IE 498: Machine Learning for Operations Research

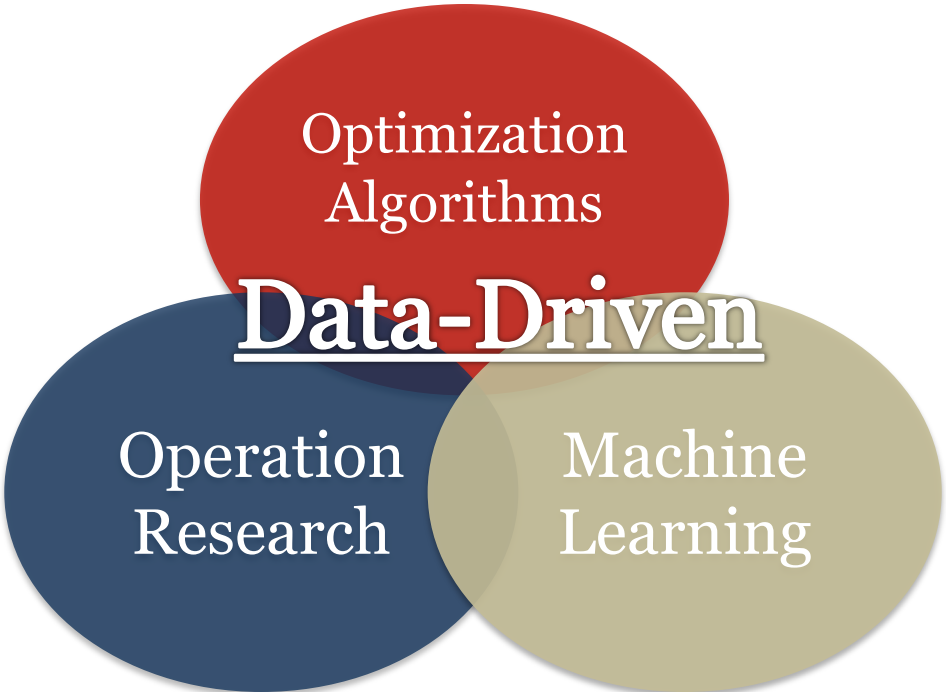
Yuan Zhou

yuanz@Illinois.edu

University of Illinois at Urbana-Champaign

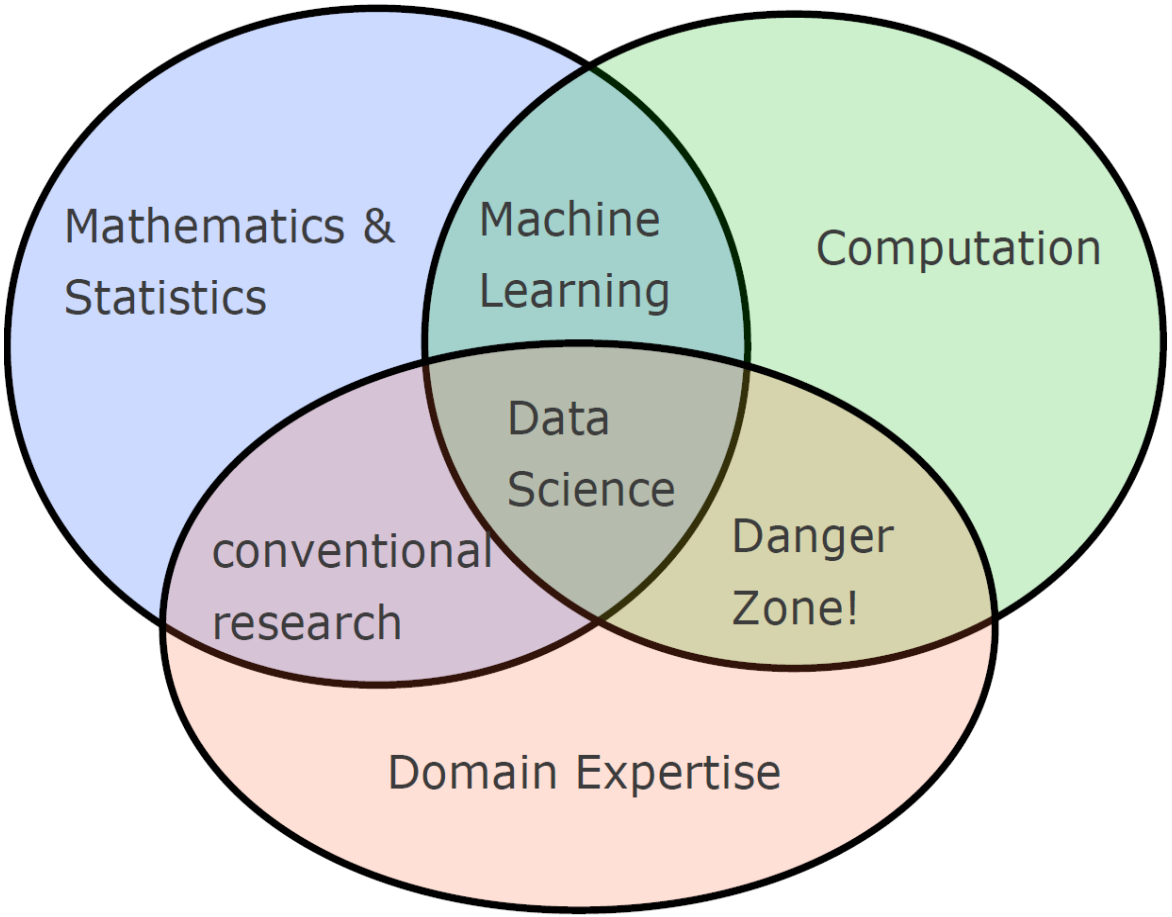
Bio Sketch

- **Ph.D. in Computer Science**
  - **Computer Science Department, Carnegie Mellon University**
- **Instructor in Applied Mathematics**
  - **Massachusetts Institute of Technology**
- **Assistant Professor**
  - **Computer Science Department, Indiana University**
  - **ISE Department and CS Department (affiliate), UIUC**



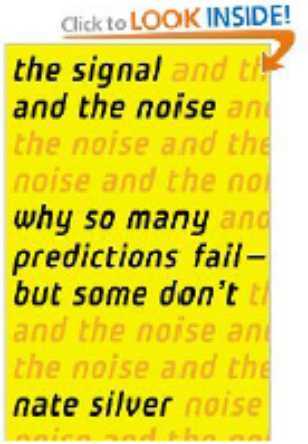
Data Science

- Machine Learning + Domain Expertise (Yann LeCun)

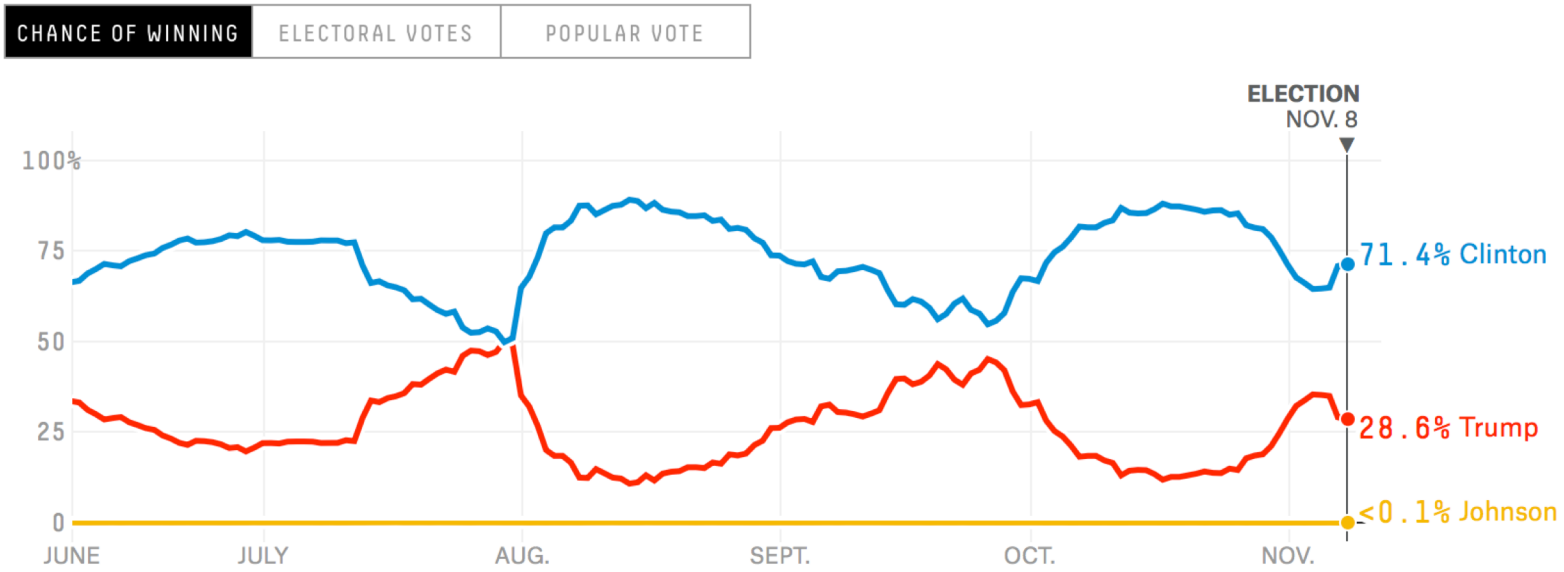


Great success in many applications

Great success in many applications



But not every time..



Great success in many applications

- How IBM built Watson, its Jeopardy-playing supercomputer (DailyFinance, 02/08/2011)

- Sample Jeopardy Q & A's.
- Question: **It's the most common last name in over half of the 50 states**
- Answer: Smith

- Question: **If you visit this country you can enter the room where Edvard Grieg did his composing in the home he built**
- Answer: Norway

Great success in many applications

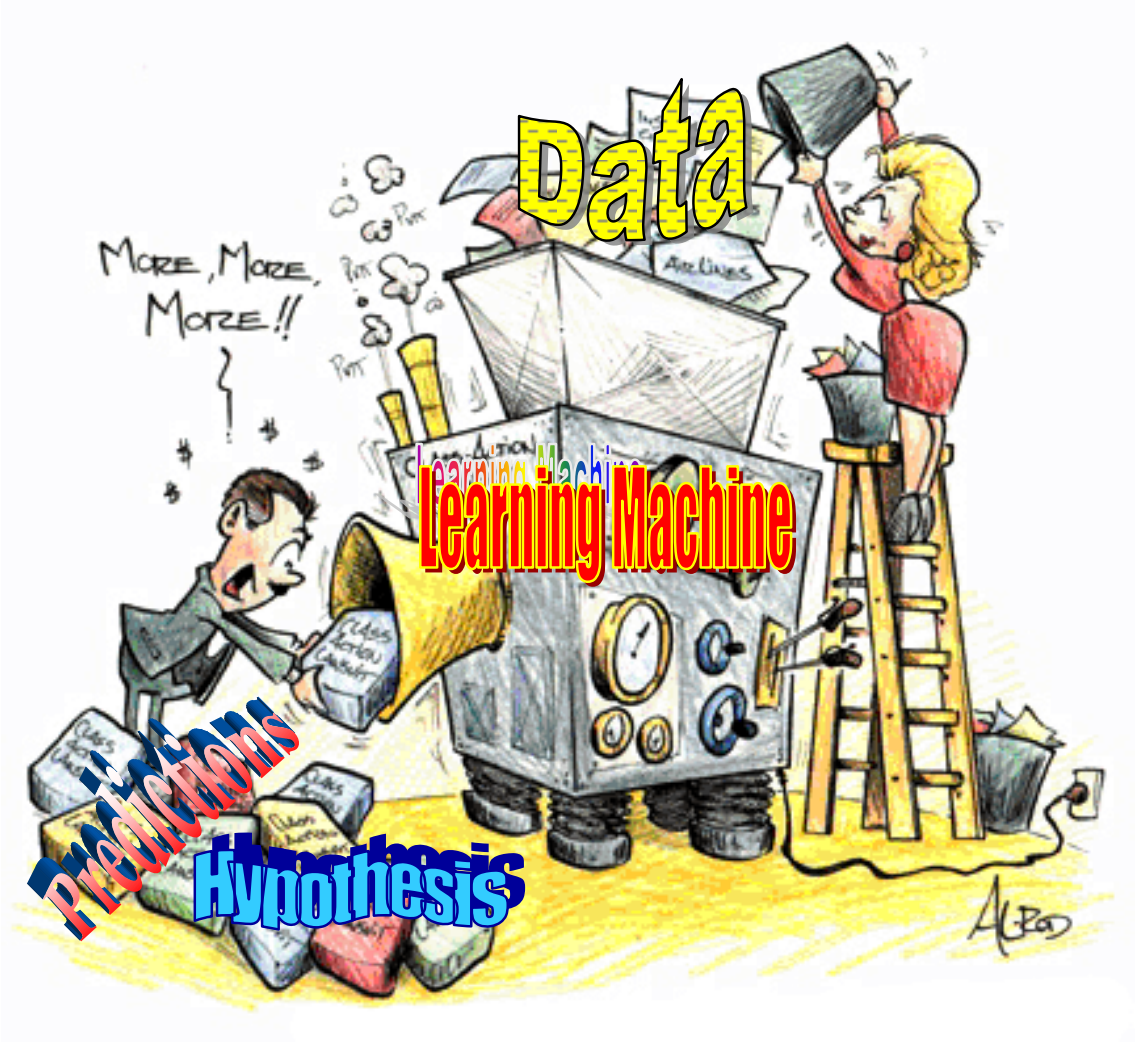
- How IBM built Watson, its Jeopardy-playing supercomputer (DailyFinance, 02/08/2011)

- According to David Ferrucci (PI of Watson DeepQA technology for IBM Research),
- "It's ***machine learning*** allows the computer to become smarter as it tries to answer questions and to ***learn*** as it gets them ***right*** or ***wrong***"

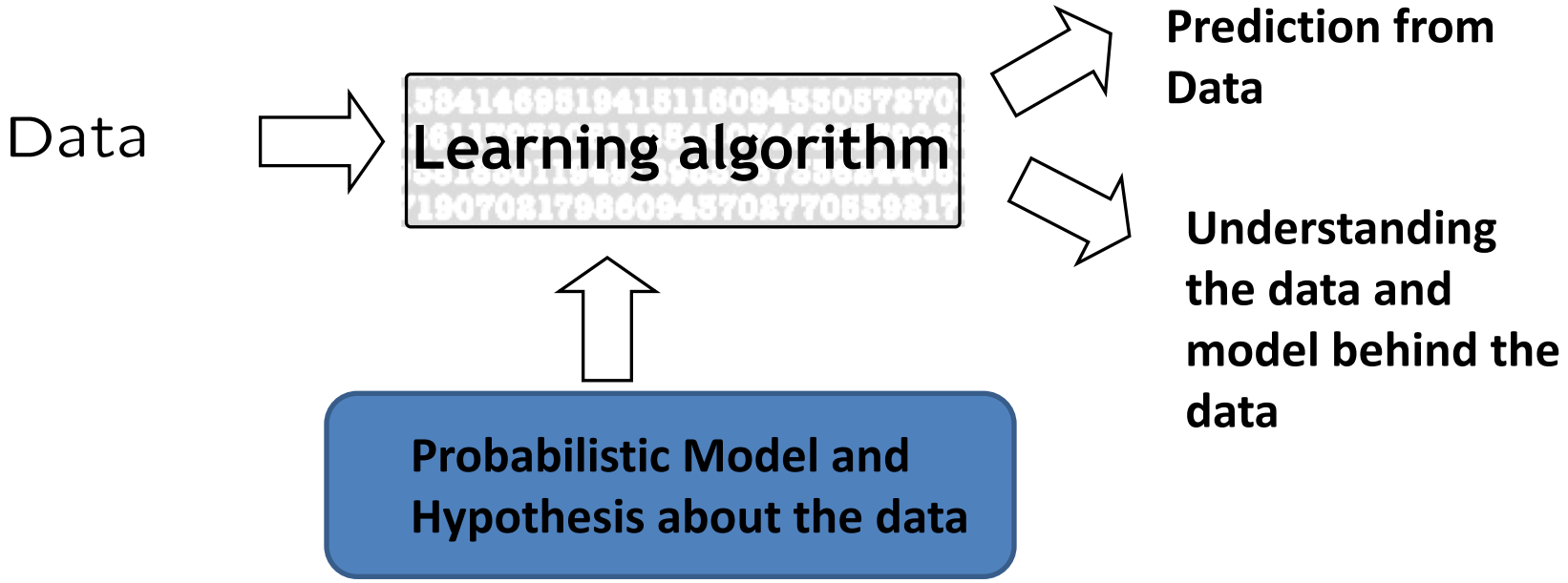
Great success in many applications

- AlphaGo

What is Machine Learning ?



What is Machine Learning?

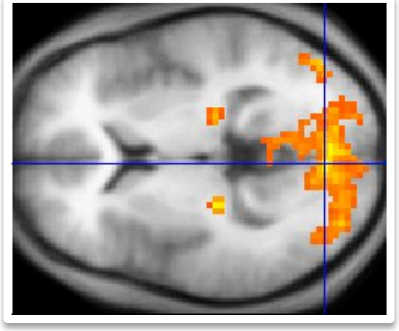


From Data to Understanding …

Machine Learning in Action

Machine Learning in Action

- Decoding thoughts from brain scans



Rob a bank …

Home » Health & Wellness

Brain Scans: Are You a Criminal?

Published February 07, 2007 by:
Andrea Okrentowich
View Profile | Follow | Add to Favorites

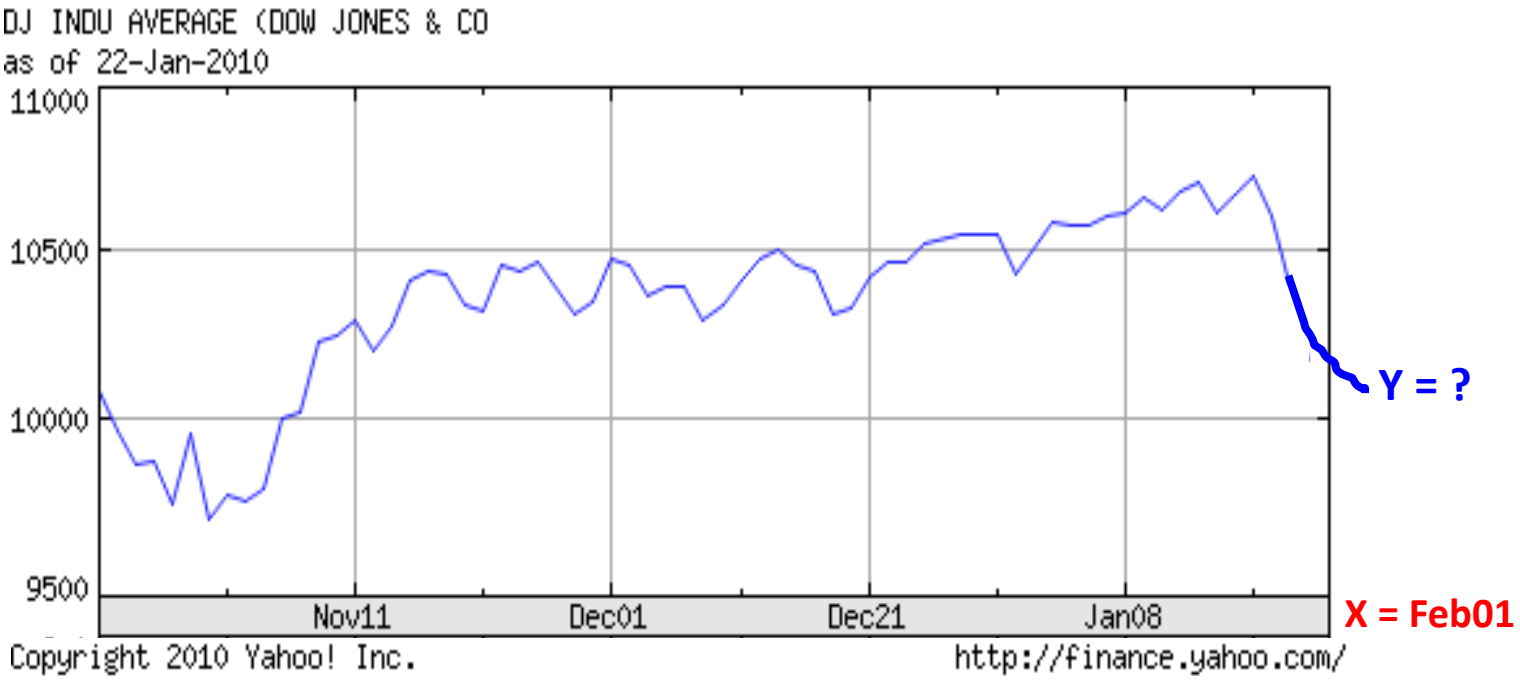
More: Brain Scans | Brain Scan | Disposition | Defendant | Criminal Behavior

MRI Scans as Courtroom Evidence

The average Joe's MRI scan can show a brain abnormality, do we proceed to check him into the nearest mental institution or prison? That would make about as much sense as trying to prove a defendant innocent of a violent

Machine Learning in Action

- Stock Market Prediction



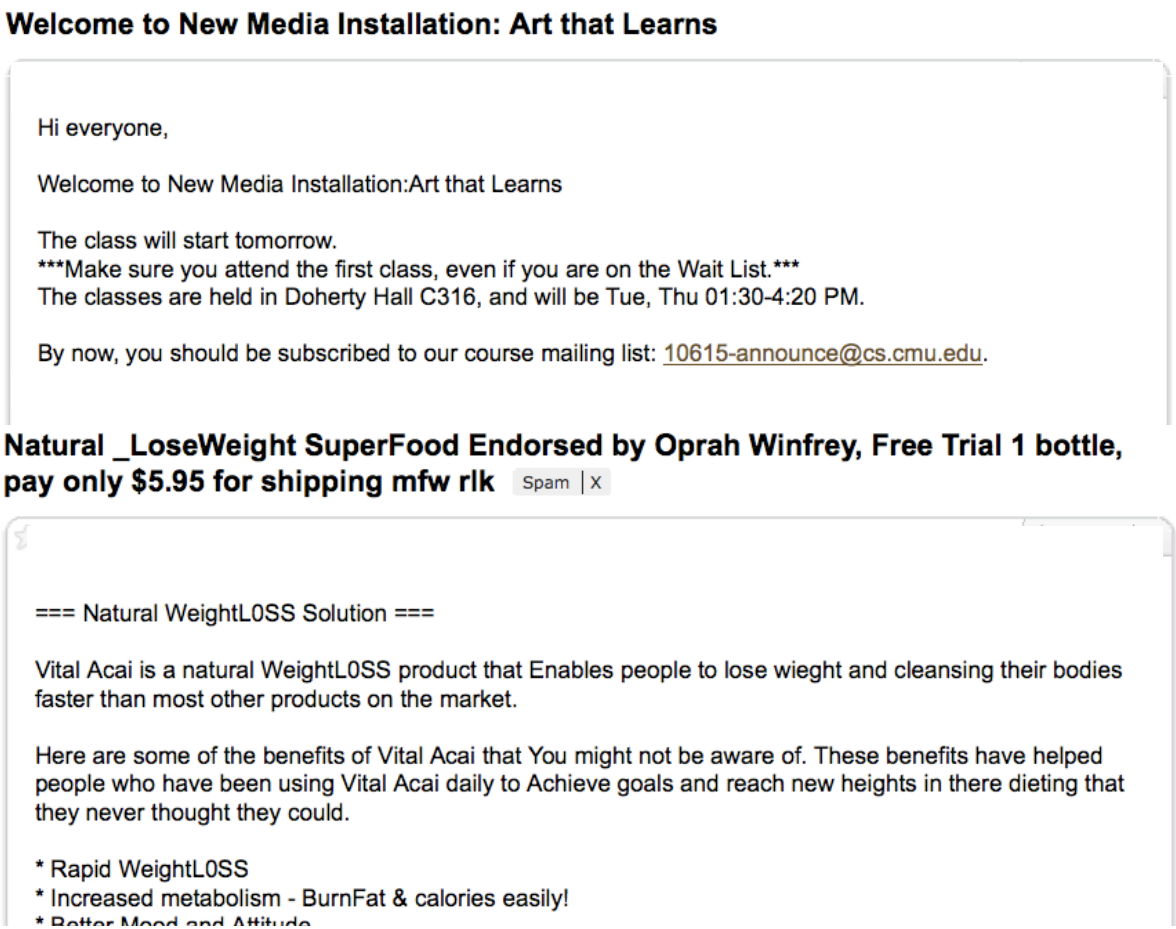
Machine Learning in Action

- Document classification

Sports
Science
News

Machine Learning in Action

- Spam filtering



Spam/
Not spam

Machine Learning in Action

- Cars navigating on their own

Boss, the self-driving SUV
1st place in the DARPA Urban Challenge.
Photo courtesy of Tartan Racing.

Machine Learning in Action

- The **best** helicopter pilot is now a computer!
  - it runs a program that learns how to fly and make acrobatic maneuver by itself!
  - Reinforcement learning!

[http://heli.stanford.edu/]

Machine Learning in Action

- Many, many more…

  Natural language processing
  Speech recognition
  Computer vision
  Medical outcomes analysis
  Computational biology
  Sensor networks
  Social networks
  Finance
  …

ML is trending!

- Wide applicability
- Very large-scale complex systems
  - Internet (billions of nodes), sensor network (new multi-modal sensing devices), genetics (human genome)
- Huge multi-dimensional data sets
  - 30,000 genes x 10,000 drugs x 100 species x …
- Large-scale data (Terabytes, Petabytes of data)
- Demand for self-customization to user, environment

What this course is about

- Covers a wide range of Machine Learning techniques – from basic to state-of-the-art

- Two Parts:
  - Foundations of Machine Learning:
    - Naïve Bayes, linear regression, logistic regression, nearest-neighbor, decision trees, overfitting, regularization, dimensionality reduction, K-means, neural networks, support vector machines, graphical models, EM algorithm,. Hidden Markov Model, semi-supervised learning
  - Learning for decision-making
    - Multi-armed Bandits
    - Reinforcement Learning

- Covers algorithms and applications

Machine Learning Tasks

Broad categories -

- **Supervised learning**

  Classification, Regression

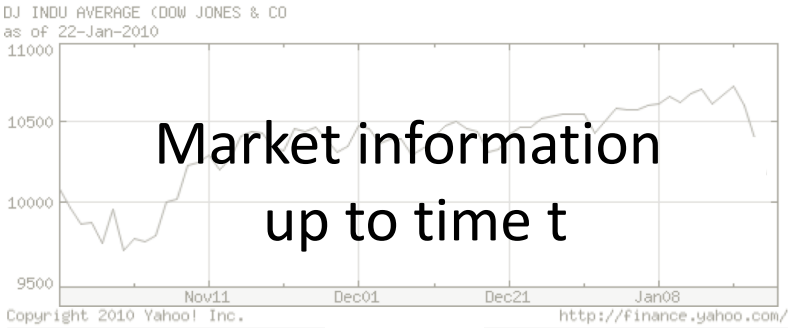
- **Unsupervised learning**

  Clustering, Dimensionality reduction

- Semi-supervised learning
- Online learning
- Reinforcement learning
- Many more …

Supervised Learning

Feature Space $\mathcal{X}$



Label Space $\mathcal{Y}$

"Sports"
"News"
"Science"
…

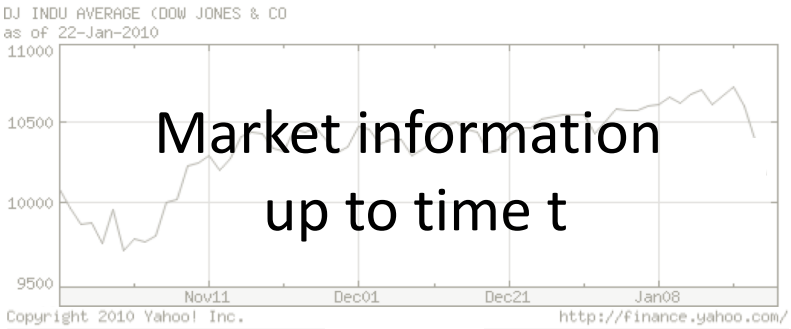
Share Price
"$ 24.50"

Task: Given $X \in \mathcal{X}$, predict $Y \in \mathcal{Y}$.

Supervised Learning - Regression

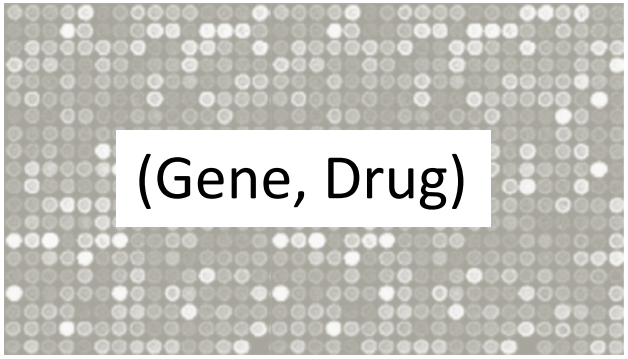
Feature Space $\mathcal{X}$

Label Space $\mathcal{Y}$



Market information up to time t

Share Price
"$ 24.50"



(Gene, Drug)

Expression level
"0.01"

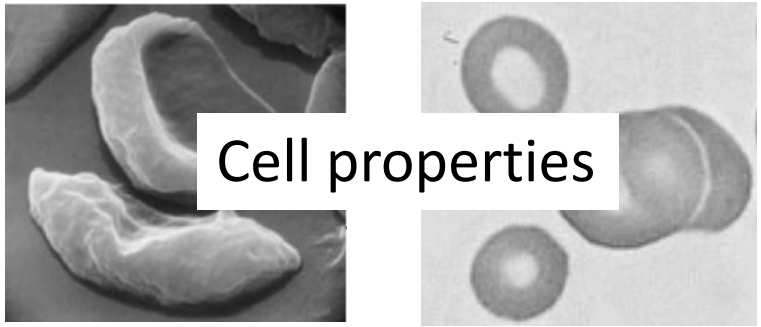
Continuous Labels

Supervised Learning - Classification

Feature Space $\mathcal{X}$

Label Space $\mathcal{Y}$

"Sports"
"News"
"Science"
...



"Anemic cell"
"Healthy cell"

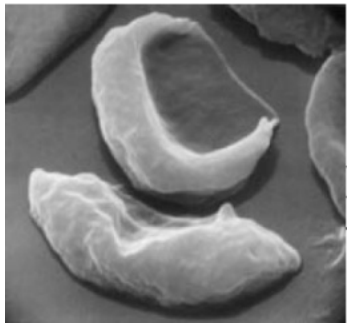
Discrete Labels

Supervised Learning Task

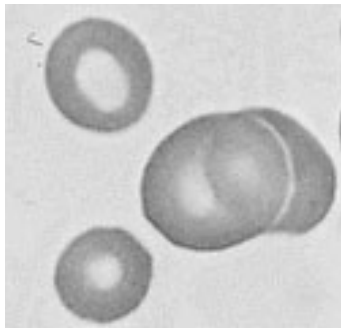
Task: Given $X \in \mathcal{X}$, predict $Y \in \mathcal{Y}$.

X - test data

$\equiv$ Construct **prediction rule** $f : \mathcal{X} \to \mathcal{Y}$



"Anemic cell (0)"

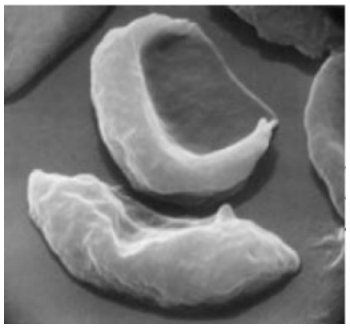


"Healthy cell (1)"

Performance Measures

Performance:

$\text{loss}(Y, f(X))$ - Measure of closeness between true label *Y* and prediction *f*(*X*)

X	*Y*	*f*(*X*)	$\text{loss}(Y, f(X))$
	"Anemic cell"	"Anemic cell"	0
		"Healthy cell"	1

$$\text{loss}(Y, f(X)) = 1_{\{f(X) \neq Y\}}$$ **0/1 loss**

Performance Measures

Performance:

$\text{loss}(Y, f(X))$ - Measure of closeness between true label *Y* and prediction *f*(*X*)

X	Share price, *Y*	*f*(*X*)	$\text{loss}(Y, f(X))$
Past performance, trade volume etc. as of Sept 8, 2010	"$24.50"	"$24.50"	0
		"$26.00"	?
		"$26.10"	?

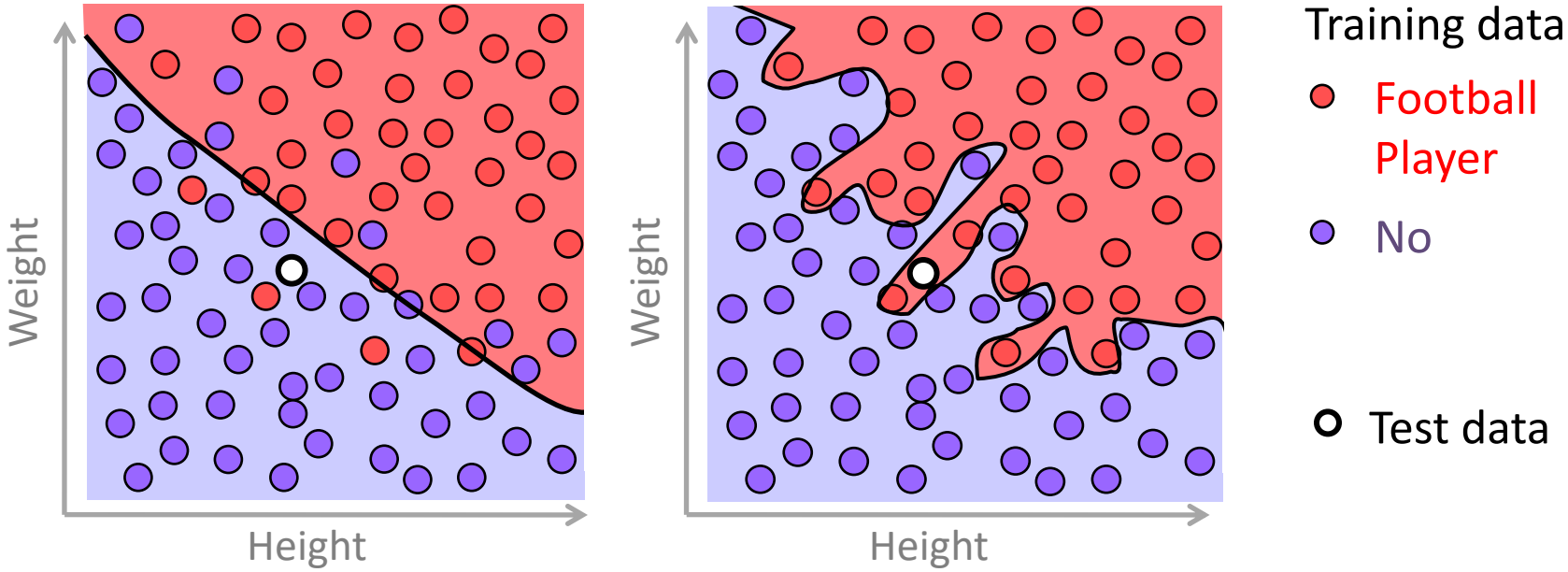
$$\text{loss}(Y, f(X)) = (Y - f(X))^2$$ **square loss**

Supervised Learning – Summary

- Label y (output, response)
  - Discrete: classification problem
  - Continuous: regression problem
- Vector of p features $x \in R^p$ (inputs, predictors)
- Training data (historical observations)
  - $(x_1, y_1), \ldots, (x_n, y_n)$
- Interested in
  - Given a new data x_{new}: how to design a learning algorithm (prediction rule) to predict y_{new}
  - How good the learning algorithm is ?
  - Measure on another set of testing data $(x'_1, y'_1), \ldots (x'_m, y'_m)$

Issues in ML

- A good machine learning algorithm
  - Does not **overfit** training data



  - **Generalizes** well to test data

More later …

Unsupervised Learning

Aka "learning without a teacher"

Feature Space $\mathcal{X}$

Words in a document

Word distribution
(Probability of a word)

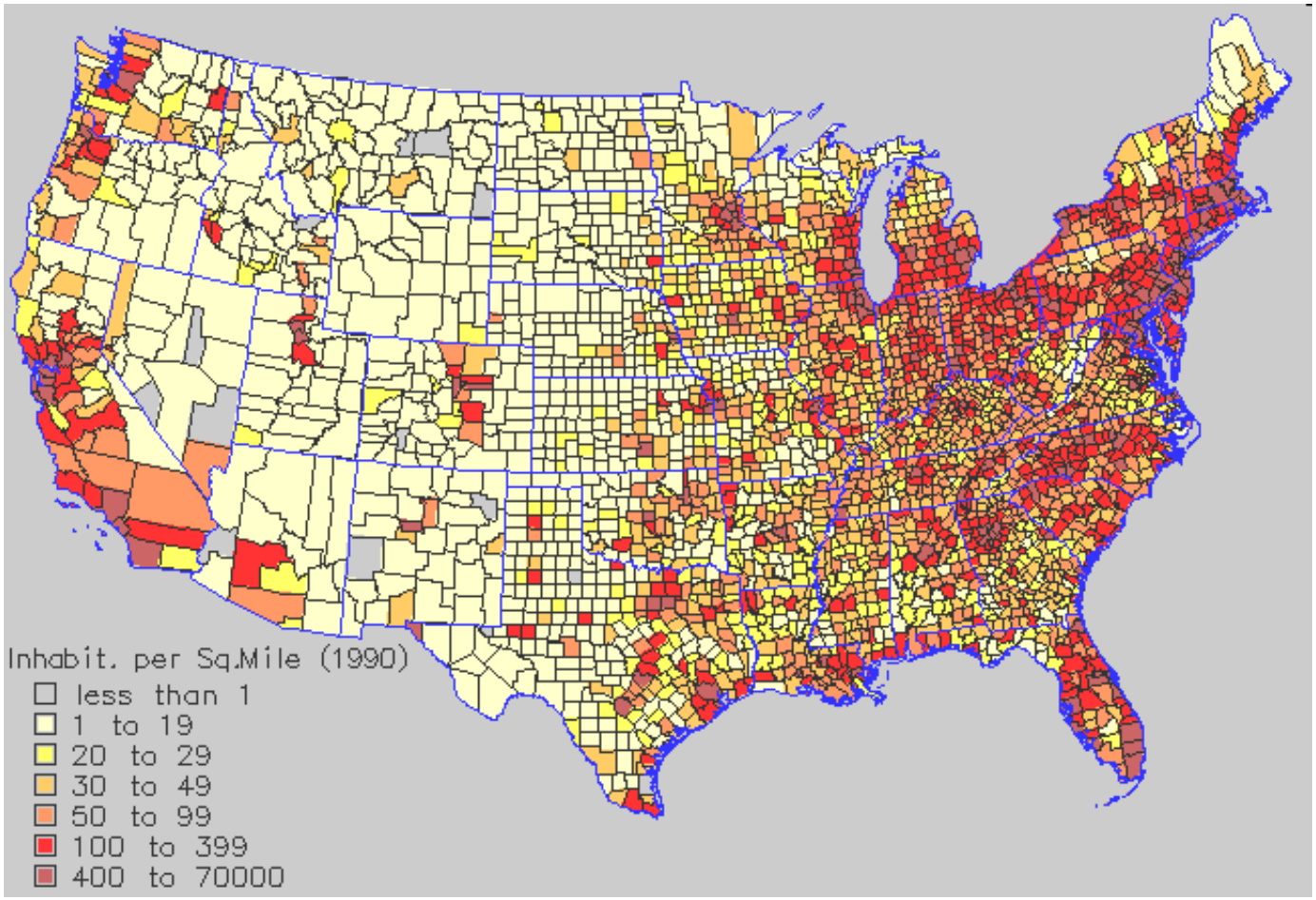
Task: Given $X \in \mathcal{X}$, learn $f(X)$.

Unsupervised Learning – more formally

- Vector of p features $x \in R^p$: no outcome
- A set of data (historical observations)
  - $x_1, \ldots, x_n$
- The goal is more diverse
- Can be served as a pre-processing step for supervised learning

Unsupervised Learning – Density Estimation

Population density

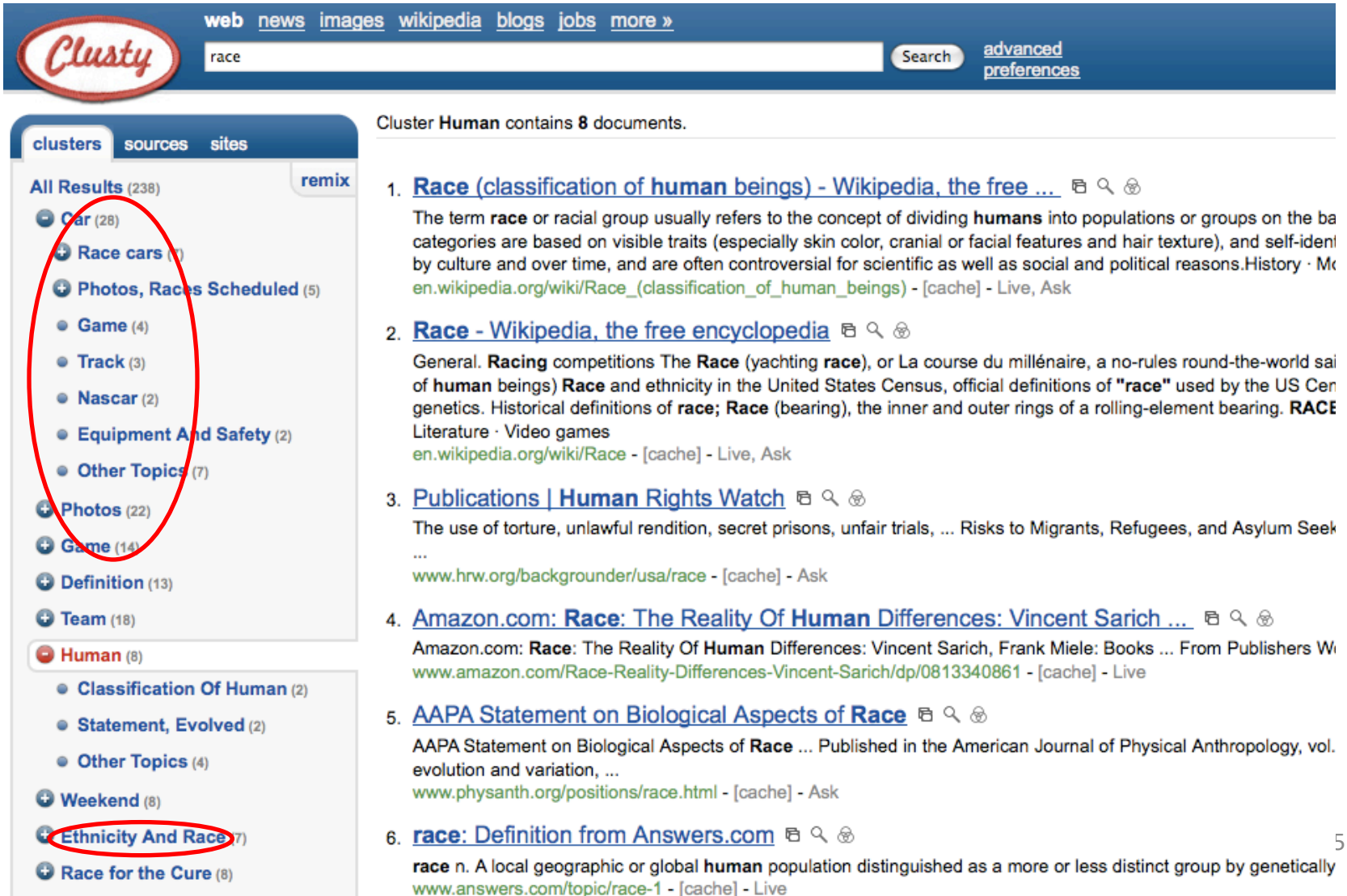


Unsupervised Learning – clustering

Group similar things e.g. images

[Goldberger et al.]

Unsupervised Learning – clustering web search results



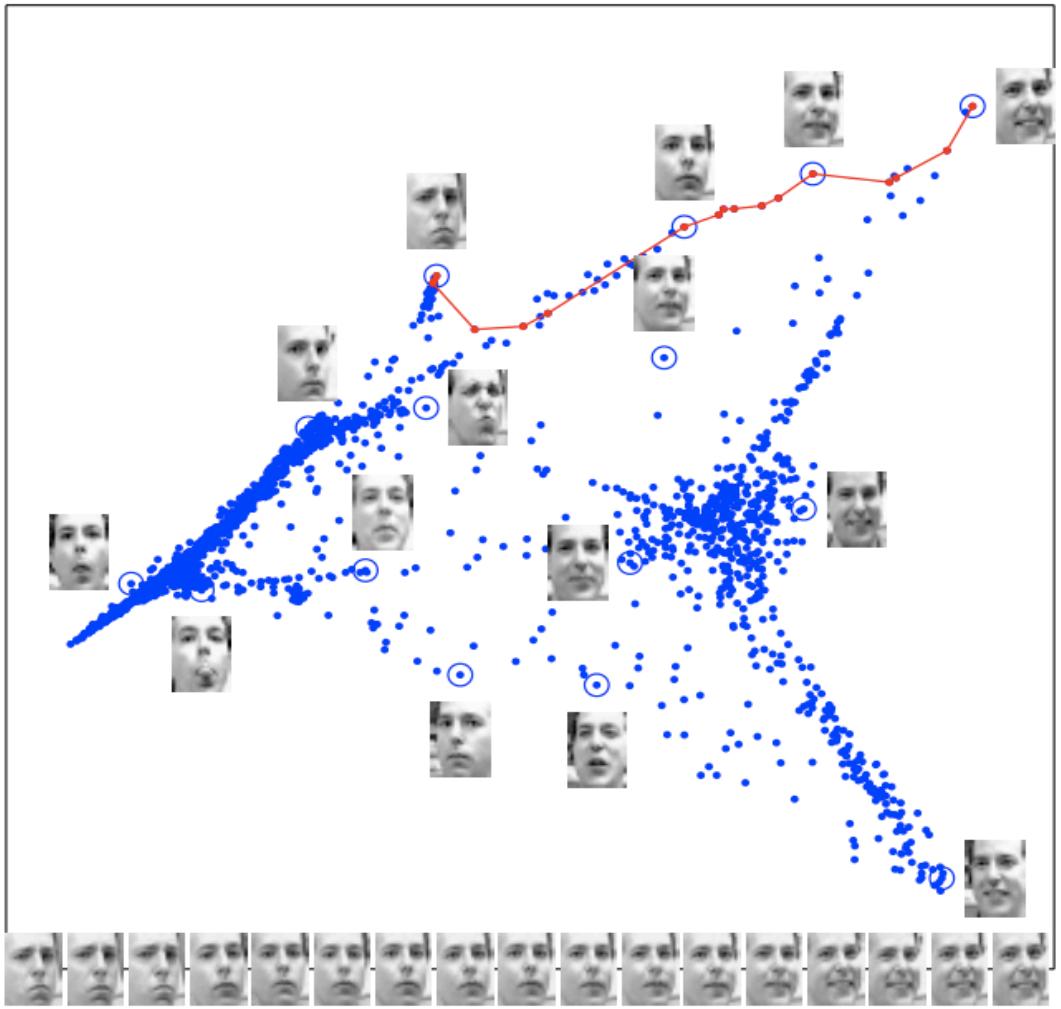
Unsupervised Learning - Embedding

Dimensionality Reduction

[Saul & Roweis '03]

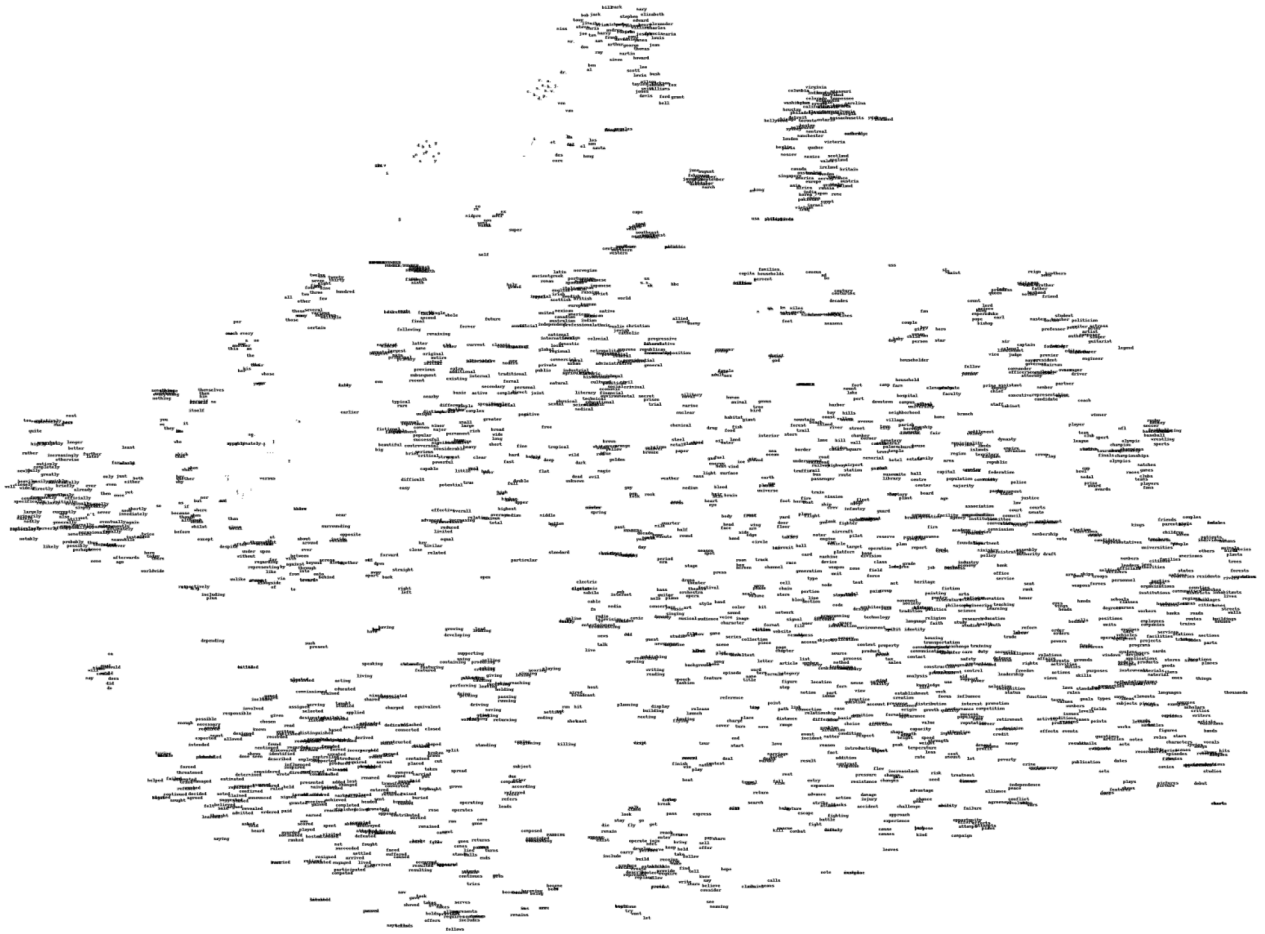
Images have thousands or millions of pixels.

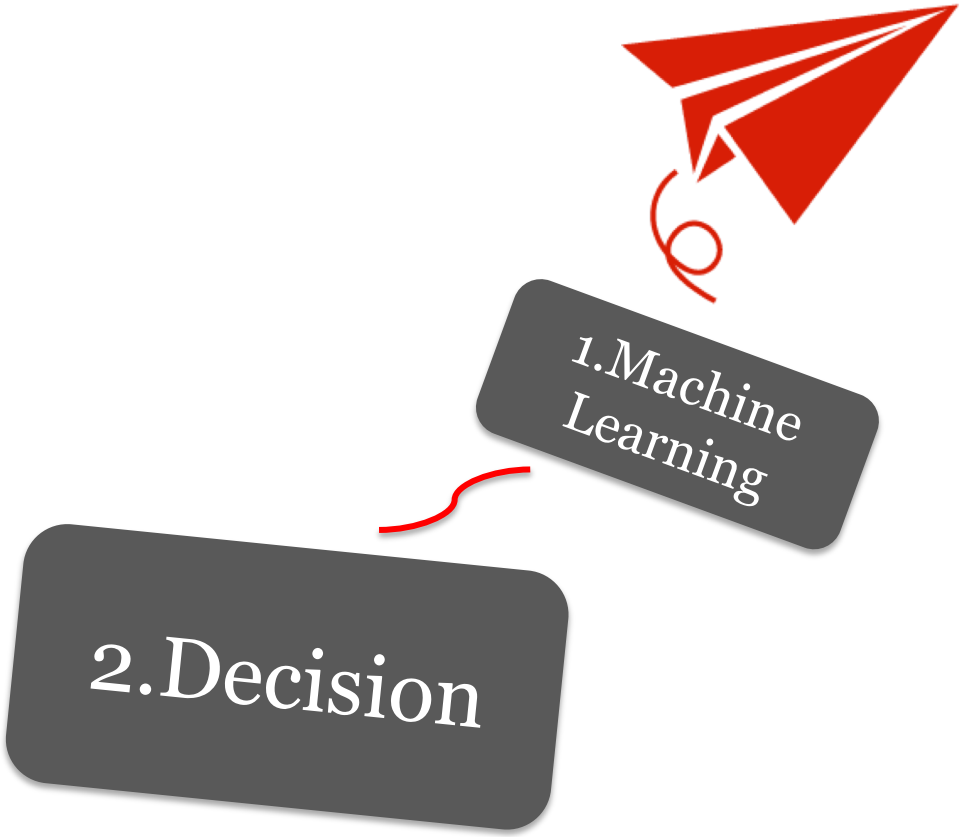
Can we give each image a coordinate,
such that similar images are near each other?



Unsupervised Learning - Embedding

Dimensionality Reduction - words [Joseph Turian]




1.Machine Learning
2.Decision

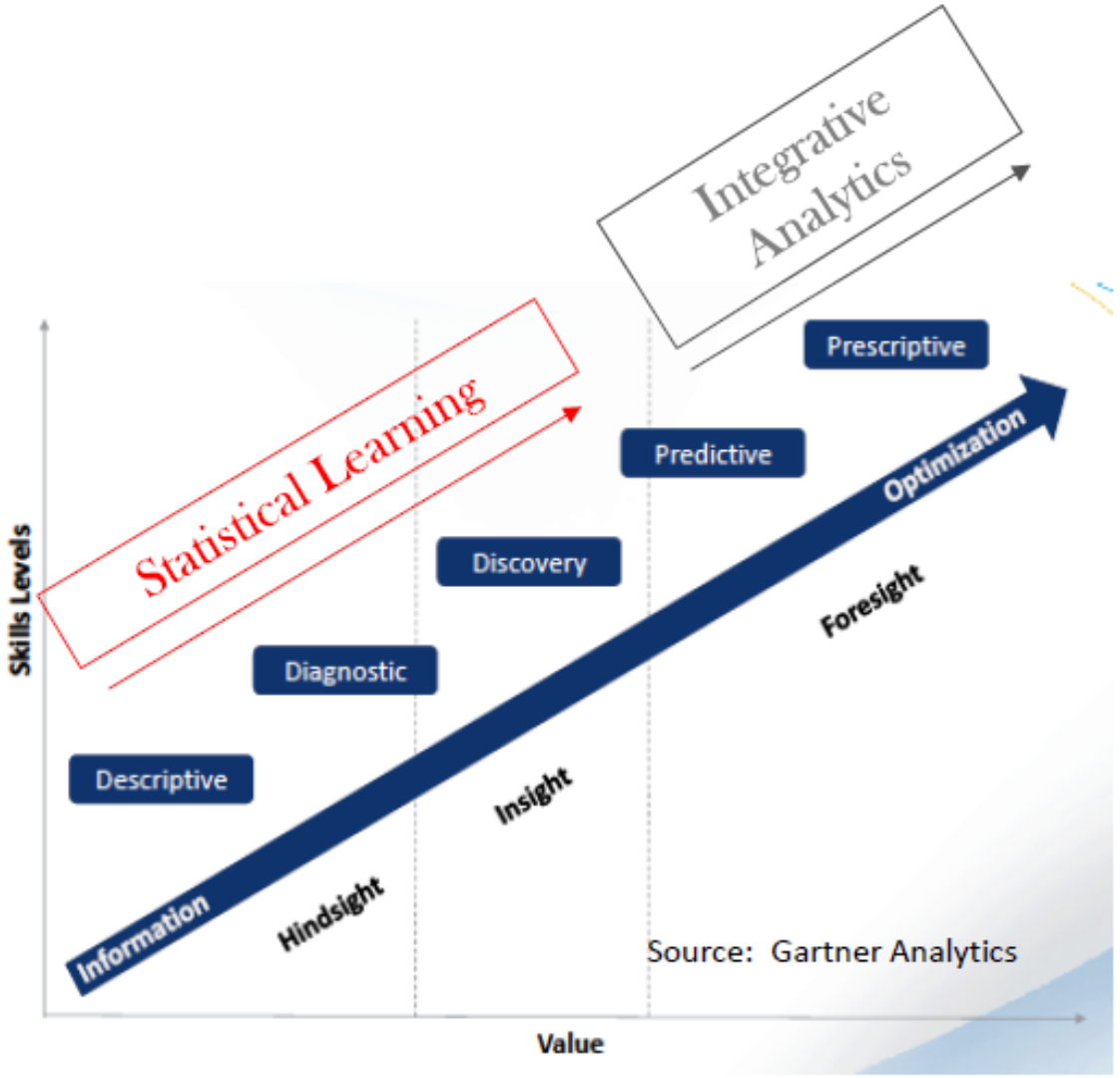
Decision-making is important for many business applications

- Decision on inventory levels
- Decision on facility locations
- Decision on routing
- Decision on price
- Decision on assortments in recommender systems
- Decision on portfolio construction

Big Data in Operations and Business

- Operations Research/Management: building stochastic **models** and trying to make optimal **decision**.
  - Build stochastic model relates price and demand
  - Assuming the demand is known, how to optimize for price
- Integrative Analytics: combine learning and decisions

The Spectrum of Analytics



Integrative Analytics Questions

- Learning Questions
- Given company data, is it more likely for a stock price to rise or fall?
- Given wind data by (location) for November morning, what is the predicted wind speed at noon today?

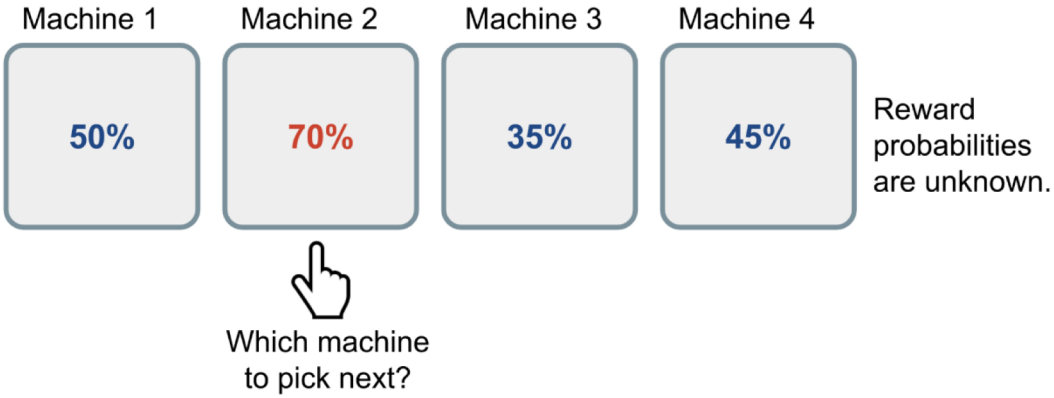
- Integrative Analytics Questions
- Given a model which predicts stock price changes, *design a portfolio*
- Given wind speed predictions, *what mix of generators "minimizes" costs and emissions?*

KPMG Survey (2014): Over 95% of CIO/CFOs (over billion dollar turnover) find that they are not as effective in Bridging the Gap Between Data and Decision-Making.

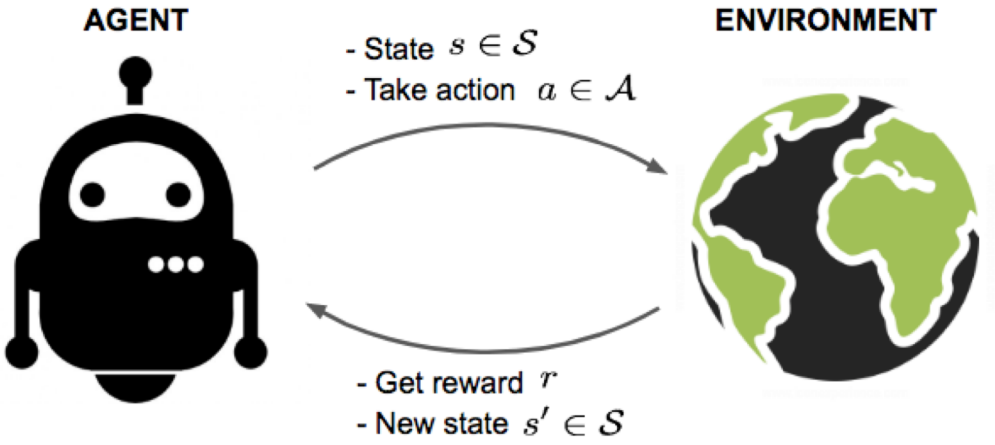
Source: Informs 2016

From Machine Learning to Decision-making

- Multi-armed Bandits



- Reinforcement Learning



Goal of this Class

- Not teaching a programming language
  - R for some demonstration
  - Any programming language is fine for homework/project (but some basic R commands is necessary)

- Unveil the black box of machine learning algorithms
  - Able to implement algorithm by yourself
  - Understand assumptions, strength, and weakness behind each algorithm

- Tools for theoretical analysis
  - Theoretical tools for establishing performance guarantee for a machine learning algorithm

Prerequisites

- Multivariate calculus: derivative, integration with respect to multivariate functions.
- Linear algebra and matrix analysis
- Probabilities
  - Distributions, densities, marginalization, law of large number, central limit theorem
- Statistics
  - Maximum likelihood estimation
- Proficient in one programming/latex
  - Mostly your choice of language, but R or Matlab will be very useful
- Ability to deal with “abstract mathematical concepts”
- Self-evaluation exam (uploaded)

Enroll in Compass2g

- https://compass2g.illinois.edu

- For course announcements, lecture slides, assignments, final projects, etc.

Tentative Syllabus

- Lecture 01: Introduction, Logistics
- Lecture 02~17: Foundations of Machine Learning
  - Introduction to Probabilities
  - Naïve Bayes, KNN, Linear Regression, Model Selection
  - Logistic Regression, SVM, Decision trees, Neural Networks
  - Semi-Supervised Learning
  - Clustering, Dimension Reduction PCA
  - Graphical Models, EM Algorithm, Hidden Markov Model

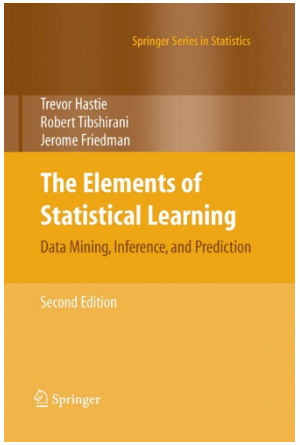
Tentative Syllabus (cont'd)

- Lecture 18~26: Online Learning and Decision Making
  - Introduction to Bandits, Exploration vs Exploitation
  - UCB Algorithm, Thompson Sampling
  - Contextual Bandits, Linear Contextual Bandits
  - Introduction to Reinforcement Learning
  - Model Free Planning/ Control, Value Function Approximations

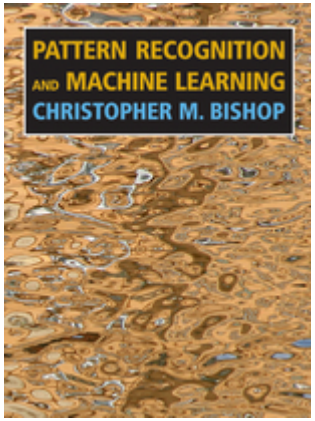
- Anticipate 2 canceled class meetings in April due to my travel plan

Textbooks (Part I)

- Recommended Textbook:
  - Elements of Statistical Learning by Trevor Hasite, Robert Tibshirani, and Jerome Friedman
    - http://web.stanford.edu/~hastie/ElemStatLearn

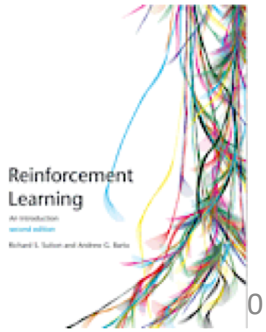


  - Pattern recognition and machine learning by Chris Bishop
    - https://www.microsoft.com/en-us/research/uploads/prod/2006/01/Bishop-Pattern-Recognition-and-Machine-Learning-2006.pdf



Textbooks (Part II)

- Recommended Textbook:
  - Bandit Algorithm by Tor Lattimore & Csaba Szepesvari
    - https://banditalgs.com/
    - https://tor-lattimore.com/downloads/book/book.pdf
  - Reinforcement Learning by Richard Sutton and Andrew Barto
    - http://incompleteideas.net/book/bookdraft2017nov5.pdf



Evaluation

- Homework (Programming Exercises) 60%
- Final Project 35%
- Participation 5%

Homework

- Programming homework assigned for 6~8 times
  - Submit the code (with annotations)
  - A concise report about your observations (including figures in your report)
  - Submit the latex code (NeurIPS templates or other templates) of the report

Final Project

- By individual or group of two
- Ideal project: combine machine learning with your own domain knowledge in business
- Submit the report and source code at the end of the semester (but starts earlier!)
- Examples
  - CMU machine learning class project guideline http://www.cs.cmu.edu/~tom/10701_sp11/proj.shtml
  - Stanford 2014 machine learning class project http://cs229.stanford.edu/projects2015.html
  - UCI Datasets https://archive.ics.uci.edu/ml/datasets.html

Garth Saloner @ Saloner 2016/9/30

“ If you’re a Stanford MBA student head over to the Engineering School and learn everything you can about AI, deep learning, automation. Now. ”

Enjoy!

- ML is becoming ubiquitous in science, engineering, business, and beyond
- This class should give you the basic foundation for applying ML and developing new methods
- The fun begins…

Slide Courtesy and Acknowledgement

- NYU Stern Course: From Machine Learning to Decision-Making with Applications to Business (by Xi Chen)
- Carnegie Mellon Course: Machine Learning (by Aarti Singh)
- Stanford Course: Introduction to Statistical Learning with Applications in R (by Daniela Witten, Trevor Hastie and Robert Tibshirani)
- John Hopkins Course: Practical Machine Learning (by Jeff Leek, Roger Peng and Brian Caffo)
- Michigan State Course: Machine Learning (by Rong Jin)
- Chicago Booth Course: Machine learning (by Carlos Carvalho, Mladen Kolar and Robert McCulloch)
- NYU Courant Course: Foundations of Machine Learning (by Mehryar Mohri)
- Tor Lattimore & Csaba Szepevari AAAI 2018 Tutorial on Bandit
- David Silver (google deep mind) Reinforcement Learning Class
- Lihong Li (google brain) reinforcement learning tutorial