

Machine Learning (CS 567)

Fall 2008

Time: T-Th, 5:00pm - 6:20pm

Location: GFS118

Instructor: Sofus A. Macskassy (macskass@usc.edu)

Office: SAL 216

Office hours: by appointment

Teaching assistant: Cheol Han

Office hours: TBA

Class web page:

<http://www-scf.usc.edu/~csci567/index.html>

Course Goals

- Introduce the background of machine learning
- Show core ML techniques
- Show proper experimental ML methodologies
- Prepare you to be able to use ML tools, contribute to the field

Not. Find out details of the state of the art...

Course Overview

- Introduction and Basics
 - Basic problems and questions in Machine Learning. Example applications.
- Linear Classifiers
 - Perceptron, Logistic Regression, Linear Discriminant Analysis (LDA)
- Five Popular Algorithms
 - Decision Trees (C4.5)
 - Neural Networks (backpropagation)
 - Probabilistic networks (Naïve bayes)
 - Nearest Neighbor (k -NN)
 - Support Vector Machines
- Theories of Learning
 - PAC, Bayesian, Bias-Variance Analysis
- Optimizing Test Set Performance
 - Overfitting, Penalty Methods, Holdout Methods, Ensembles
- Evaluation and Methodologies
 - Statistical tests, comparison of models and algorithms, confidence bounds

Administrative issues

- Class materials
 - *Introduction to Machine Learning*, Ethem Alpaydin, 2004, MIT Press
 - Class notes are meant to complement readings
- Mailing List
 - Please send me (macskass@usc.edu) mail with the following subject line:
cs567 student

Administrative issues

- Prerequisites
 - Basic linear algebra and calculus
 - Some knowledge of probabilities and statistics
 - Some AI background is recommended *but not required*
 - The Weka machine learning toolkit will be used in assignments. *Prior knowledge not required*

Administrative

- Please stop me if I talk too fast
- Please stop me if you have any questions
- *Any* feedback is better than *no* feedback
 - Sooner is better
 - End of the semester is too late to make it easier on you

Evaluation

- Weekly quizzes 10%
- 5 homework assignments 20%
- Midterm 20%
- Final Exam (not cumulative) 20%
- Class Project 30%
 - Groups of 1-4 people
 - Research paper to be written
 - Presentations at the last two classes
 - Details later in the course

Evaluation

- Grades
- Based on absolute scores
 - A 90.0%
 - A- 87.5%
 - B+ 85.0%
 - B 80.0%
 - B- 77.5%
 - C+ 75.0%
 - C 70.0%

Lecture 1 Outline

- The *what* and *why* of machine learning
- Why study machine learning
- Types of machine learning
- Supervised learning defined

What is learning?

- H.Simon: Any process by which a system improves its performance
- M.Minsky: Learning is making useful changes in our minds
- Michalsky: Learning is constructing or modifying representations of what is being experienced
- Valiant: Learning is the process of knowledge acquisition in the absence of explicit programming
- **Generally:** A process by which a computer algorithm finds an approximate solution to a problem

What is The Learning Problem?

Learning = Improving with experience at some task

- Improve over task T ,
- with respect to performance measure P ,
- based on experience E .

E.g., Learn to play checkers

- T : Play checkers
- P : % of games won in world tournament
- E : opportunity to play against self

Compare to Human Learning

- Psychologists use “learning” somewhat differently.
- Typically the task changes, whereas in machine learning, typically we assume the task is stationary.
- Seems like a pretty fundamental difference!

Why study machine learning?

- Easier to build a learning system than to hand-code a working program. E.g.:
 - Robot that learns a map of the environment by wandering around it
 - Programs that learn to play games by playing against themselves or others
- Improving on existing programs, e.g.
 - Instruction scheduling and register allocation in compilers
 - Combinatorial optimization problems
- Discover knowledge and patterns in highly dimensional, complex data

Why study machine learning?

- Solving tasks that require a system to be adaptive, e.g.
 - Speech and handwriting recognition
 - “Intelligent” user interfaces
- Understanding animal and human learning
 - How do we learn language?
 - How do we recognize faces?
- Creating “real” AI!
 - An expert system—brilliantly designed, engineered and implemented—cannot learn not to repeat its mistakes.

Time is right to study machine learning

- Recent progress in algorithms and theory
- Growing flood of online data
- Computational power is available
- Growing industry demand

Three example ML tasks

- Data mining: using historical data to improve decisions
 - medical records → medical knowledge
- Software applications we can't program by hand
 - autonomous driving
 - speech recognition
- Self customizing programs
 - Newsreader that learns user interests

Typical Learning Task

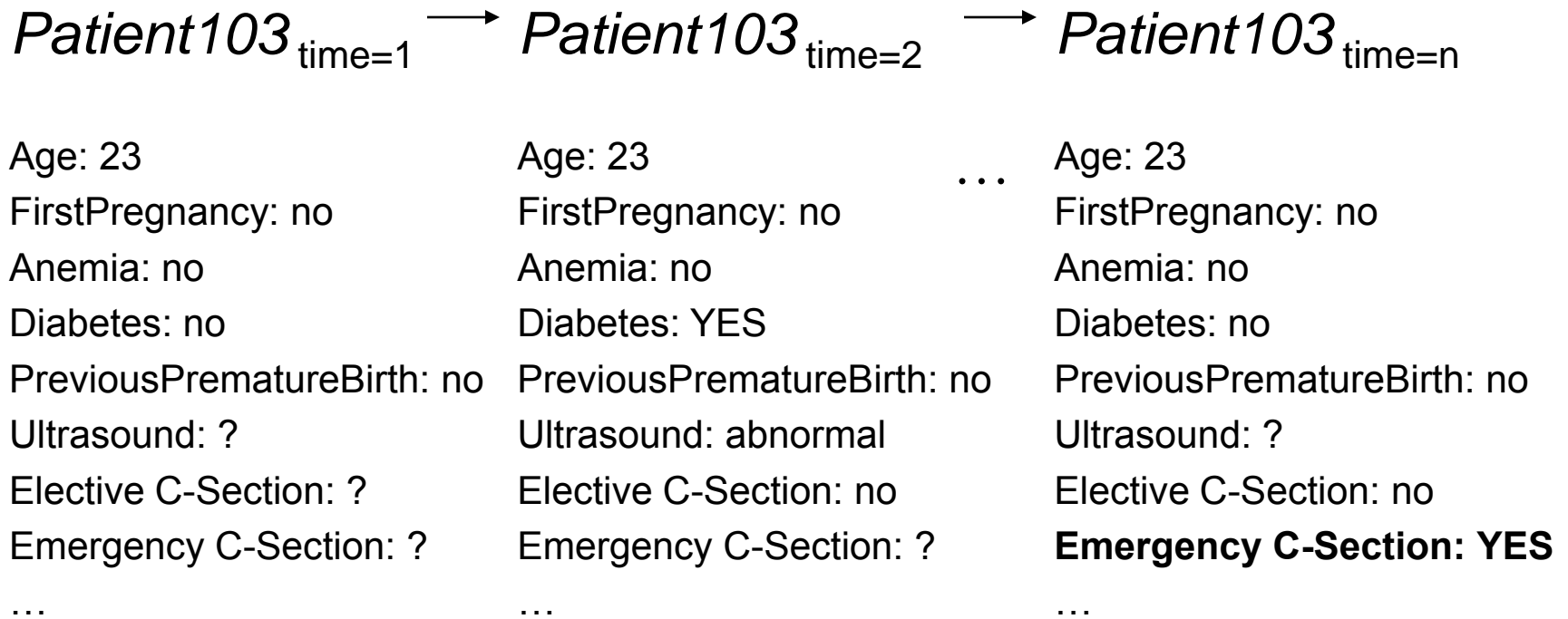
Given:

- 9714 patient records, each describing a pregnancy and birth
- Each patient record contains 215 features

Learn to predict:

- Classes of future patients at high risk for Emergency Cesarean Section

Patient Data



Datamining Result

One of 18 learned rules:

If

No previous vaginal delivery, and
Abnormal 2nd Trimester Ultrasound, and
Malpresentation at admission

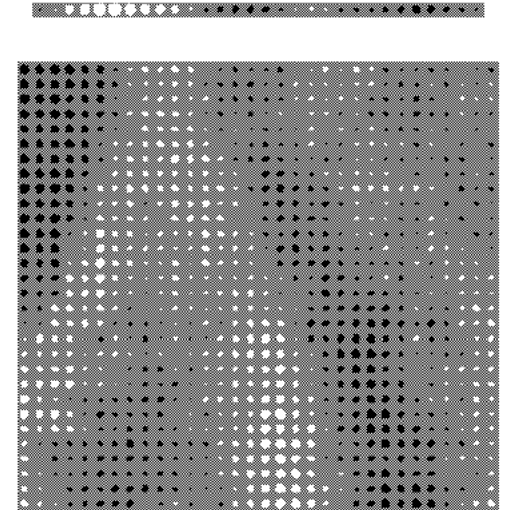
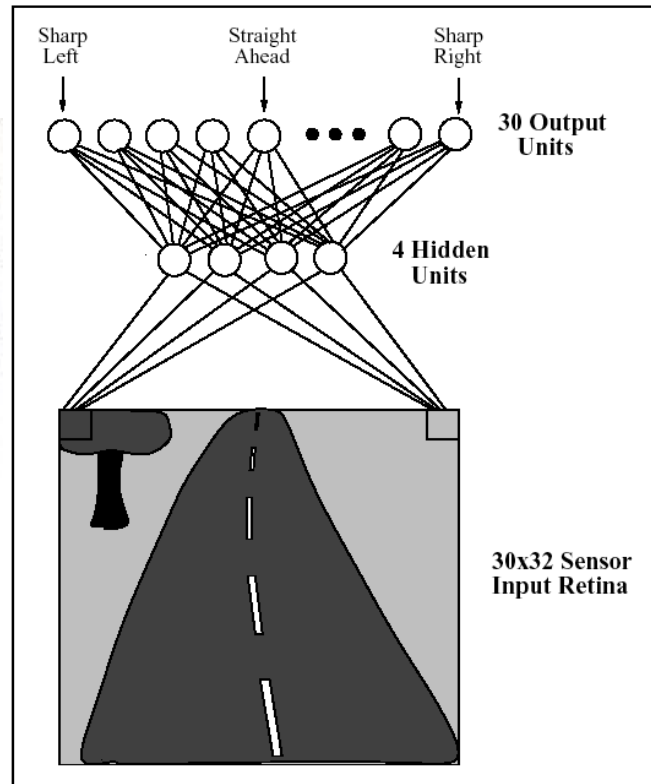
Then

Probability of Emergency C-Section is 0.6

- Accuracy in training data: $26/41 = .63$,
- Accuracy in test data: $12/20 = .60$

Too Difficult to Program

Problems too difficult to program by hand
ALVINN [Pomerleau] drives 70mph on highways



Software Customizes to User

<http://www.wisewire.com>

April 30, 1998

Lycos Acquires WiseWire

By internetnews.com Staff

Lycos, Inc. today announced the acquisition of targeted-content provider WiseWire Corp. for around \$39.75 million in Lycos shares.

Under the acquisition, the three-year old, Pittsburgh-based WiseWire's automated directory will be incorporated into Lycos' search page. Lycos said it will be the first online service using the technology which is based on user input and an intelligent agent product.



Where Is This Headed?

Today: tip of the iceberg

- First-generation algorithms: neural nets, decision trees, regression ...
- Applied to well-formatted database
- Budding industry

Looking to Tomorrow

Opportunity for tomorrow: enormous impact

- Learn across full mixed-media data
- Learn across multiple internal databases, plus the web and newsfeeds
- Learn by active experimentation
- Learn decisions rather than predictions
- Cumulative, lifelong learning
- Programming languages with learning embedded?

Relevant Disciplines

- Artificial intelligence
- Cognitive Science
- Probability theory and statistics
- Computational complexity theory
- Control theory
- Information theory
- Philosophy
- Psychology and neurobiology
- ...

Important Application Areas

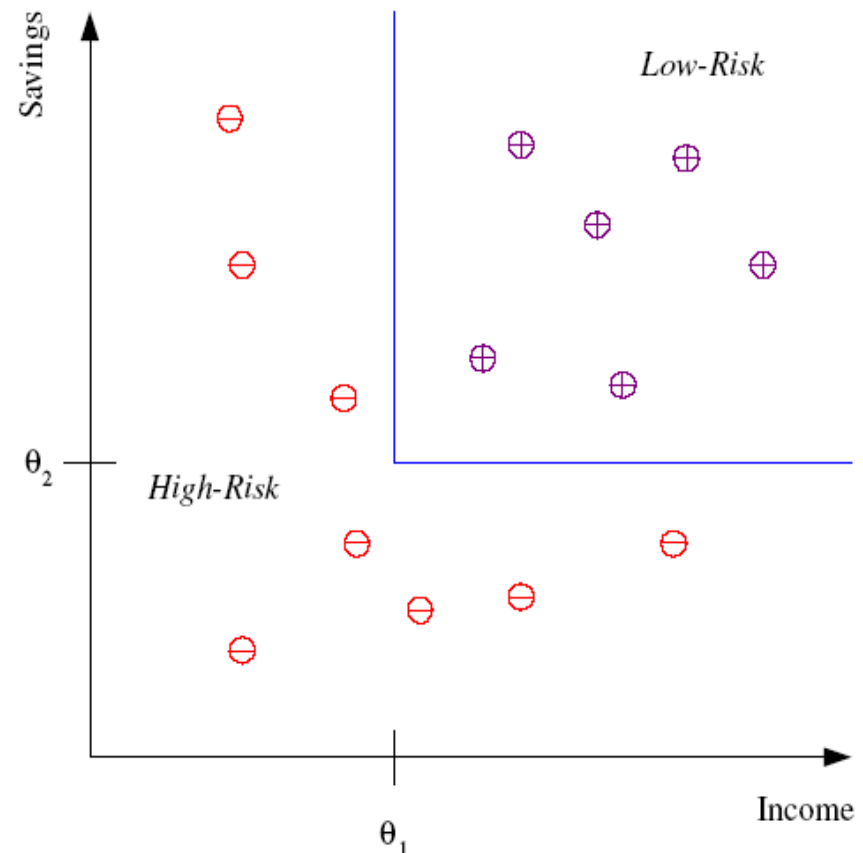
- **Bioinformatics:** sequence alignment, analyzing micro-array data, information integration, ...
- **Computer vision:** object recognition, tracking, segmentation, active vision, ...
- **Robotics:** state estimation, map building, decision making
- **Graphics:** building realistic simulations
- **Speech:** recognition, speaker identification
- **Financial analysis:** option pricing, portfolio allocation, ...
- **E-commerce:** automated trading agents, data mining, spam, ...
- **Medicine:** diagnosis, treatment, drug design, ...
- **Computer games:** building adaptive opponents
- **Multimedia:** retrieval across diverse databases
- **Web:** information extraction
- **Industry:** record linkage, social network analysis, anomaly detection, tracking entities, ...

Kinds of Learning

- Based on *information available*
 - Supervised – true labels provided
 - Reinforcement – Only indirect labels provided (reward/punishment)
 - Unsupervised – No feedback & no labels
- Based on the *role of the learner*
 - Passive – given a set of data, produce a model
 - Online – given one data point at a time, update model
 - Active – ask for specific data points to improve model
- Based on *type of output*
 - Concept Learning – Binary output based on +ve/-ve examples
 - Classification – Classifying into one among many classes
 - Regression – Numeric, ordered output

Type of Output: Classification

- Example: Credit scoring
- Differentiating between **low-risk** and **high-risk** customers from their *income* and *savings*



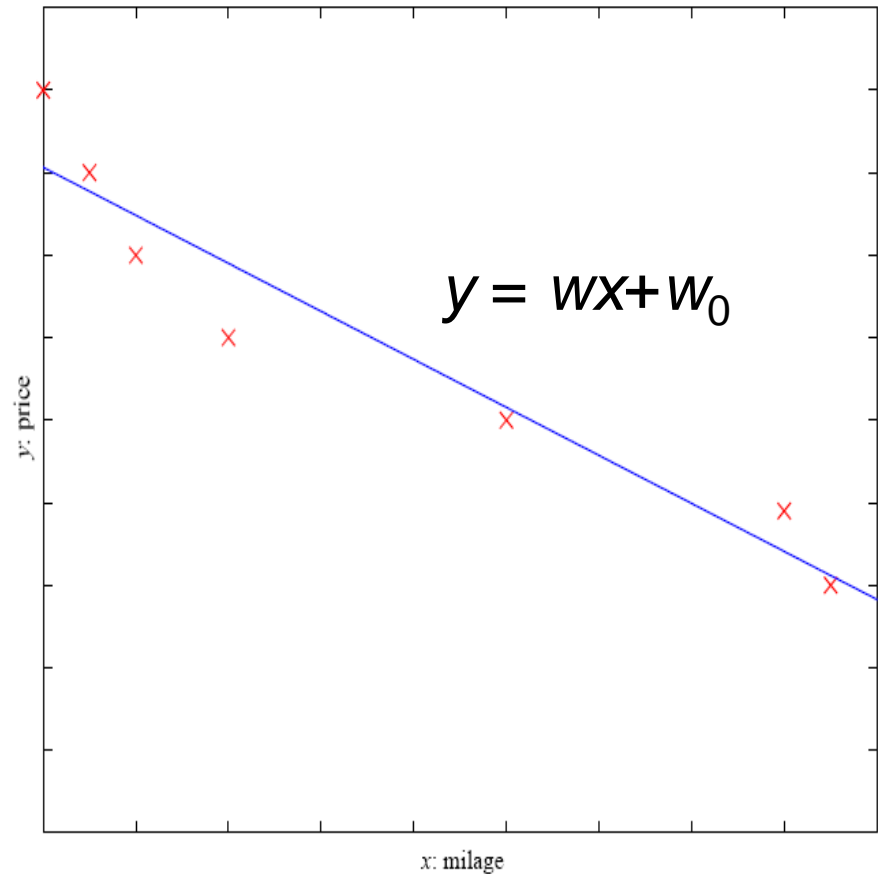
Discriminant: IF $income > \theta_1$ AND $savings > \theta_2$
THEN **low-risk** ELSE **high-risk**

Classification: Applications

- Aka Pattern recognition
- Face recognition: Pose, lighting, occlusion (glasses, beard), make-up, hair style
- Character recognition: Different handwriting styles.
- Speech recognition: Temporal dependency.
 - Use of a dictionary or the syntax of the language.
 - Sensor fusion: Combine multiple modalities; eg, visual (lip image) and acoustic for speech
- Medical diagnosis: From symptoms to illnesses
- ...

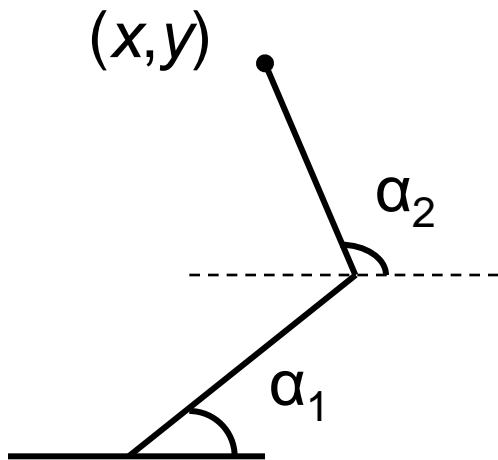
Type of Output: Regression

- Example: Price of a used car
- x : car attributes
- y : price
- $y = g(x | \theta)$
- $g(\)$ model,
- θ parameters



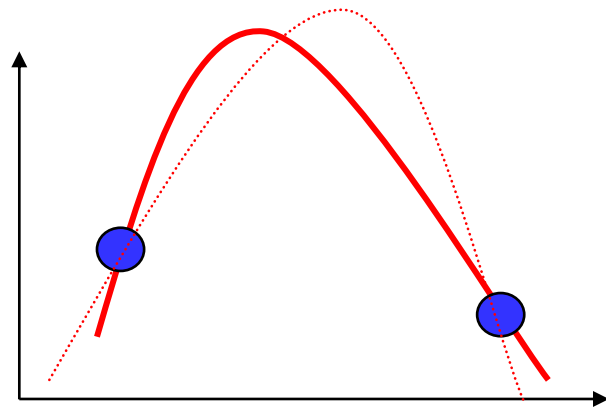
Regression Applications

- Navigating a car: Angle of the steering wheel (CMU NavLab)
- Kinematics of a robot arm



$$a_1 = g_1(x, y)$$

$$a_2 = g_2(x, y)$$



- Response surface design

Supervised Learning: Uses

- Prediction of future cases: Use the rule to predict the output for future inputs
- Knowledge extraction: The rule is easy to understand
- Compression: The rule is simpler than the data it explains
- Outlier detection: Exceptions that are not covered by the rule, e.g., fraud

Reinforcement Learning

- Learning a policy: A sequence of outputs
- No supervised output but delayed reward
- Credit assignment problem
- Game playing
- Robot in a maze
- Multiple agents, partial observability, ...

Unsupervised Learning

- Learning “what normally happens”
- No output
- Clustering: Grouping similar instances
- Example applications
 - Customer segmentation in CRM
 - Image compression: Color quantization
 - Bioinformatics: Learning motifs

Focus for this Course

- Based on *information available*
 - Supervised – true labels provided
 - Reinforcement – Only indirect labels provided (reward/punishment)
 - Unsupervised – No feedback & no labels
- Based on the *role of the learner*
 - Passive – given a set of data, produce a model
 - Online – given one data point at a time, update model
 - Active – ask for specific data points to improve model
- Based on *type of output*
 - Concept Learning – Binary output based on +ve/-ve examples
 - Classification – Classifying into one among many classes
 - Regression – Numeric, ordered output