Machine Learning

Slides use Tony Jebara's materials

Topic 1

- Introduction
- Machine Learning: What, Why and Applications
- Syllabus, policies, texts, web page
- Historical Perspective
- Machine Learning Tasks and Tools
- Digit Recognition Example
- Machine Learning Approach
- Deterministic or Probabilistic Approach
- Why Probabilistic?

Machine Learning: What/Why

Statistical Data-Driven Computational Models

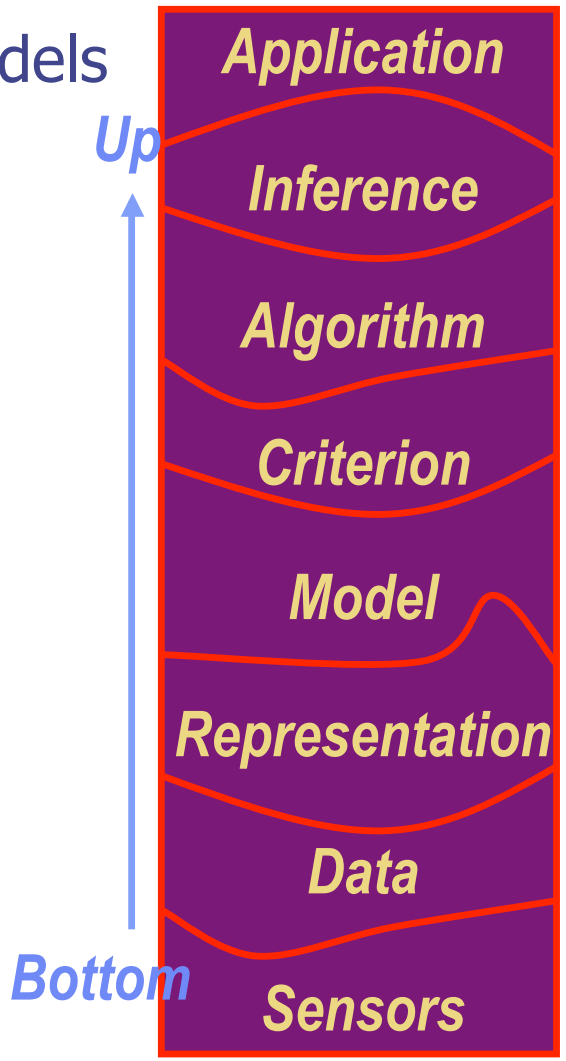
Real domains (vision, speech, behavior):
- no $E=MC^2$
- noisy, complex, nonlinear
- have many variables
- non-deterministic
- incomplete, approximate models

Need: statistical models driven by data & sensors, a.k.a Machine Learning

Bottom-Up: use data to form a model

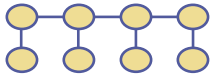
Why? Complex data everywhere, audio, video, internet

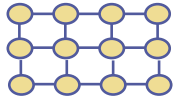
Intelligence = Learning = Prediction



Machine Learning Applications

- ML: Interdisciplinary (CS, Math, Stats, Physics, OR, Psych)
- Data-driven approach to AI
- Many domains are too hard to do manually

Speech Recognition (HMMs, ICA) 

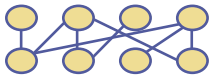
Computer Vision (face rec, digits, MRFs, super-res) 

Time Series Prediction (weather, finance) 

Genomics (micro-arrays, SVMs, splice-sites)

NLP and Parsing (HMMs, CRFs, Google)

Text and InfoRetrieval (docs, google, spam, TSVMs)

Medical (QMR-DT, informatics, ICA) 

Behavior/Games (reinforcement, gammon, gaming)

Course Details & Requirements

- Probability/Stats, Linear Algebra, Calculus, AI
- Mathematical & Data Driven approach to AI
- Lots of Equations!

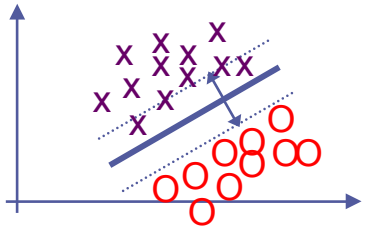
- Required Text: Introduction to Graphical Models by M. Jordan & C. Bishop (Online)
  Pattern Recognition & Machine Learning by C. Bishop (Spring 2006 Edition)
- Reference Text: Pattern Classification (3rd Edition) by Duda, Hart and Stork

- Homework: Every 2-3 weeks
- Grading: homework, midterm, 2 quizzes & final examination

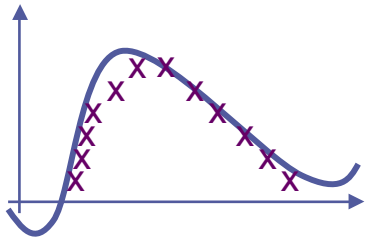
- Software Requirements: Matlab software & Acis account

Machine Learning Tasks

Classification y=sign(f(x))

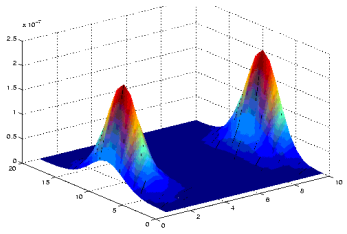


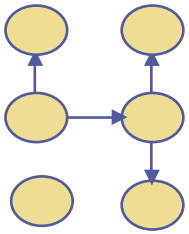
Regression y=f(x)



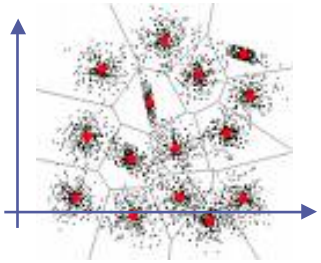
Supervised

Modeling p(x)

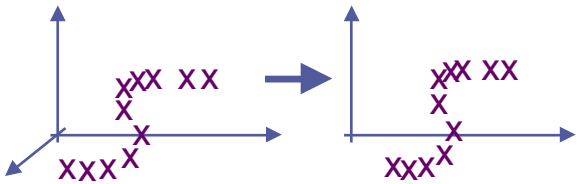




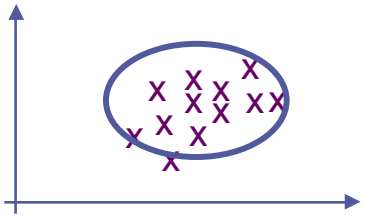
Clustering



Feature Selection

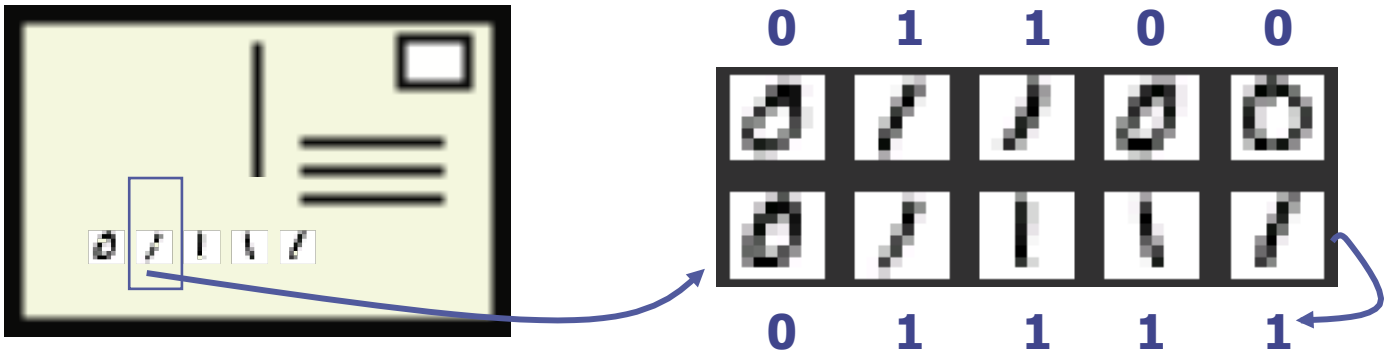


Detection p(x)<t

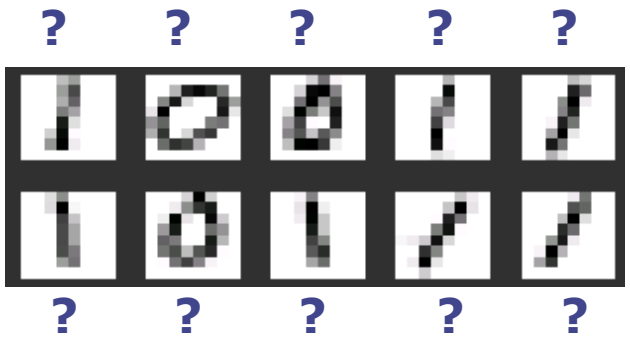


Unsupervised

ML Example: Digit Recognition



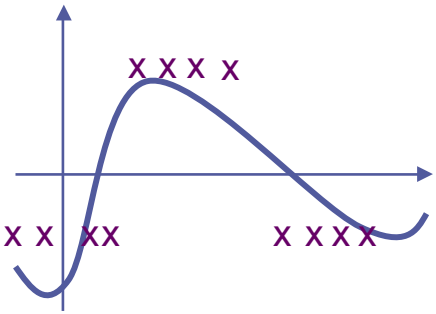
- Want to automate zipcode reading in post office
- Look at an image and say if it is a '1' or '0'
- 8x8 pixels of gray-level (0.0=dark, 0.5=gray, 1.0=white)
- Learn from above labeled training images
- Predict labels on testing images
- Binary Classification [0,1]
- What to do?



Ex: Two Approaches

In ML, we will consider two complementary approaches:

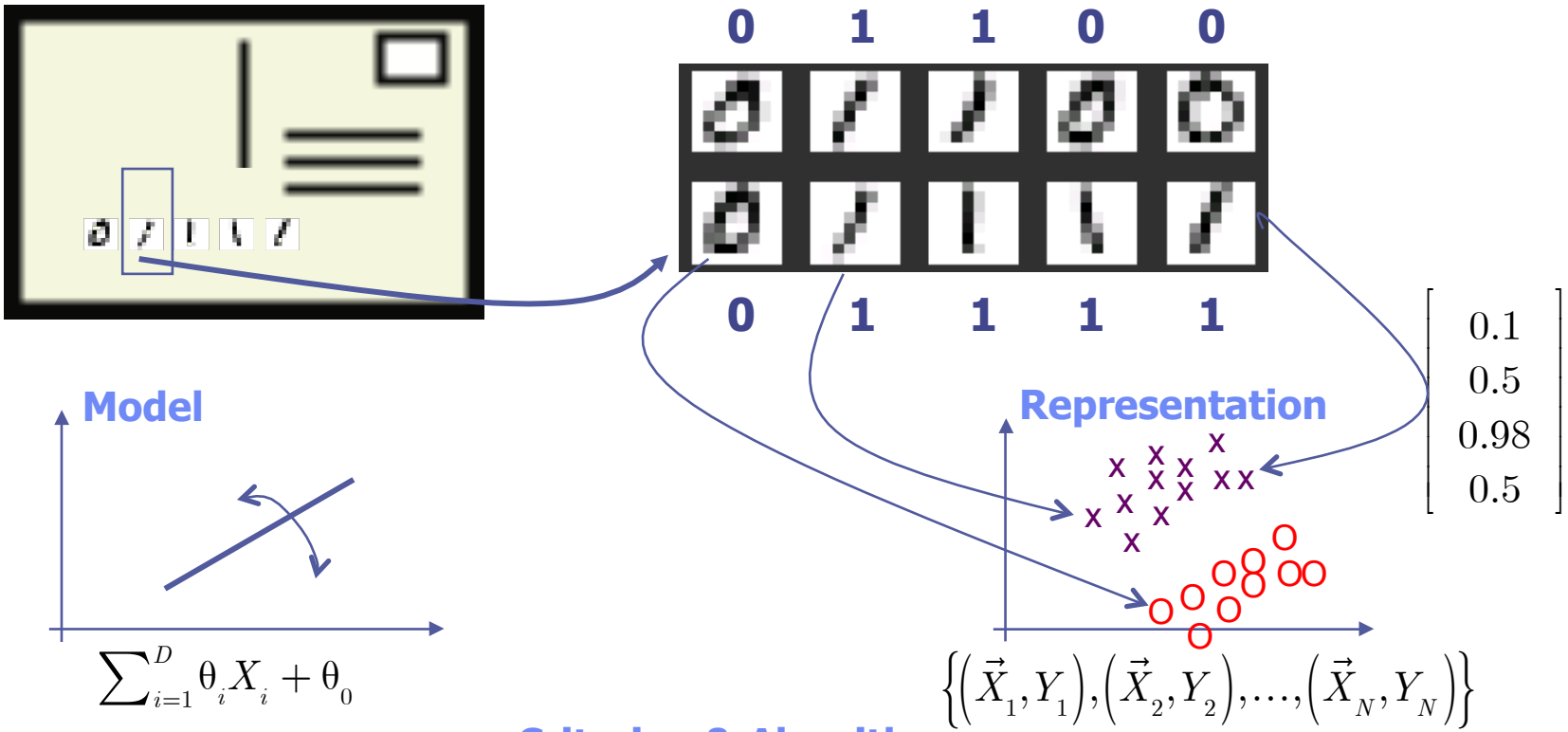
1) Deterministic:
   All variables/observables are treated as certain/exact
   Find/fit a function f(X) on an image X
   Output 0 or 1 depending on input
   Class label given by y=sign(f(X))/2 + 1/2



2) Probabilistic/Bayesian/Stochastic:
   Variables/observables are random (R.V.) and uncertain
   Probability image is a '0' digit: p(y=0|X) = 0.43
   Probability image is a '1' digit: p(y=1|X) = 0.57
   Output label with larger p(y=0|image) or p(y=1|image)

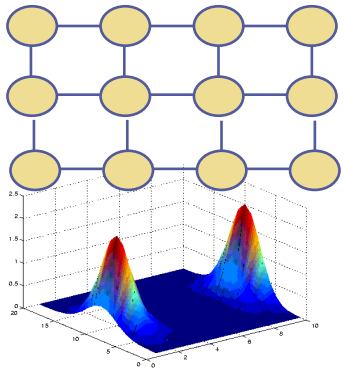
These are interconnected! Deterministic approaches can be generated from (more general) probabilistic approaches

Ex: 1) Deterministic Approach

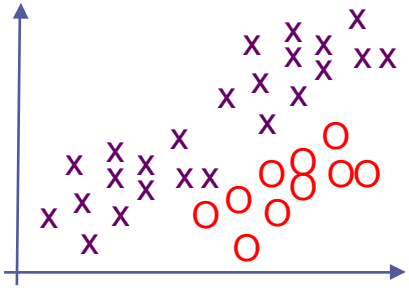


Ex: 2) Probabilistic Approach

a) Provide Prior Model
Parameters & Structure
e.g. nearby pixels are
co-dependent

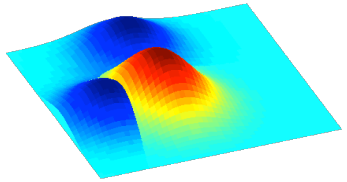


b) Obtain Data and Labels $\left\{\left(X_1, Y_1\right),\ldots,\left(X_T, Y_T\right)\right\}$



c) Learn a probability model with data
p(all system variables)

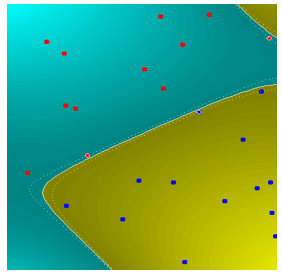
$p\left(X, Y\right)$



d) Use model for inference (classify/predict)

Probability image is '0': p(y=0|X)
Probability image is '1': p(y=1|X)
Output: arg max$_i$ p(y=i|X)

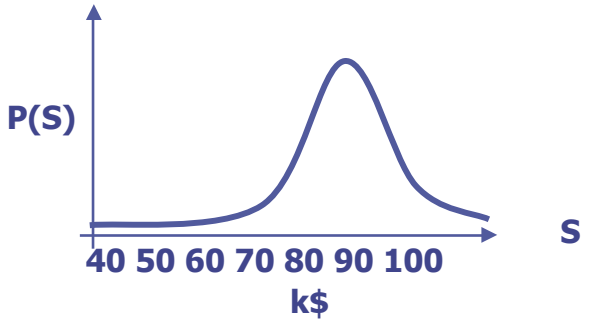
$p\left(Y \mid X\right)$



Why Probabilistic Approach?

- Decision making often involves uncertainty
- Hidden variables, complexity, randomness in system
- Input data is noisy and uncertain
- Estimated model is noisy and uncertain
- Output data is uncertain (no single correct answer)

- Example: Predict your salary in the future
- Inputs: Field, Degree, University, City, IQ
- Output: $Amount
- There is uncertainty and hidden variables
- No one answer (I.e. $84K) is correct
- Answer = a distribution over salaries



Why Probabilistic? Monty Hall

- Behind one door is a prize (car? 1$?)
- Pick a door

Door A Door B Door C

Monty Hall Solution

Probabilistic Graphical Model Bayesian Network

Probabilistic Interpretation is Best

Bayesian Solution: Change your mind!

Assume we always start by picking A.

If prize behind A: Opens B/C → Change A to C/B → Lose

If prize behind B: Opens C → Change A to B → Win

If prize behind C: Opens B → Change A to C → Win

Probability of winning if change your mind = 66%
Probability of winning if stick to your guns = 33%