Outline

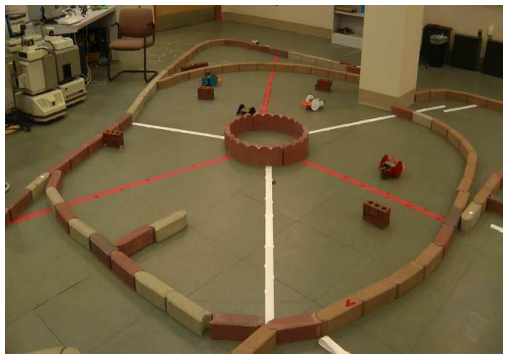
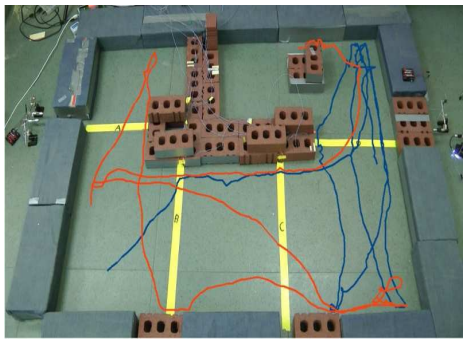
- Introduce yourself!!
- What is Machine Learning?
- What is CAP-5610 about?
- Class information and logistics

About the instructor:

- **Name:** Leonardo Bobadilla, Ph.D

B.E Computer Systems and Engineering. (National University of Colombia)

- **M.Sc Statistics** (National University of Colombia)
- **Ph.D Computer Science** (University of Illinois at Urbana-Champaign)
- **Research interests**: Robotics, Artificial Intelligence, Cyber-Physical Systems

About the instructor:

Leonardo Bobadilla, Assistant Professor
Meeting times: Tuesday/Thursday 9:30am-10:45am

bobadilla@cs.fiu.edu

Office:ECS 212b

Phone: 217-778-4009

Tuesday/Thursday 11:00am-12:00pm or by appointment

Introduce yourself!

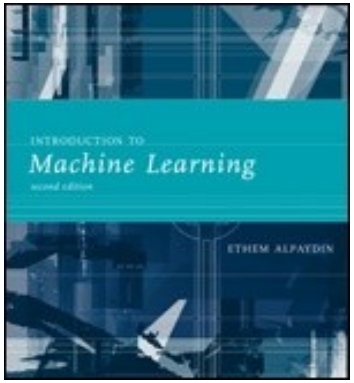
- -Name
- -Year of PhD
- -Area
- -Advisor
- -Why are you interested in the class?
- -Classes you have taken in Statistics and Linear Algebra

Lecture Slides for

INTRODUCTION TO

Machine Learning

2nd Edition



ETHEM ALPAYDIN, modified by Leonardo Bobadilla

alpaydin@boun.edu.tr
http://www.cmpe.boun.edu.tr/~ethem/i2m

Why “Learn” ?

- Machine learning is programming computers to optimize a performance criterion using example data or past experience.
- There is no need to “learn” to calculate payroll
- Learning is used when:
  - Human expertise does not exist (navigating on Mars),
  - Humans are unable to explain their expertise (speech recognition)
  - Solution changes in time (routing on a computer network)
  - Solution needs to be adapted to particular cases (user biometrics)

What We Talk About When We Talk About “Learning”

- Learning general models from a data of particular examples
- Data is cheap and abundant (data warehouses, data marts); knowledge is expensive and scarce.
- Example in retail: Customer transactions to consumer behavior:

  People who bought “Blink” also bought “Outliers” (www.amazon.com)
- Build a model that is *a good and useful approximation* to the data.

Domains

Machine learning is everywhere!! One of the most useful areas of CS!

- Retail: Market basket analysis, Customer relationship management (CRM)
- Finance: Credit scoring, fraud detection
- Manufacturing: Control, robotics, troubleshooting
- Medicine: Medical diagnosis
- Telecommunications: Spam filters, intrusion detection
- Bioinformatics: Motifs, alignment
- Web mining: Search engines

What is Machine Learning?

- Optimize a performance criterion using example data or past experience.
- Role of Statistics: Inference from a sample
- Role of Computer science: Efficient algorithms to
  - Solve the optimization problem
  - Representing and evaluating the model for inference

Applications

- Association
- Supervised Learning
  - Classification
  - Regression
- Unsupervised Learning
- Reinforcement Learning

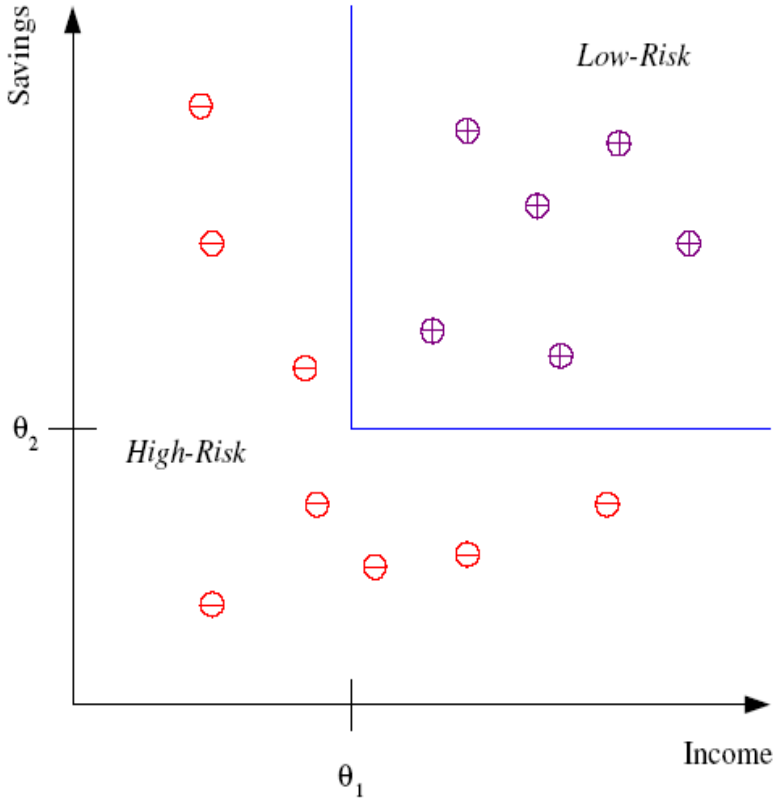
Learning Associations

- Basket analysis:
  $P(Y \mid X)$ probability that somebody who buys X also buys Y where X and Y are products/services.

  Example: $P(\text{chips} \mid \text{beer}) = 0.7$

Classification

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



Discriminant: IF *income* > θ_1 AND *savings* > θ_2
THEN low-risk ELSE high-risk

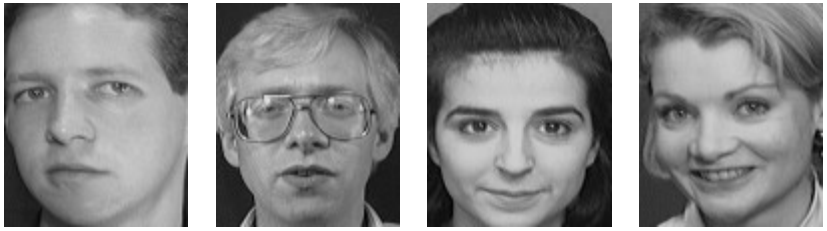
Classification: Applications

- Aka Pattern recognition
- Face recognition: Pose, lighting, occlusion (glasses, beard), make-up, hair style
- Character recognition: Different handwriting styles.
- Speech recognition: Temporal dependency.
- Medical diagnosis: From symptoms to illnesses
- Biometrics: Recognition/authentication using physical and/or behavioral characteristics: Face, iris, signature, etc
- ...

Face Recognition

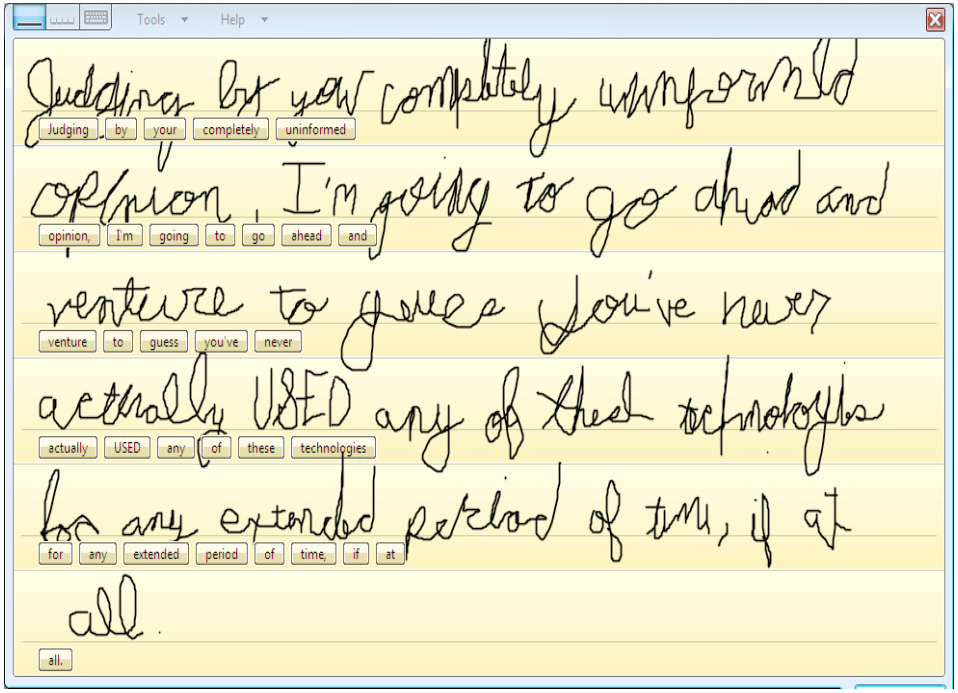
Training examples of a person

Test images



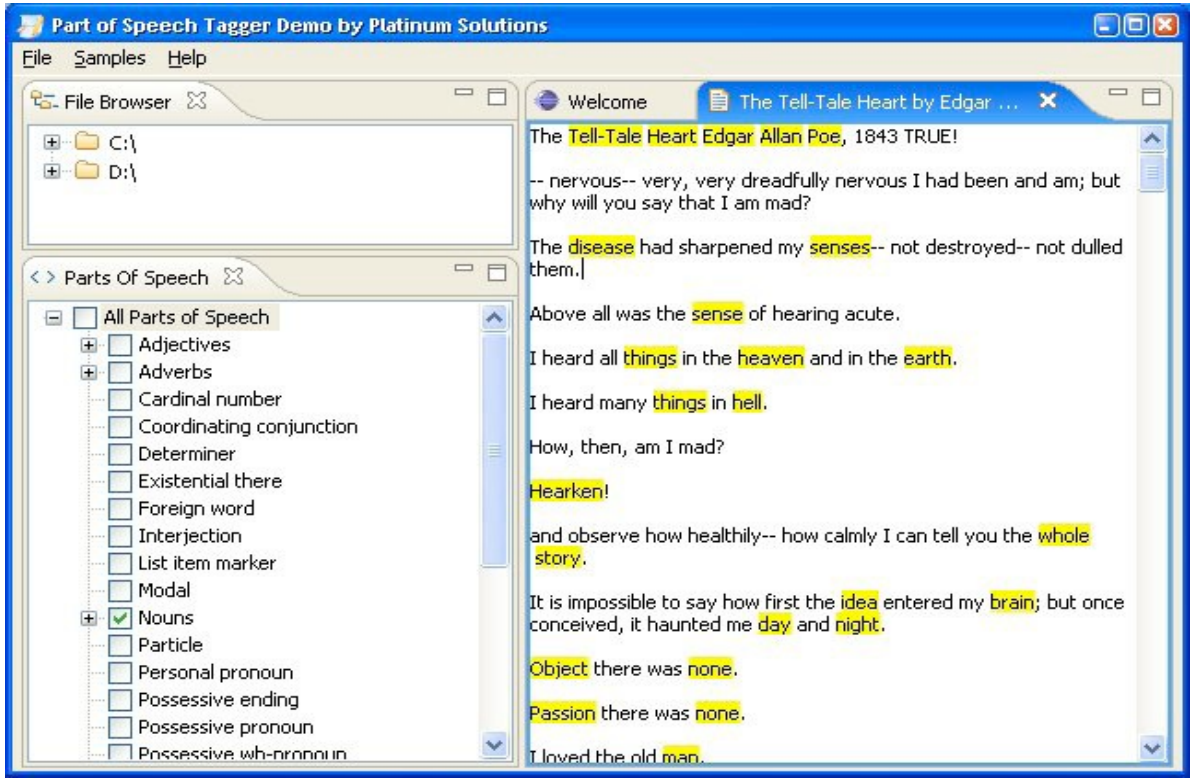
ORL dataset,
AT&T Laboratories, Cambridge UK

Applications of Classification



Taken rom http://www.cs.uccs.edu/~jkalita/work/cs586/2010/
http://www.platerecognition.info/
http://www.gottabemobile.com/forum/uploads/322/recognition.png

Applications of Classification: POS Tagging

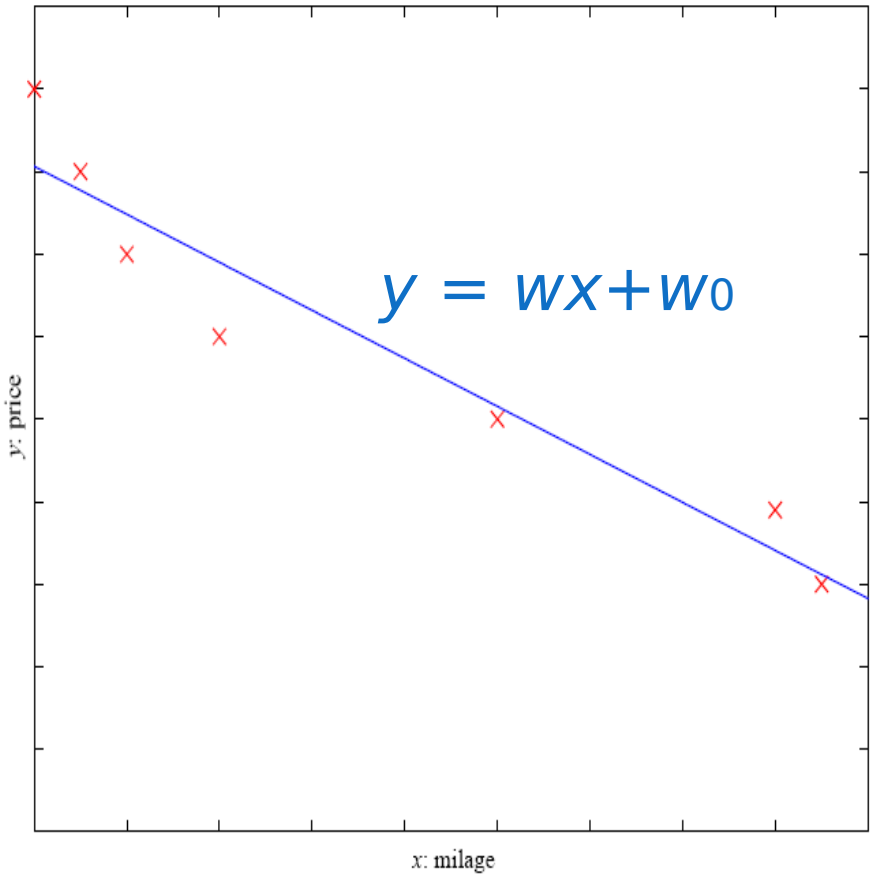


Taken from http://www.cs.uccs.edu/~jkalita/work/cs586/2010/

http://blog.platinumsolutions.com/files/pos-tagger-screenshot.jpg.

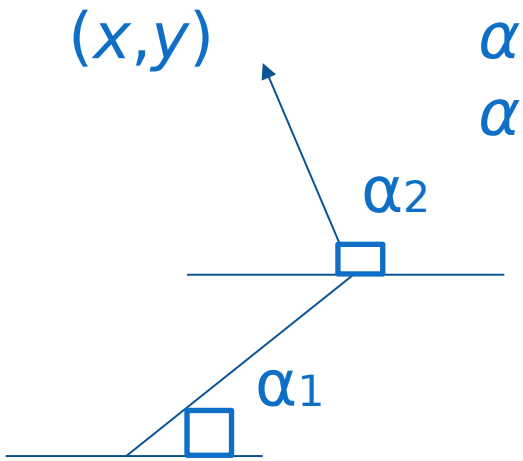
Regression

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

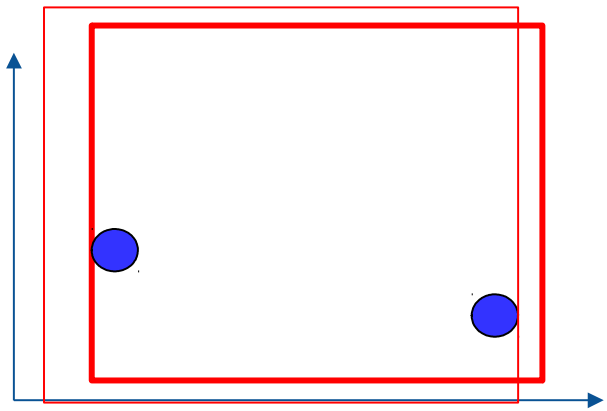


Regression Applications

- Navigating a car: Angle of the steering
- Kinematics of a robot arm



$\alpha_1 = g_1(x,y)$
$\alpha_2 = g_2(x,y)$



- 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

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

Unsupervised Learning

Bioinformatics: Clustering genes according to gene array expression data.

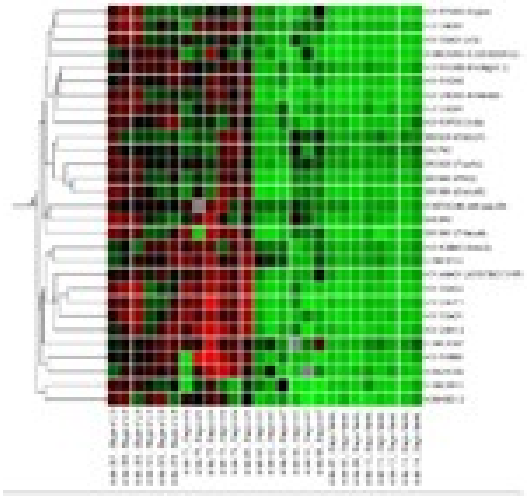
• Finance: Clustering stocks or mutual based on characteristics of company or companies involved

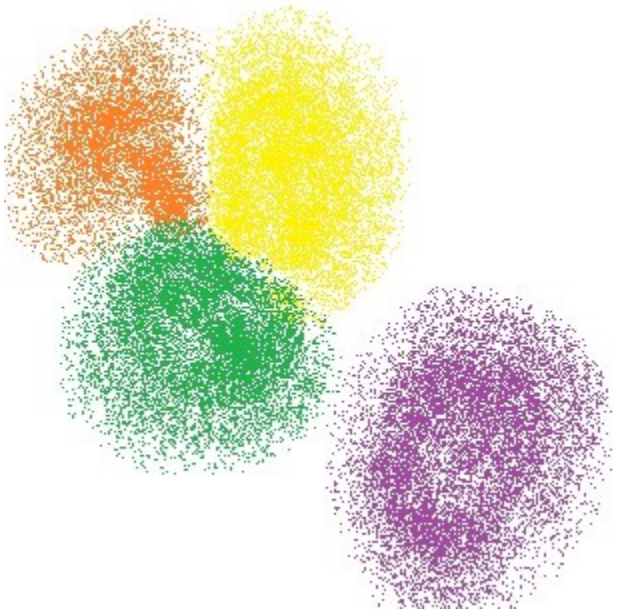
Document clustering: Cluster documents based on the words that are contained in them.

• Customer segmentation: Cluster customers based on demographic information, buying habits, credit information, etc. Companies advertise differently to different customer segments. Outliers may form niche markets.

Taken from http://www.cs.uccs.edu/~jkalita/work/cs586/2010/

Unsupervised Learning Examples





Reinforcement Learning

In some applications, the output of the system is a sequence of actions.

- A single action is not important alone.
- What is important is the policy or the sequence of correct actions to reach the goal.
- In reinforcement learning, reward or punishment comes usually at the very end or infrequent intervals.
- The machine learning program should be able to assess the goodness of ”policies”; learn from past good action sequences to generate a ”policy”.

Applications of Reinforcement Learning

Game Playing: Games usually have simple rules and environments although the game space is usually very large. A single move is not of paramount importance; a sequence of good moves is needed. We need to learn good game playing policy.

Robot navigating in an environment: A robot is looking for a goal location to charge, or to pick up trash, to pour a liquid, to hold a container or object. At any time, the robot can move in many in one of a number of directions, or perform one of several actions. After a number of trial runs, it should learn the correct sequence of actions to reach the goal state from an initial state, and do it efficiently.

Taken from http://www.cs.uccs.edu/~jkalita/work/cs586/2010/

Resources: Datasets

UCI Repository:
http://www.ics.uci.edu/~mlearn/MLRepository.html

- UCI KDD Archive:
http://kdd.ics.uci.edu/summary.data.application.html

Course Logistics: Prerequisites

- Statistics and Linear Algebra useful, but I can provide refreshers if needed
- A programming language
- **Most important:** Desire and motivation to learn and apply Machine Learning.

Course Logistics: Final Project

- A research project, up to 2 students
- An opportunity to apply what you have learned
- Pick a problem that you find **interesting**
- Ideally should be a publishable effort
- 1) Project proposal. 2) Midterm progress. 3) Project Presentation. 4) Final Report.
- I will provide feedback all the way

Course Logistics: Homeworks

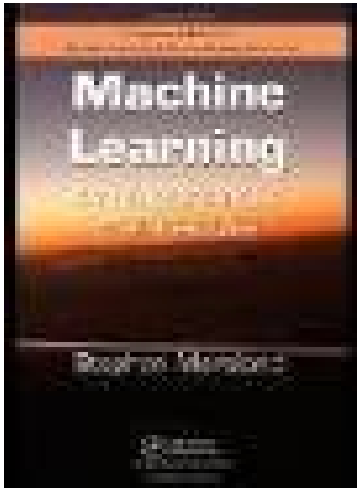
- Cheating No!
- Homework: Collaboration and study groups are encouraged;
- However, write your own solutions and program and do not use old solutions
- Midterm and Final will be based on homeworks
-

Course Logistics: Class participation

- I will put the corresponding sections of the book on the calendar. Please read them before class.
- Ask questions, participate, discuss in class!

Course Logistics: Programming Languages

- **Python:** Interesting book Machine Learning: An Algorithmic Perspective
- R
- Octave
- Matlab
- Java



Next class: Supervised Learning!

- Sections 2.1-2.4 I2ML