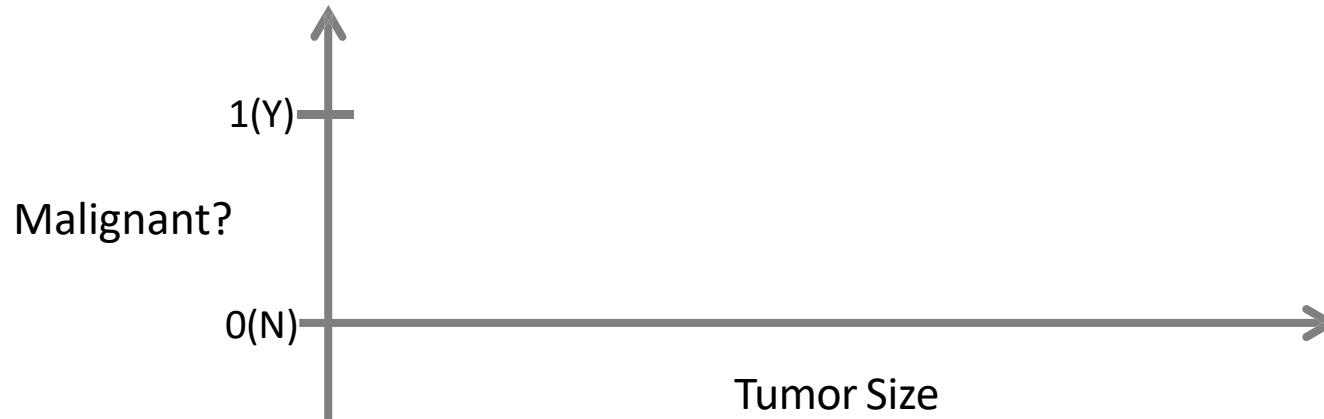
## Housing price prediction



Regression: Predict continuous valued output (price)

# Classification

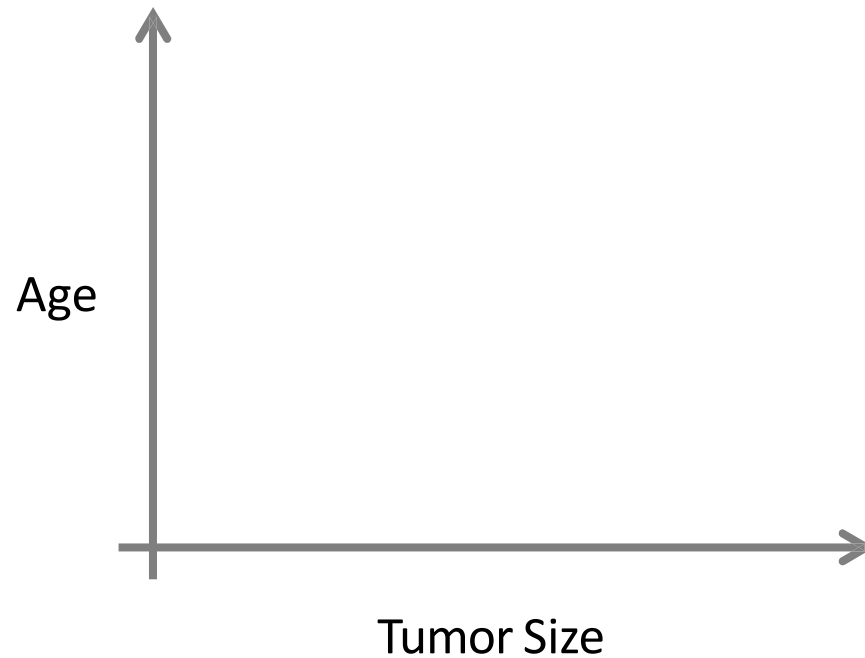
Breast cancer (malignant, benign)



## Classification

Discrete valued output (0 or 1)

# Classification



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

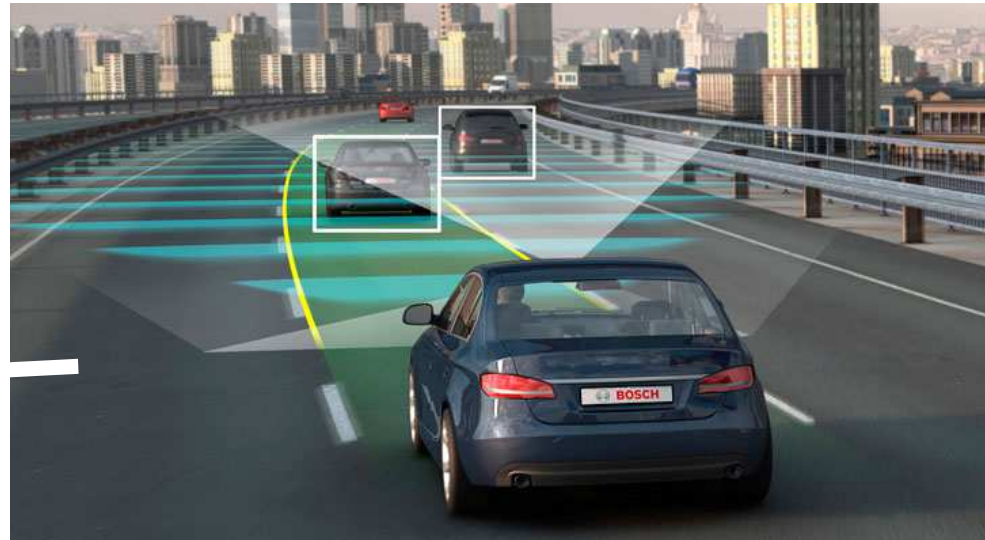
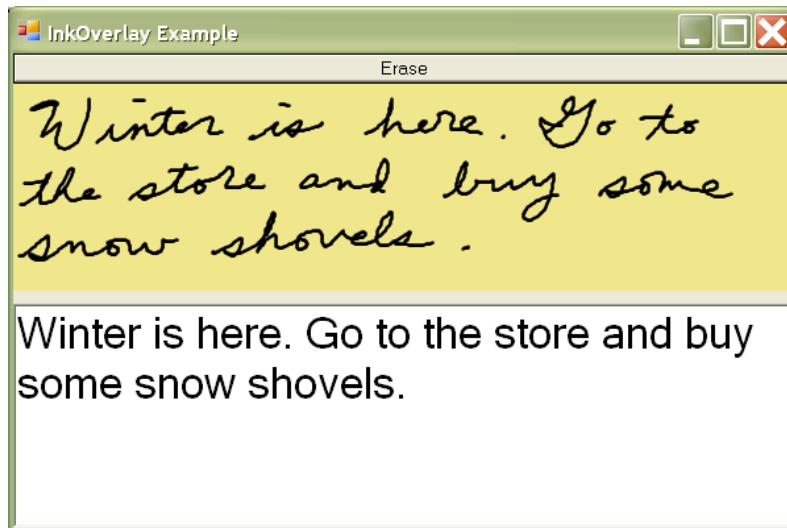
# Classification

- Logistic Regression
- Neural Network
- Support Vector Machine
- Decision Tree

# Supervised learning: Applications

- Automated system

E.g., Autonomous helicopter, handwriting recognition, most of Natural Language Processing (NLP), Computer Vision.



# Sequential Learning

- Named Entity Recognition (NER)

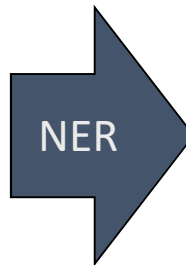
October 14, 2002, 4:00 a.m. PT

For years, Microsoft Corporation CEO Bill Gates railed against the economic philosophy of open-source software with Orwellian fervor, denouncing its communal licensing as a "cancer" that stifled technological innovation.

Today, Microsoft claims to "love" the open-source concept, by which software code is made public to encourage improvement and development by outside programmers. Gates himself says Microsoft will gladly disclose its crown jewels--the coveted code behind the Windows operating system--to select customers.

"We can be open source. We love the concept of shared source," said Bill Veghte, a Microsoft VP. "That's a super-important shift for us in terms of code access."

Richard Stallman, founder of the Free Software Foundation, countered saying...



person company jobTitle

October 14, 2002, 4:00 a.m. PT

For years, Microsoft Corporation CEO Bill Gates railed against the economic philosophy of open-source software with Orwellian fervor, denouncing its communal licensing as a "cancer" that stifled technological innovation.

Today, Microsoft claims to "love" the open-source concept, by which software code is made public to encourage improvement and development by outside programmers. Gates himself says Microsoft will gladly disclose its crown jewels--the coveted code behind the Windows operating system--to select customers.

"We can be open source. We love the concept of shared source," said Bill Veghte, a Microsoft VP. "That's a super-important shift for us in terms of code access."

Richard Stallman, founder of the Free Software Foundation, countered saying...



# Sequential Learning

person company jobTitle

October 14, 2002, 4:00 a.m. PT

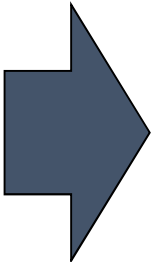
For years, [Microsoft Corporation](#) [CEO Bill Gates](#) [es](#) railed against the economic philosophy of open-source software with Orwellian fervor, denouncing its communal licensing as a "cancer" that stifled technological innovation.

Today, [Microsoft](#) claims to "love" the open-source concept, by which software code is made public to encourage improvement and development by outside programmers. [Gates](#) himself says [Microsoft](#) will gladly disclose its crown jewels--the coveted code behind the Windows operating system--to select customers.

"We can be open source. We love the concept of shared source," said [Bill Veghte](#), a [Microsoft VP](#). "That's a super-important shift for us in terms of code access."

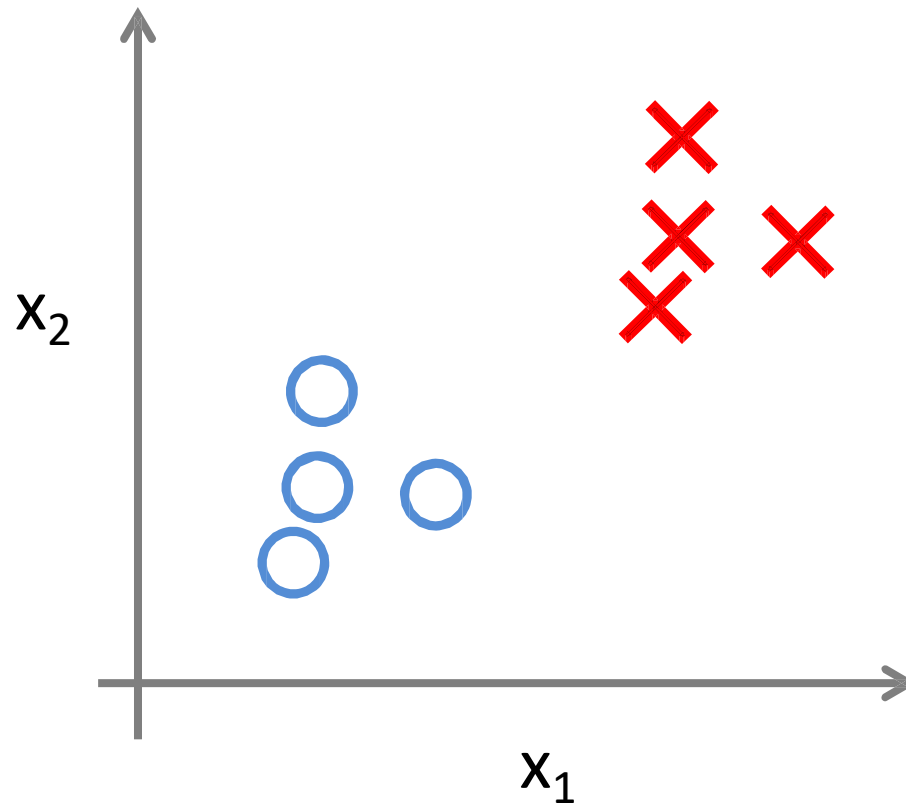
[Richard Stallman](#), [founder](#) of the [Free Software Foundation](#), countered saying...

NER



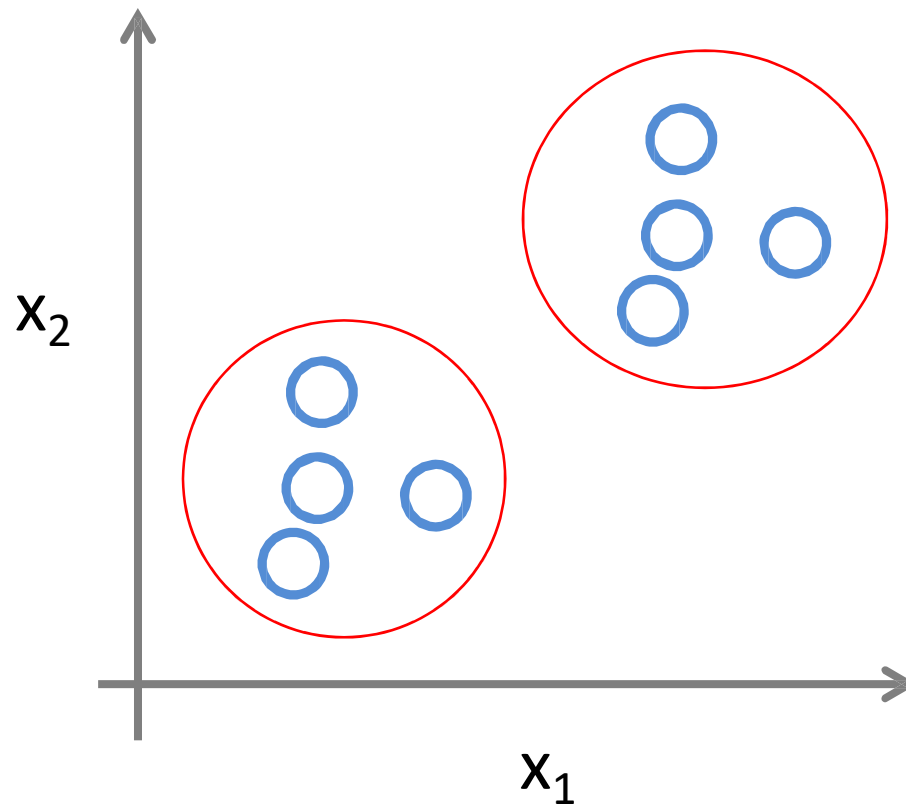
NAME	TITLE	ORGANIZATION
<a href="#">Bill Gates</a>	<a href="#">CEO</a>	<a href="#">Microsoft</a>
<a href="#">Bill Veghte</a>	<a href="#">VP</a>	<a href="#">Microsoft</a>
<a href="#">Richard St...</a>	<a href="#">founder</a>	<a href="#">Free Soft..</a>

# Unsupervised Learning



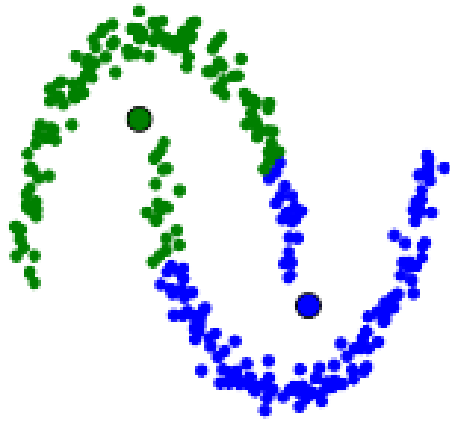
Supervised Learning

# Unsupervised Learning

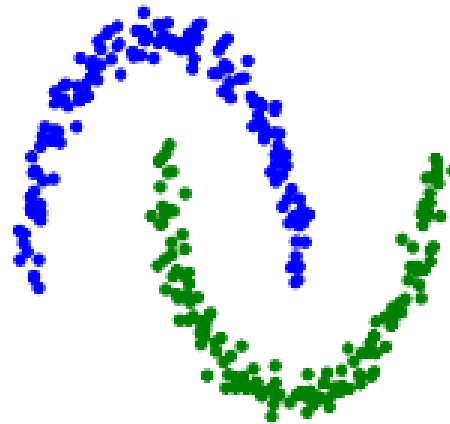


Unsupervised Learning

# Unsupervised Learning

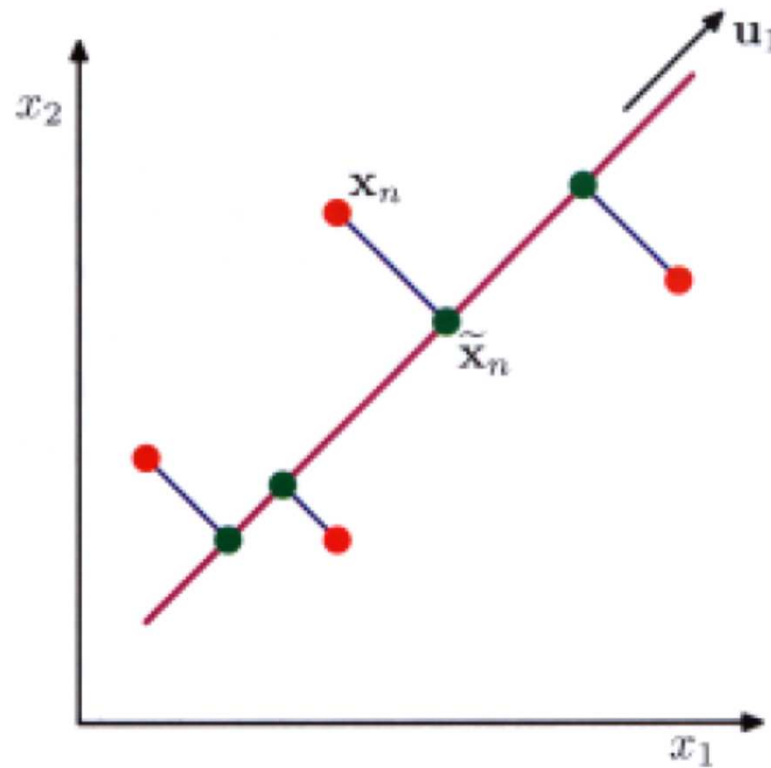


K-means  
clustering



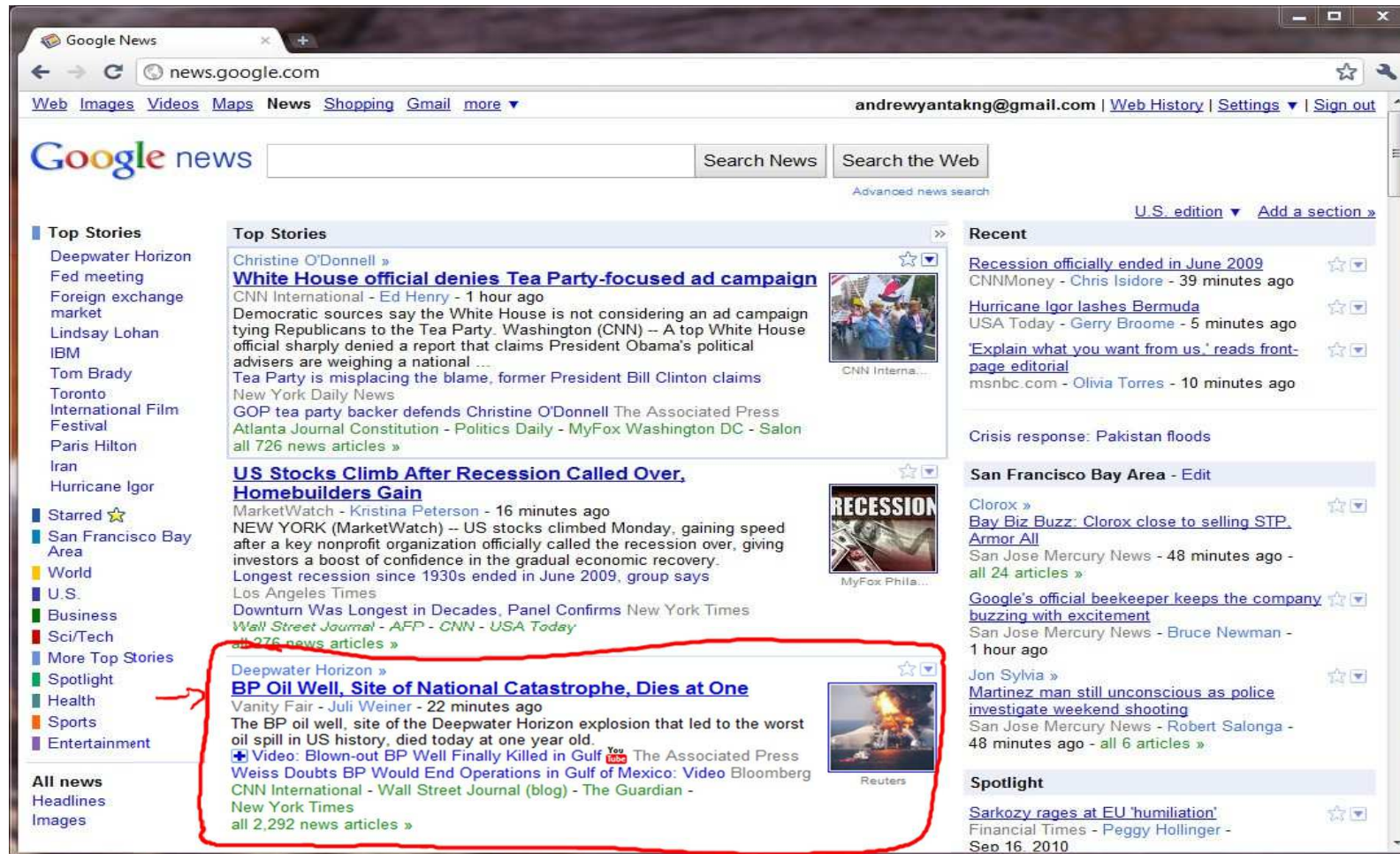
Spectral  
clustering

# Unsupervised Learning

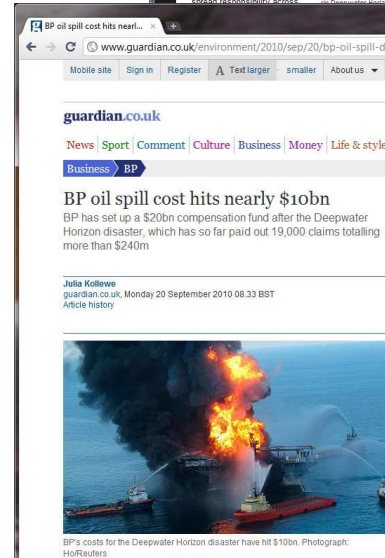
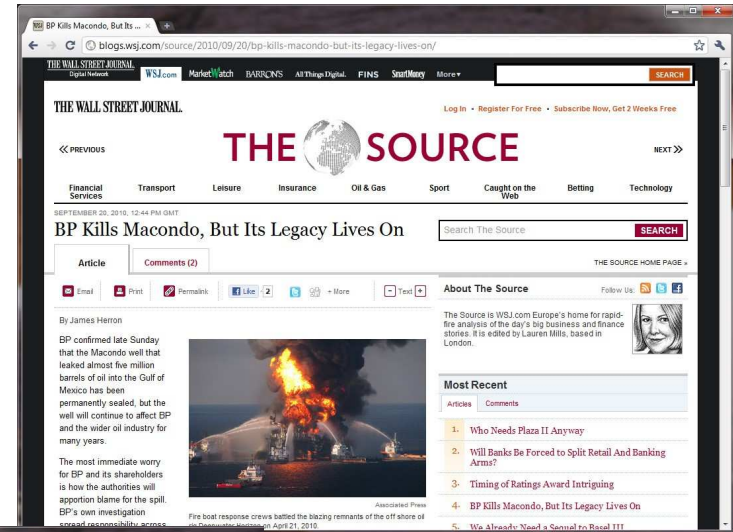
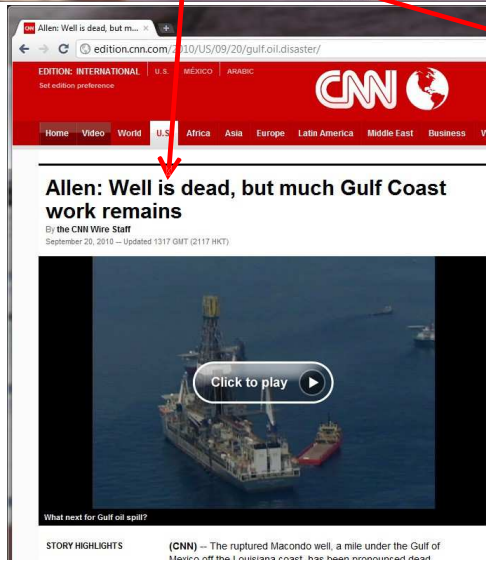
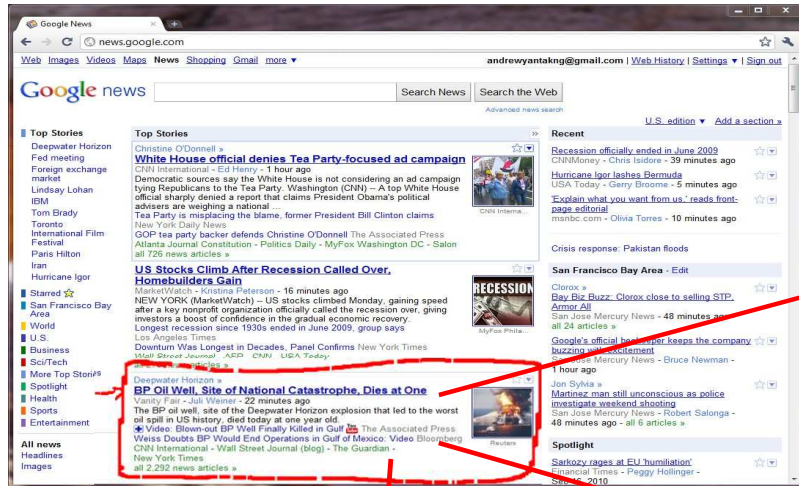


Seek a space of lower dimensionality

# Unsupervised Learning



# Unsupervised Learning





# Machine learning: Applications

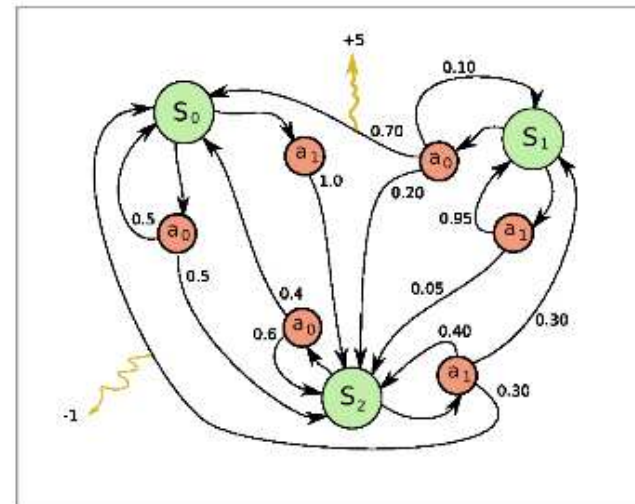
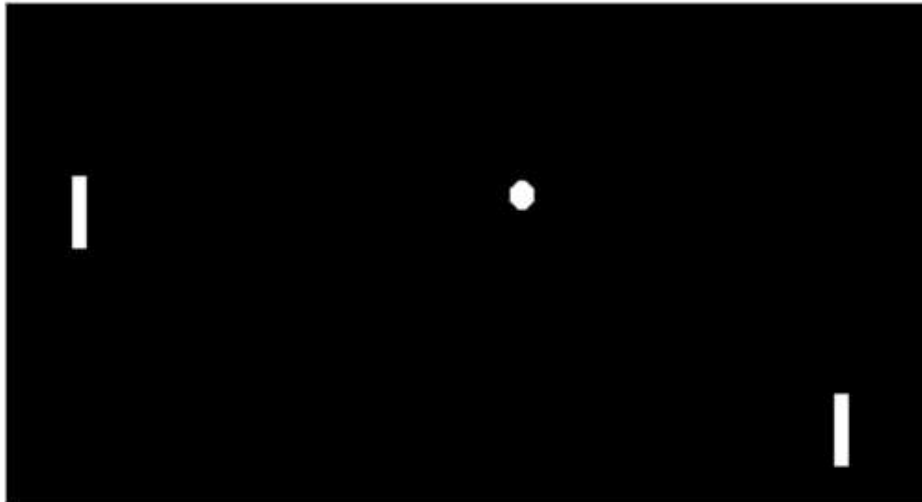
- Self-customizing programs  
E.g., Amazon, Netflix product recommendations





# Reinforcement Learning

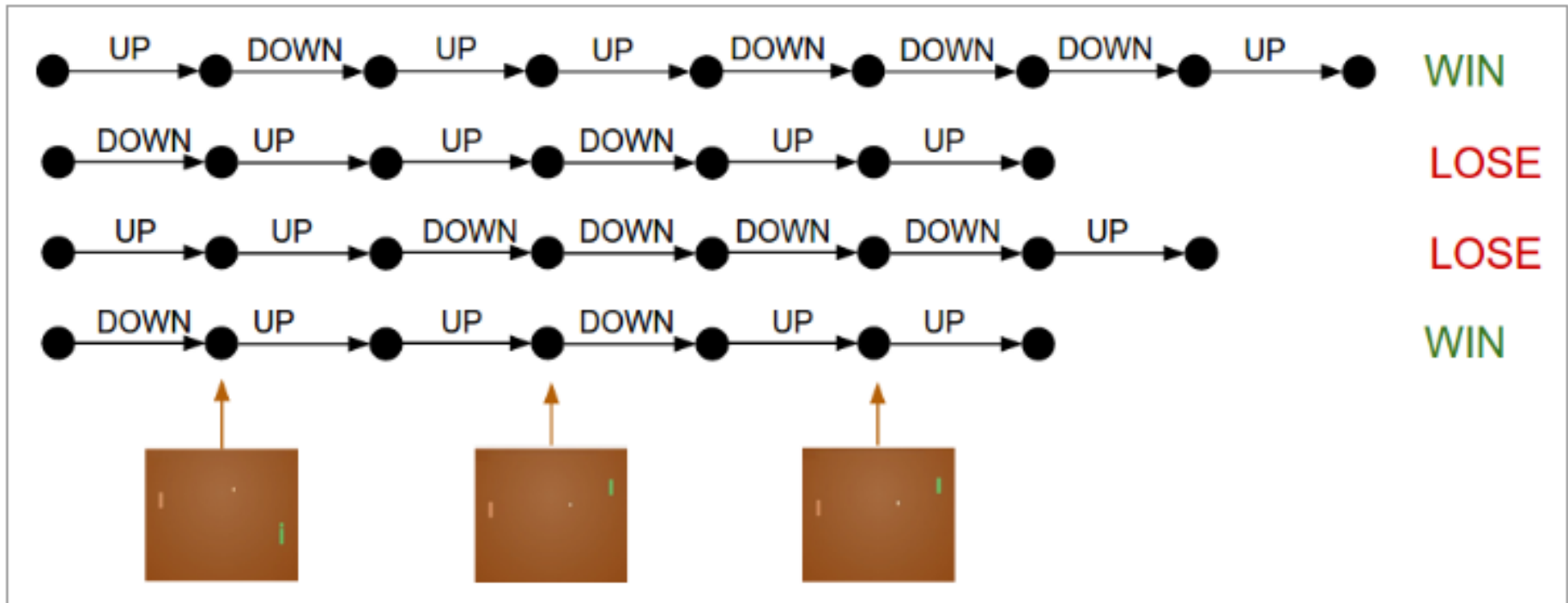
## Atari Game Example



Left: The game of Pong. Right: Pong is a special case of a [Markov Decision Process \(MDP\)](#): A graph where each node is a particular game state and each edge is a possible (in general probabilistic) transition. Each edge also gives a reward, and the goal is to compute the optimal way of acting in any state to maximize rewards.

[<http://karpathy.github.io/2016/05/31/rl/>]

# Reinforcement Learning



Cartoon diagram of 4 games. Each black circle is some game state (three example states are visualized on the bottom), and each arrow is a transition, annotated with the action that was sampled. In this case we won 2 games and lost 2 games. With Policy Gradients we would take the two games we won and slightly encourage every single action we made in that episode. Conversely, we would also take the two games we lost and slightly discourage every single action we made in that episode.

# Summary

- Artificial intelligent
- Machine learning
- Datamining
- Supervised learning
- Unsupervised learning
- Sequential learning
- Reinforcement learning

# References

- Christopher M. Bishop, “Pattern recognition and Machine Learning,” Springer 2006
- Kevin P. Murphy, “Machine Learning: A probabilistic Perspective,” MIT Press 2012
- Lecture slices on <https://www.coursera.org/learn/machine-learning/>
- Lecture slices on <http://www.cedar.buffalo.edu/~srihari/CSE574/>
- Lecture slices on [http://www.cs.cmu.edu/~tom/10701\\_sp11/lectures.shtml](http://www.cs.cmu.edu/~tom/10701_sp11/lectures.shtml)
- Lecture slices on [http://curtis.ml.cmu.edu/w/courses/index.php/Syllabus\\_for\\_Machine\\_Learning\\_10-601B\\_in\\_Spring\\_2016](http://curtis.ml.cmu.edu/w/courses/index.php/Syllabus_for_Machine_Learning_10-601B_in_Spring_2016)