ISTA 421/521
Introduction to Machine Learning

Clay Morrison
clayton@sista.arizona.edu
Gould-Simpson 819
Phone 621-6609

26 August 2014

Welcome to ISTA 421/521

- Today:
  - Introductions
  - Syllabus
  - Course Structure & Goals
  - Intro to ML

Your Instructors

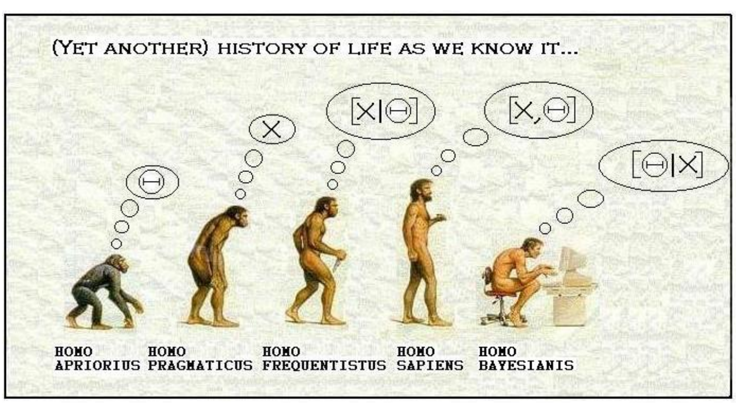
- **Clay Morrison**
  - GS 819, 621-6609, clayton@sista.arizona.edu
  - Office Hours:
    - By appointment (email me!)

- **Leon Palafox**
  - http://www.leonpalafox.com/
  - leonp@lpl.arizona.edu
  - Office Hours: Tues & Fri 1-2pm, GS 918

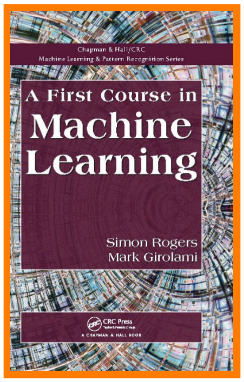
Please contact us ahead of time if you plan to attend office hours!

Course Goals

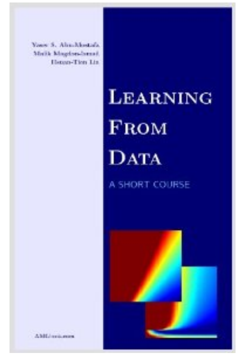
- Basic literacy in core, modern ML methods
- Practical experience implementing ML algorithms and using them on data

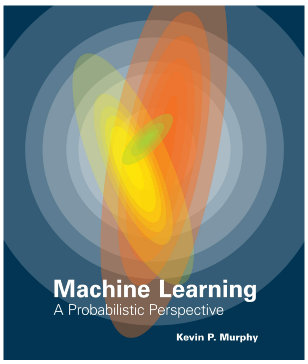


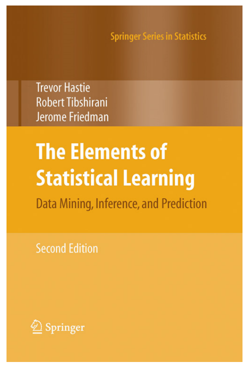
Text

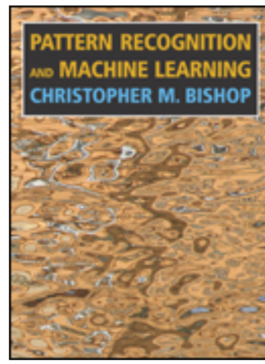


Tom Mitchell, 1997

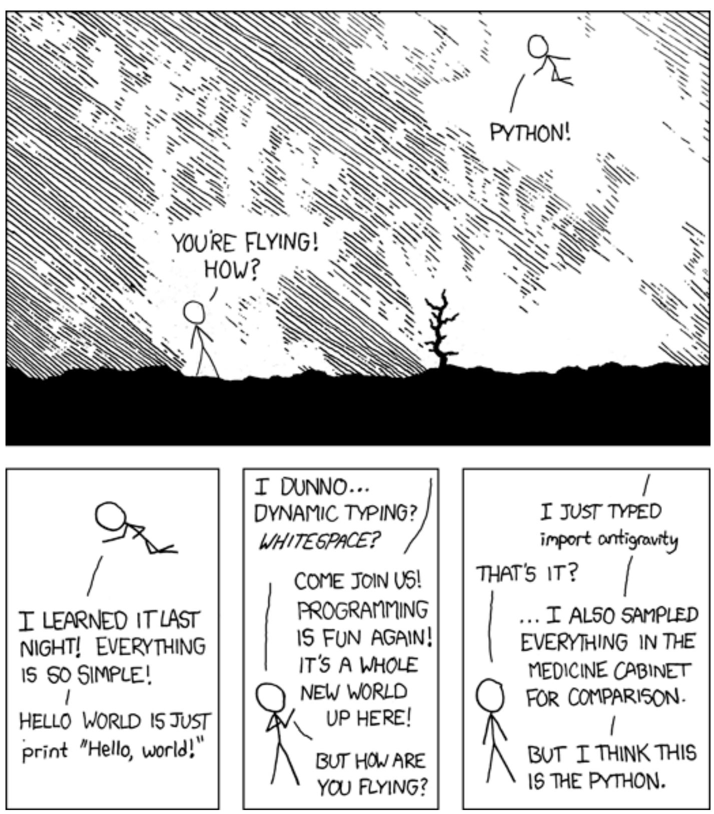








Python 2.7
- numpy
- scipy
- matplotlib



What is Machine Learning?

- The goal of machine learning is to build computer systems that can adapt and learn from their experience. (Dietterich, 1999)
- Machine learning usually refers to changes in systems that perform tasks associated with artificial intelligence. Such tasks involve recognition, diagnosis, planning, robot control, prediction, etc. (Nilsson, 1996)
- Some reasons for adaptation:
  - Some tasks can be hard to define except via examples
  - Adaptation can improve a human-built system, or track changes over time
- Goals can be autonomous machine performance, or enabling humans to learn from and understand data (data mining and modeling)

Ack: this and some following content adapted from Chris Williams 2006

Some of the Roots of Machine Learning

- Philosophy: epistemology, philosophy of science, logical inference: the Problem of Induction
- Mathematics, Physics
- Statistics
- Psychological models (of learning and development)
- Brain models, e.g. neural networks
- Artificial Intelligence: e.g., discovering rules using decision trees, inductive logic programming, autonomy
- Engineering: Statistical pattern recognition, adaptive control theory

From Rob Tibshirani

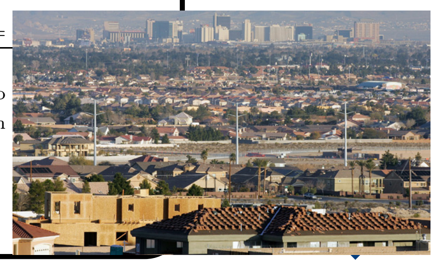
Greater focus on
- prediction
- analysis of learning algorithms (not just large dataset issues)

What is the difference between
Machine Learning and Statistics?

Glossary

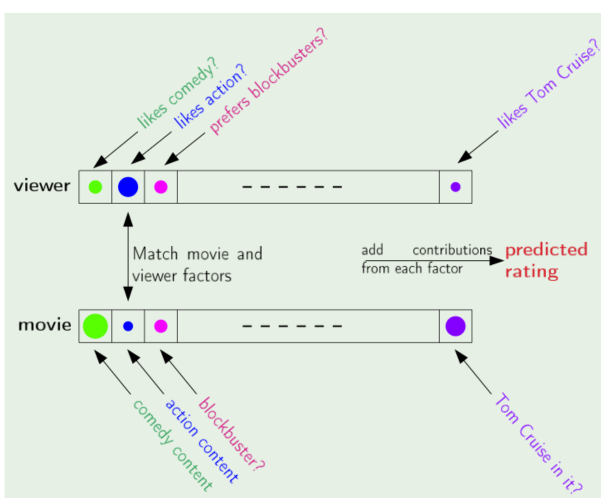
Machine learning	Statistics
network, graphs	model
weights	parameters
learning	fitting
generalization	test set performance
supervised learning	regression/classification
unsupervised learning	density estimation, clustering
e grant = $1,000,000	large grant=
place to have a meeting: vbird, Utah, French Alps	nice place to Las Vegas in

Greater focus on
- understanding data in terms of models
- interpretability, hypothesis testing



Common use-case of machine learning

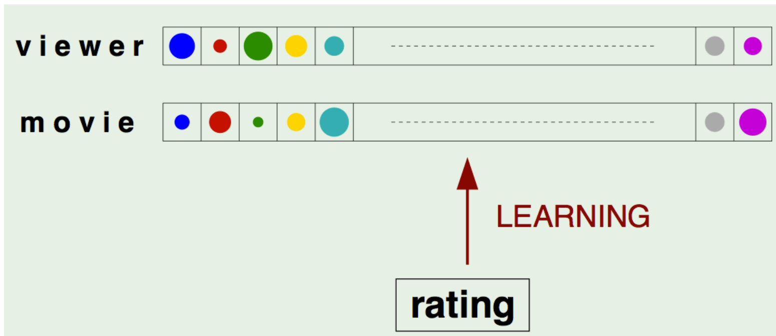
1. A pattern exists
2. We don't have a closed-form, analytic model or rule
3. We have data!



Adapted from Yaser S. Abu-Mostafa et al., *Learning from Data*

Common use-case of machine learning

1. A pattern exists
2. We don't have a closed-form, analytic model or rule
3. We have data!



Adapted from Yaser S. Abu-Mostafa et al., *Learning from Data*

Another Example

- Credit approval
- Applicant information:

age	23 years
gender	male
annual salary	\$30,000
years in residence	1 year
years in job	1 year
current debt	\$15,000
···	···

- Approve credit?

Adapted from Yaser S. Abu-Mostafa et al., *Learning from Data*

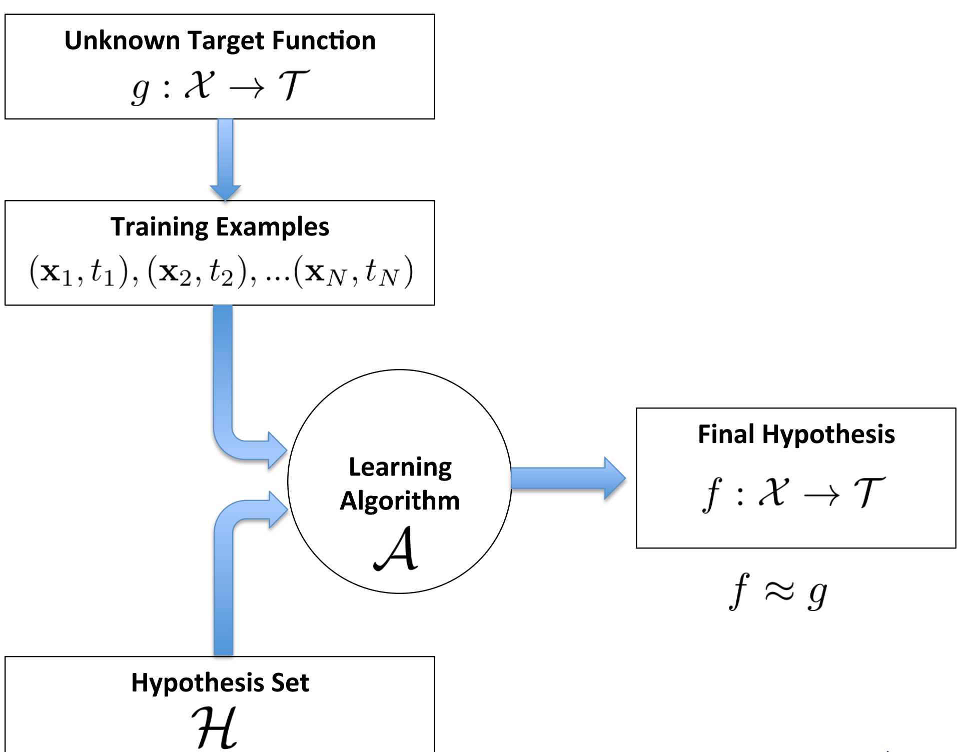
Some Terminology

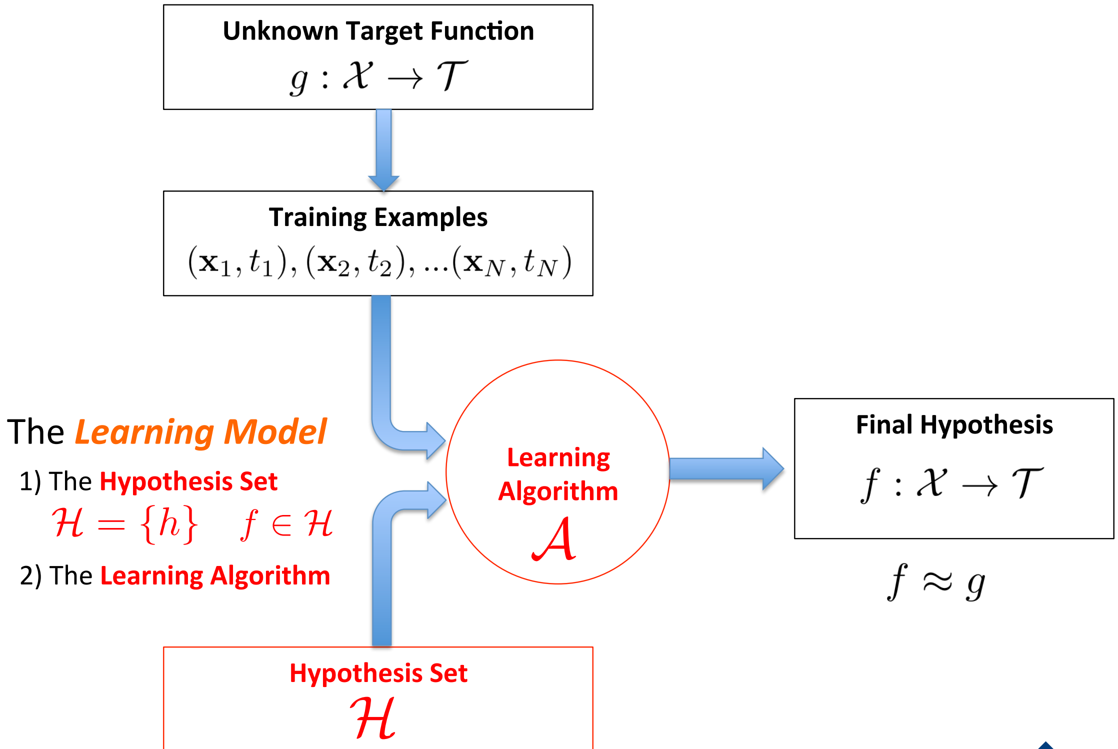
- Input: $\mathbf{x}$ (customer application)
- Output: t (good/bad customer?)
- Target function: $g : \mathcal{X} \rightarrow \mathcal{T}$ (*ideal* credit approval fn)
- Data: $(\mathbf{x}_1, t_1), (\mathbf{x}_2, t_2), ...(\mathbf{x}_N, t_N)$ (historical records)

learning

- Hypothesis: $f : \mathcal{X} \rightarrow \mathcal{T}$ (formula to be used)

Adapted from Yaser S. Abu-Mostafa et al., *Learning from Data*



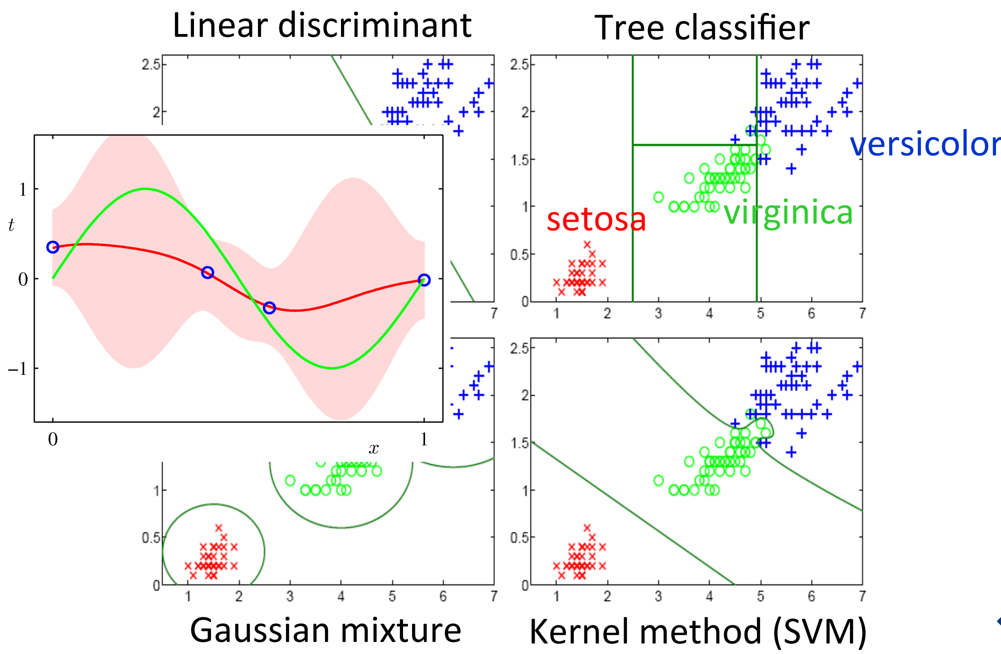


Some Learning Models

- Linear models
- Kernel methods
- Neural networks
- Decision trees

Iris Data (Fisher, 1936)

Figure from Norbert Jankowski and Krzysztof Grabczewski



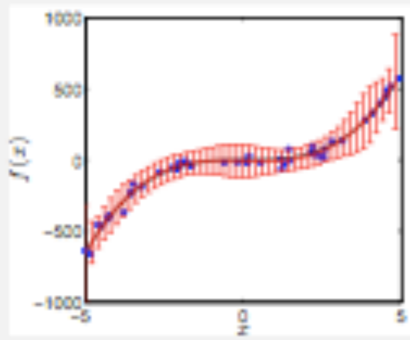
Three General Classes of ML

- **Supervised learning** – model $p(y|\mathbf{x})$
  - Given data and model, or data with correct output (label)
  - Regression, Classification, etc.
- **Unsupervised Learning** – model $p(\mathbf{x})$
  - Only given input data (no output)
  - Clustering, Latent Models, Projection methods, etc.
- **Reinforcement Learning** – model $p(\mathbf{s}_{t+1}|\mathbf{s},a)$
  - Given input data, *some* output, and *grade* for output
  - Learning to choose better actions
  - Markov decision processes, POMDPs, planning

Supervised Learning

Regression

Learning a continuous function from a set of examples.



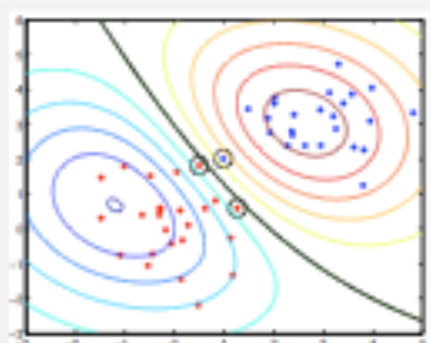
Example

Predicting stock prices (x might be time or some other variable of interest).

Supervised Learning

Classification

Learning a rule that can separate objects of different types from one another.



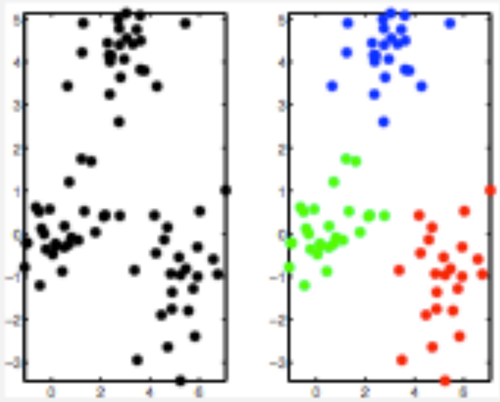
Examples

Disease diagnosis, spam email detection.

Unsupervised Learning

Clustering

Finding groups of similar objects.



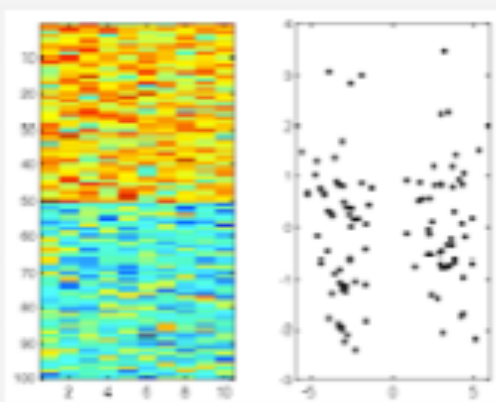
Examples

People with similar 'taste', genes with similar function.

Unsupervised Learning

Projection

Reducing the number of variables – e.g. from 10 to 2.

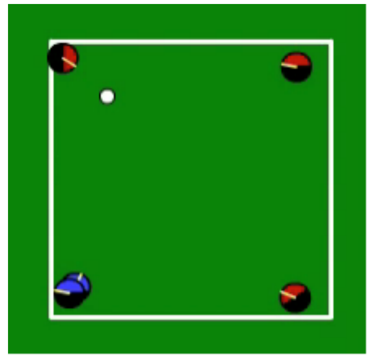


Examples

Visualising complex data.

Reinforcement Learning

- Example: Simulated Robot Soccer Keepaway
- Input: angles, distances, clock



- What is the best next action to take given what I know now?
- Allows system to learn by interacting with environment with human only providing tips about "goodness"

Topics

- The linear model
  – Regression, Classification
- Classification
  – Probabilistic:
    - Bayes Classifier, Naïve Bayes
    - Logistic Regression
  – Other, *non*-probabilistic
    - K-nearest neighbors
    - Support Vector Machines and kernel methods
- Clustering
  – K-means
  – Mixture Models and EM
- Other Unsupervised methods:
  – Principle Components Analysis
  – Latent Variable Models
- Additional topics (time permitting)
  – Neural networks, Deep networks
  – Ensemble methods, Boosting
  – Gaussian processes
- Probability
  – Quantifying uncertainty
  – Bayesian Approach: Prior, Marginal Likelihood, MAP
- Inference Methods
  – Least Squares
  – Maximum Likelihood
  – Bayesian Inference: Direct and Sampling
- Machine Learning algorithm evaluation
- Learning theory
- Feature Selection and Model Selection