CSE 446
Machine Learning

Daniel Weld
Xiao Ling
Congle Zhang

What is Machine Learning ?

Machine Learning

Study of algorithms that

- improve their performance
- at some task
- with experience

Why?

Is this topic important?

Exponential Growth in Data

Supremacy of Machine Learning

- Machine learning is preferred approach to
  - Speech recognition, Natural language processing
  - Web search – result ranking
  - Computer vision
  - Medical outcomes analysis
  - Robot control
  - Computational biology
  - Sensor networks
  - …
- This trend is accelerating
  - Improved machine learning algorithms
  - Improved data capture, networking, faster computers
  - Software too complex to write by hand
  - New sensors / IO devices
  - Demand for self-customization to user, environment

Logistics

Syllabus

- Covers a wide range of Machine Learning techniques – from basic to state-of-the-art
- You will learn about the methods you heard about:
  - Naïve Bayes, logistic regression, nearest-neighbor, decision trees, boosting, neural nets, overfitting, regularization, dimensionality reduction, error bounds, loss function, VC dimension, SVMs, kernels, margin bounds, K-means, EM, mixture models, semi-supervised learning, HMMs, graphical models, active learning
- Covers algorithms, theory and applications
- **It's going to be fun and hard work ☺**

Prerequisites

- Probabilities
  - Distributions, densities, marginalization…
- Basic statistics
  - Moments, typical distributions, regression…
- Algorithms
  - Dynamic programming, basic data structures, complexity…
- Programming
  - Mostly your choice of language, but Python (NumPy) + Matlab will be very useful
- We provide some background, but the class will be fast paced
- Ability to deal with "abstract mathematical concepts"

Staff

- Two Great TAs:
  Fantastic resource for learning, interact with them!
  - **Xiao Ling**, CSE 610, xiaoling@cs
  - Office hours: TBA

  - **Congle Zhang**, CSE 524, clzhang@cs
  - Office hours: TBA

- Administrative Assistant
  - Alicen Smith, CSE 546, asmith@cs

Text Books

- Required Text:
  - Pattern Recognition and Machine Learning; Chris Bishop
- Optional:
  - Machine Learning; Tom Mitchell
  - The Elements of Statistical Learning: Data Mining, Inference, and Prediction; Trevor Hastie, Robert Tibshirani, Jerome Friedman
  - Information Theory, Inference, and Learning Algorithms; David MacKay
  - Website: Andrew Ng's AI class videos
  - Website: Tom Mitchell's AI class videos

Grading

- 4 homeworks (55%)
  - First one goes out Fri 1/6/12
    - Start early, Start early, Start early, Start early, Start early, Start early, Start early, Start early, Start early, Start early
- Midterm (15%)
  - Circa Feb 10 in class
- Final (30%)
  - TBD by registrar

Homeworks

- Homeworks are hard, start early ☺
- Due at the beginning of class
  - Minus 33% credit for each day (or part of day) late
- All homeworks **must be handed in**, even for zero credit
- Collaboration
  - You may **discuss** the questions
  - Each student writes their own answers
  - Write on your homework anyone with whom you collaborate
  - Each student must write their own code for the programming part
  - **Please don't search for answers on the web, Google, previous years' homeworks, etc.**
    - Ask us if you are not sure if you can use a particular reference

Communication

- Main discussion board
  - https://catalyst.uw.edu/gopost/board/xling/25219/
- Urgent announcements
  - cse446@cs
  - Subscribe: http://mailman.cs.washington.edu/mailman/listinfo/cse446
- To email instructors, always use:
  - cse446_instructor@cs

Space of ML Problems

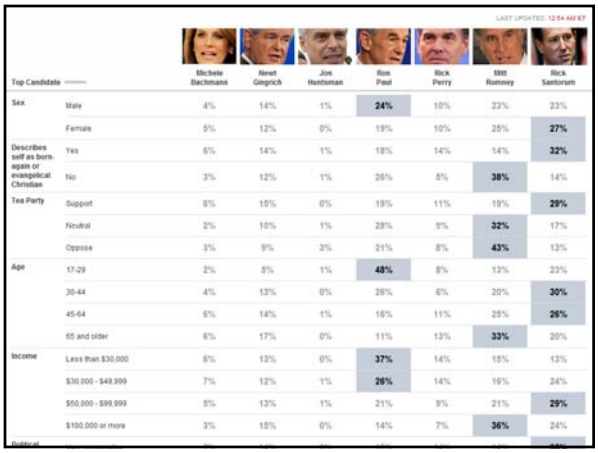
Type of Supervision
(eg, Experience, Feedback)

What is Being Learned?

	Labeled Examples	Reward	Nothing
Discrete Function	Classification		Clustering
Continuous Function	Regression		
Policy	Apprenticeship Learning	Reinforcement Learning	

Classification

from data to discrete classes



Spam filtering

data | prediction

Text classification

→ Company home page

vs

Personal home page

vs

Univeristy home page

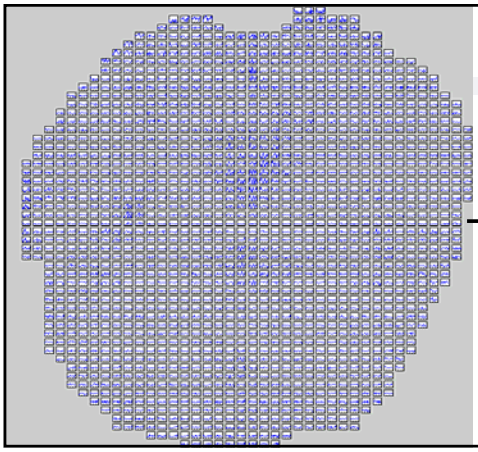
vs

…

Object detection

(Prof. H. Schneiderman)

Example training images for each orientation



Reading a noun (vs verb)

[Rustandi et al., 2005]

Weather prediction

The classification pipeline

Training

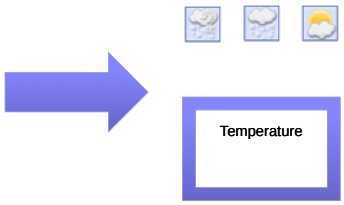
Testing

Regression

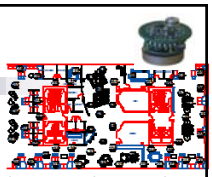
predicting a numeric value

Stock market

Weather prediction revisted



Modeling sensor data



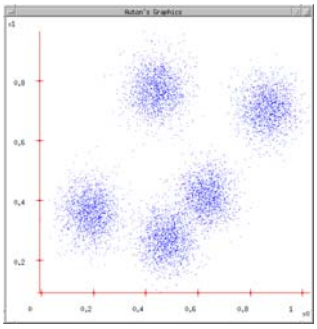
- Measure temperatures at some locations
- Predict temperatures throughout the environment

[Guestrin et al. '04]

Clustering

discovering structure in data

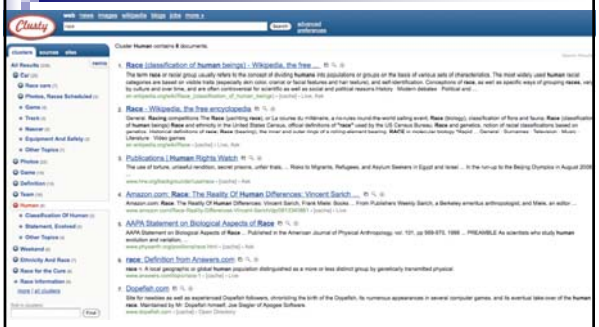
Clustering Data: Group similar things



Clustering images

[Goldberger et al.]

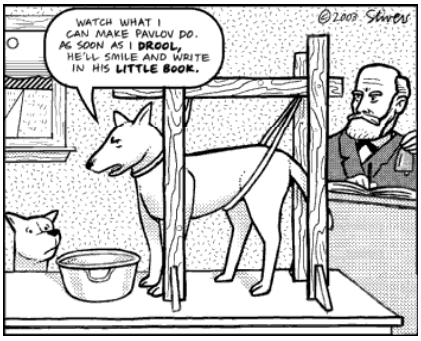
Clustering web search results



Reinforcement Learning

training by feedback

Reinforcement Learning



Learning to act

- Reinforcement learning
- An agent
  - Makes sensor observations
  - Must select action
  - Receives rewards
    - positive for "good" states
    - negative for "bad" states

[Ng et al. '05]

In Summary

Type of Supervision
(eg, Experience, Feedback)

What is Being Learned?

	Labeled Examples	Reward	Nothing
Discrete Function	Classification		Clustering
Continuous Function	Regression		
Policy	Apprenticeship Learning	Reinforcement Learning	

In Summary

Type of Supervision
(eg, Experience, Feedback)

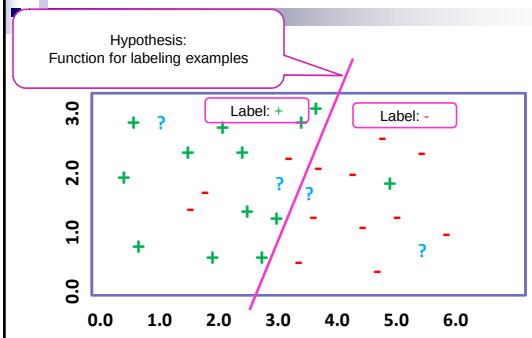
What is Being Learned?

	Labeled Examples	Reward	Nothing
Discrete Function	Classification		Clustering
Continuous Function	Regression		
Policy	Apprenticeship Learning	Reinforcement Learning	

Key Concepts

38

Classifier



Generalization

- Hypotheses must ***generalize*** to correctly classify instances not in the training data.
- Simply memorizing training examples is a consistent hypothesis ***that does not generalize***.

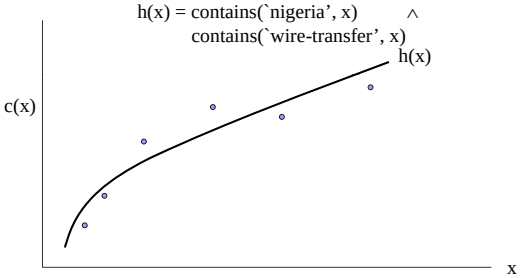
40

ML = Function Approximation

May not be any perfect fit

Classification ~ discrete functions

h(x) = contains(`nigeria', x) ∧ contains(`wire-transfer', x)



41

Why is Learning Possible?

Experience alone never justifies any conclusion about any unseen instance.

Learning occurs when

PREJUDICE meets DATA!

Learning a "Frobnitz"

42

Bias

- The nice word for prejudice is "bias".
  - Different from "Bias" in statistics
- What kind of hypotheses will you ***consider***?
  - What is allowable ***range*** of functions you use when approximating?
- What kind of hypotheses do you ***prefer***?

43

Some Typical Biases

- Occam's razor
  - *"It is needless to do more when less will suffice"*
  - *– William of Occam,*
  - *died 1349 of the Black plague*
- MDL – Minimum description length
- Concepts can be approximated by
- ... **conjunctions** of predicates
  - ... by **linear** functions
  - ... by **short** decision trees

Frobnitz?

ML as Optimization

- Specify Preference Bias
  - aka "Loss Function"
- Solve using optimization
  - Combinatorial
  - Convex
  - Linear
  - Nasty

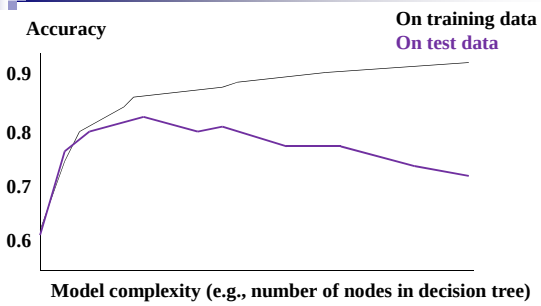
Overfitting

- Hypothesis H is *overfit* when ∃ H' and
  - H has ***smaller*** error on training examples, but
  - H has ***bigger*** error on test examples

Overfitting

- Hypothesis H is *overfit* when ∃ H' and
  - H has ***smaller*** error on training examples, but
  - H has ***bigger*** error on test examples
- Causes of overfitting
  - Training set is too small
  - Large number of features
- Big problem in machine learning
  - One solution: Validation set

Overfitting



The Road Ahead