CS446 Introduction to Machine Learning (Spring 2015)
University of Illinois at Urbana-Champaign
http://courses.engr.illinois.edu/cs446

Lecture 1: Introduction

Prof. Julia Hockenmaier
juliahmr@illinois.edu

Welcome to CS 446!

Professor:

Julia Hockenmaier juliahmr@illinois.edu

Teaching assistants:

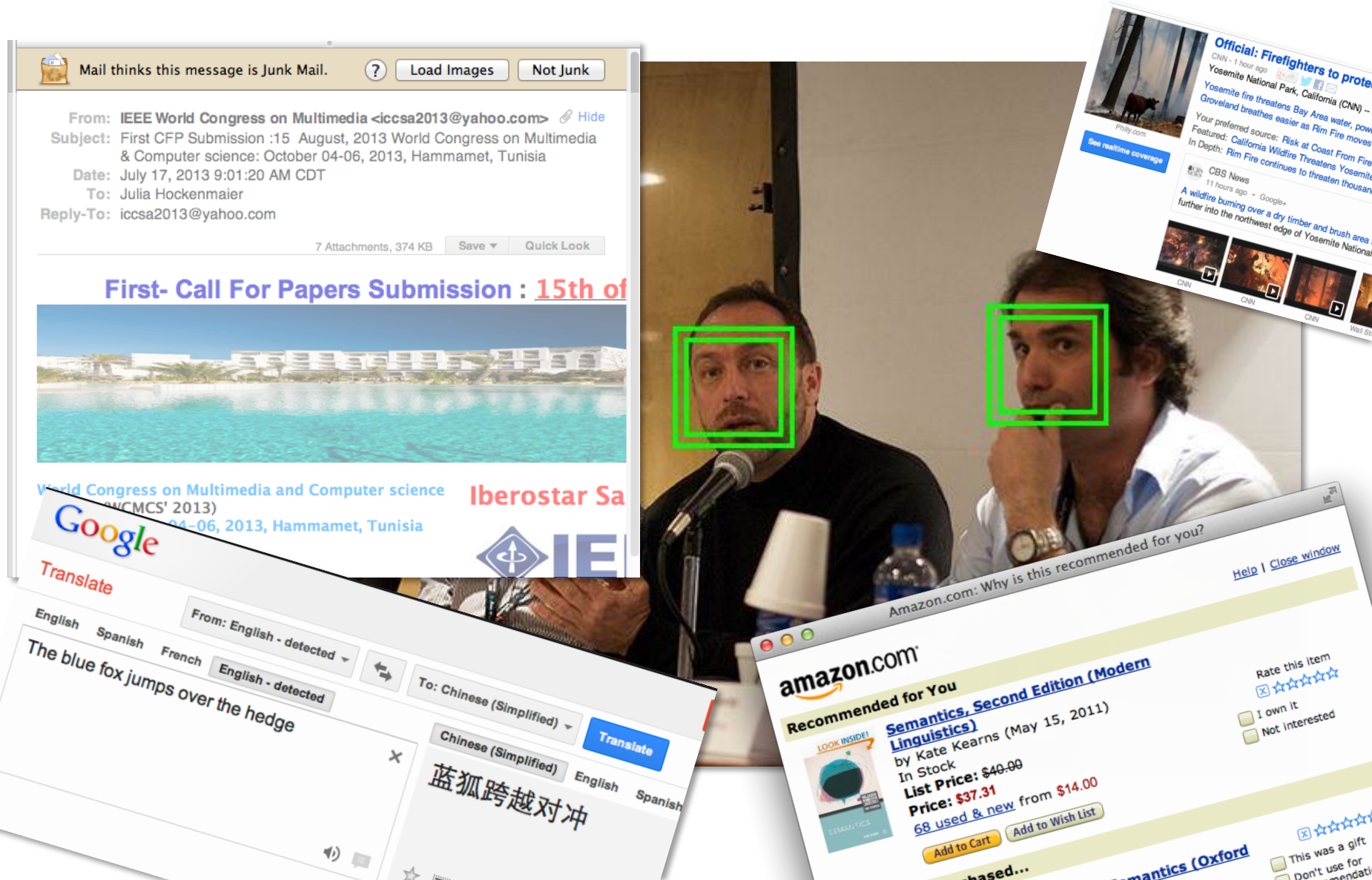
Ryan Musa ramusa2@illinois.edu

Stephen Mayhew mayhew2@illinois.edu

Ruichen (Ray) Wang rwang11@illinois.edu

What is machine learning?

Machine learning is everywhere



Applications: Spam Detection

This is a binary classification task:
Assign one of two labels (i.e. yes/no)
to the input (here, an email message)

Applications: Spam Detection

Classification requires a model (a classifier) to determine which label to assign to items.

Applications: Spam Detection

In this class, we study algorithms and techniques to learn such models from data.

Learning = Generalization

Mail thinks this message is junk mail.

Not junk

The learner has to be able to classify items it has never seen before.

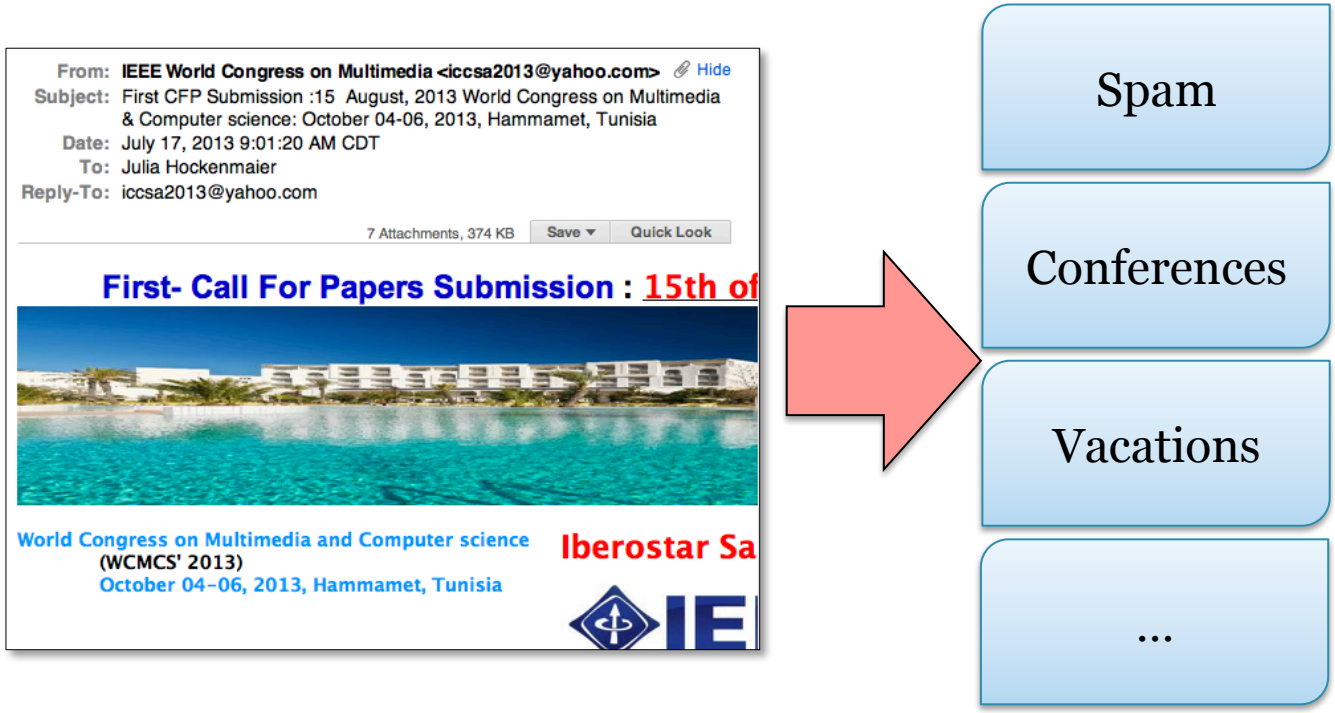
Learning = Adaptation

Mail thinks this message is junk mail.

Not junk

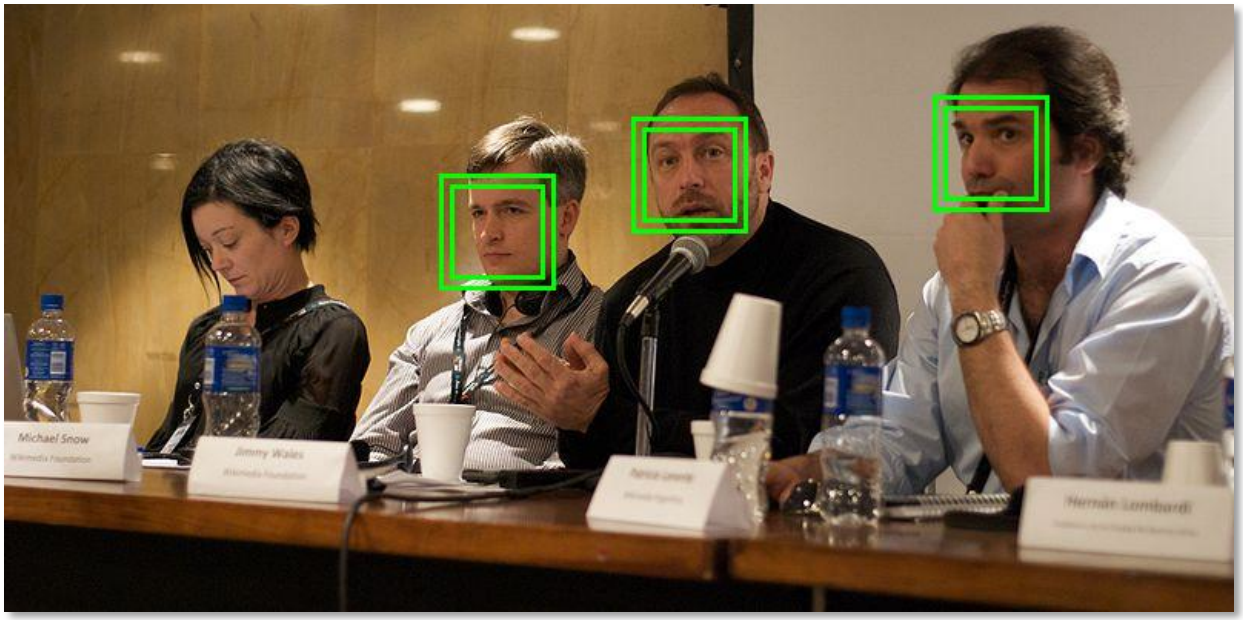
The learner should adapt its model to feedback (supervision) it receives.

Applications: Text classification



This is a multiclass classification task:
Assign one of k labels to the input
{Spam, Conferences, Vacations,…}

Applications: Face recognition



This is also a binary classification task: Decide for each rectangular image region whether it shows a face or not.

What will we cover in this class?

CS446: Key questions

- What kind of tasks can we learn models for?
- What kind of models can we learn?
- What algorithms can we use to learn?
- How do we evaluate how well we have learned to perform a particular task?
- How much data do we need to learn models for a particular task?

Learning scenarios

Supervised learning:

The focus of CS446

Learning to predict labels from correctly labeled data

Unsupervised learning:

Learning to find hidden structure (e.g. clusters) in input data

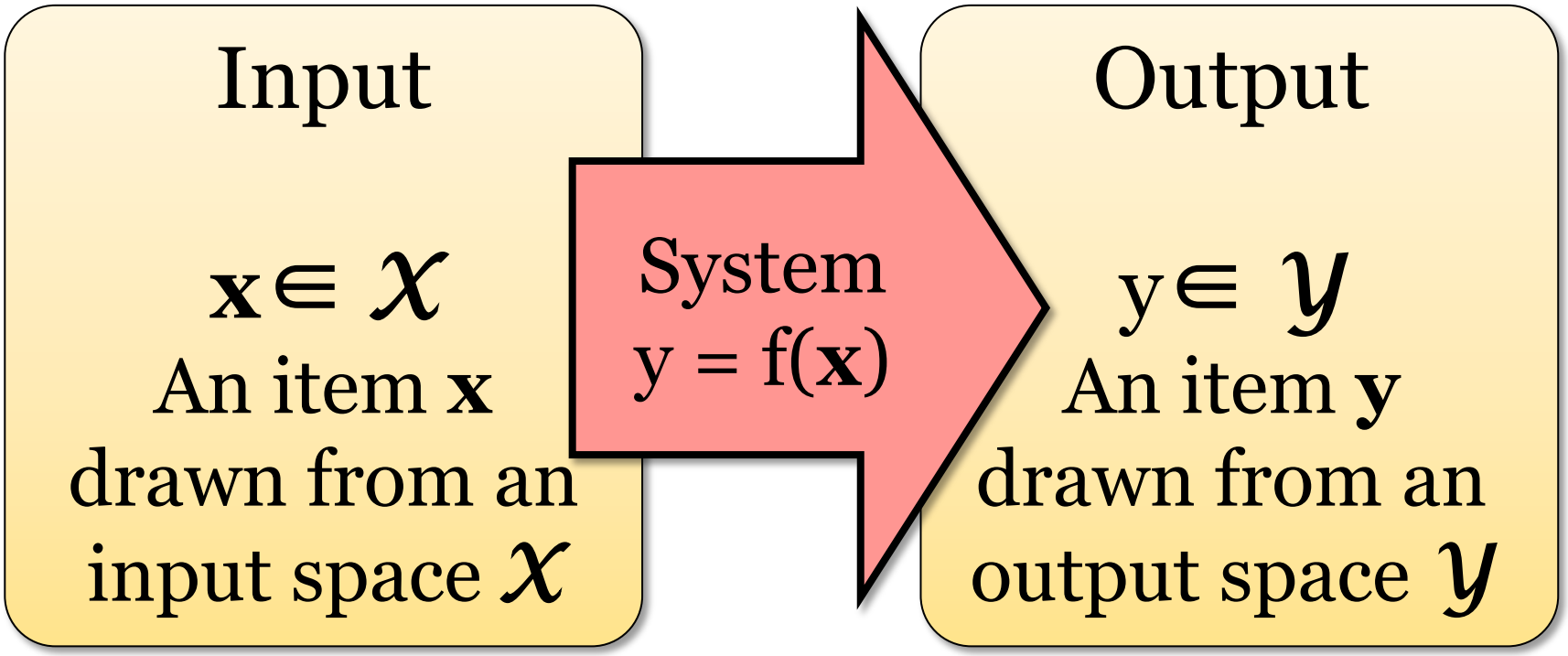
Semi-supervised learning:

Learning to predict labels from (a little) labeled and (a lot of) unlabeled data

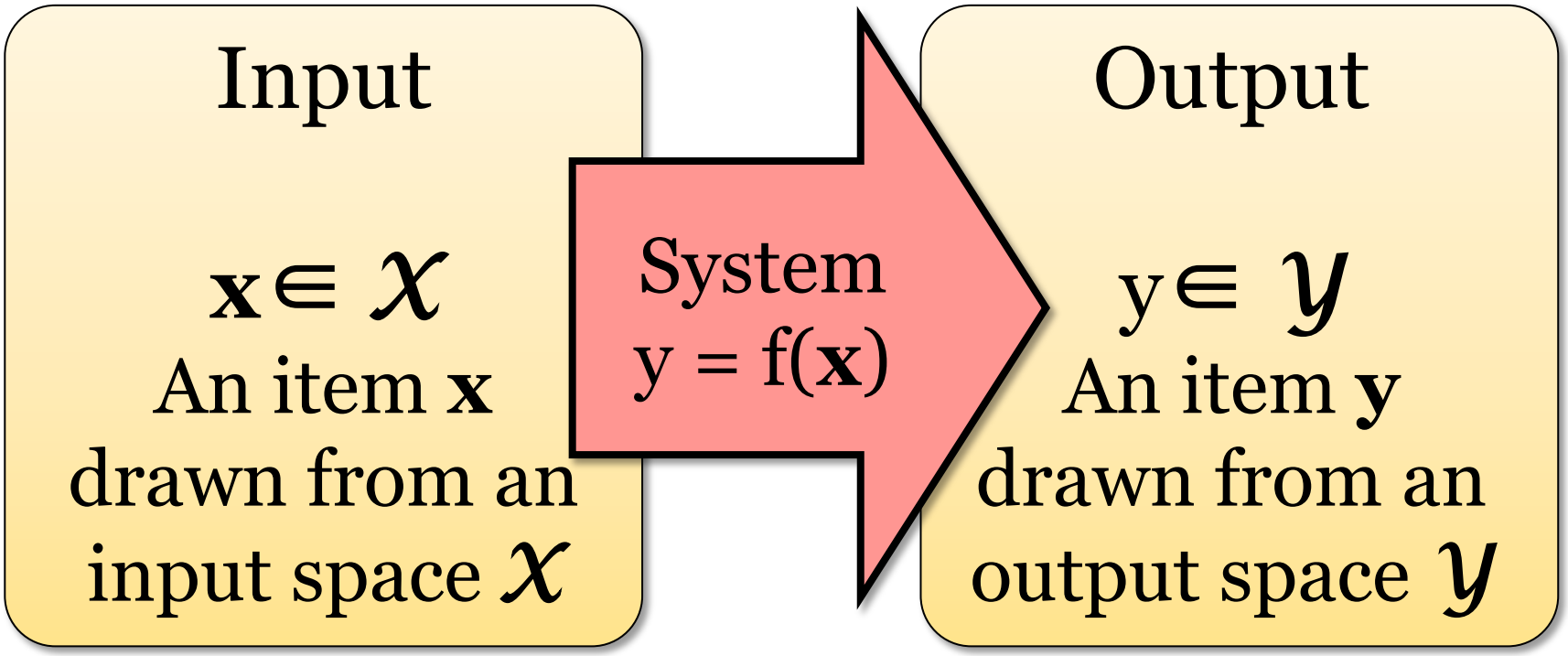
Reinforcement learning:

Learning to act through feedback for actions (rewards/punishments) from the environment

Supervised learning



We consider systems that apply a function f() to input items **x** and return an output **y** = f(**x**).

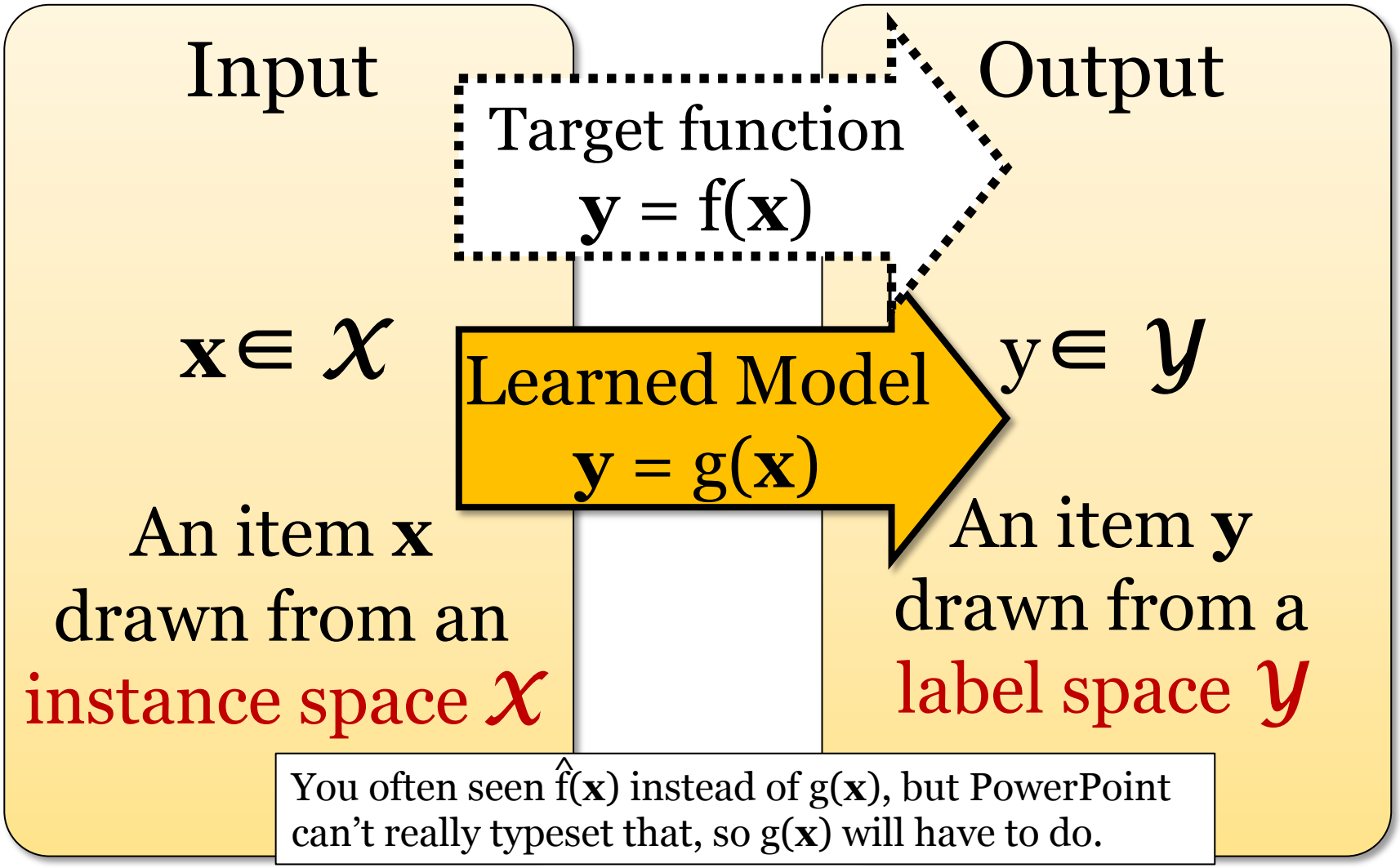


In (supervised) machine learning, we deal with systems whose $f(\mathbf{x})$ is learned from examples.

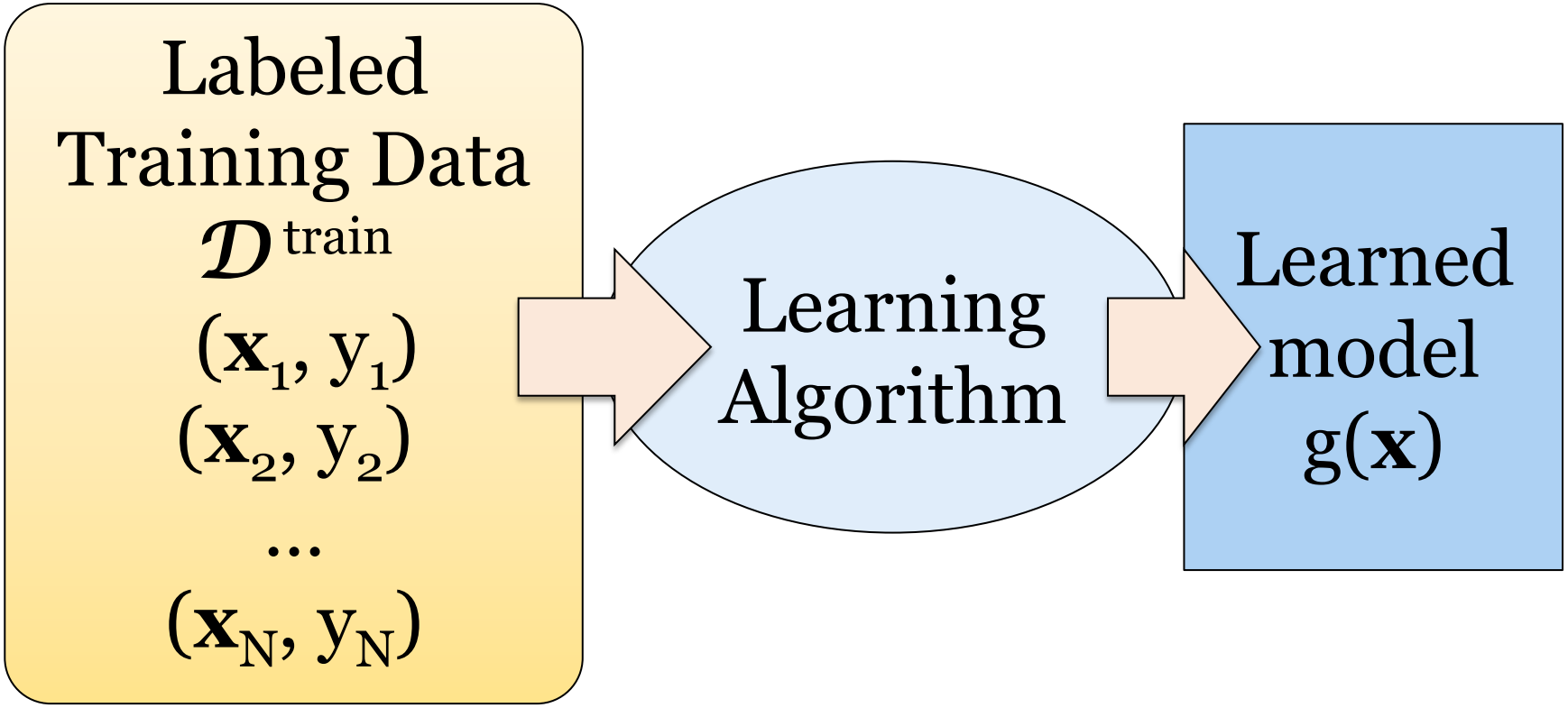
Why use learning?

We typically use machine learning when the function f(**x**) we want the system to apply is too complex to program by hand.

Supervised learning



Supervised learning: Training



Give the learner examples in $\mathcal{D}^{\text{train}}$

The learner returns a model $g(\mathbf{x})$

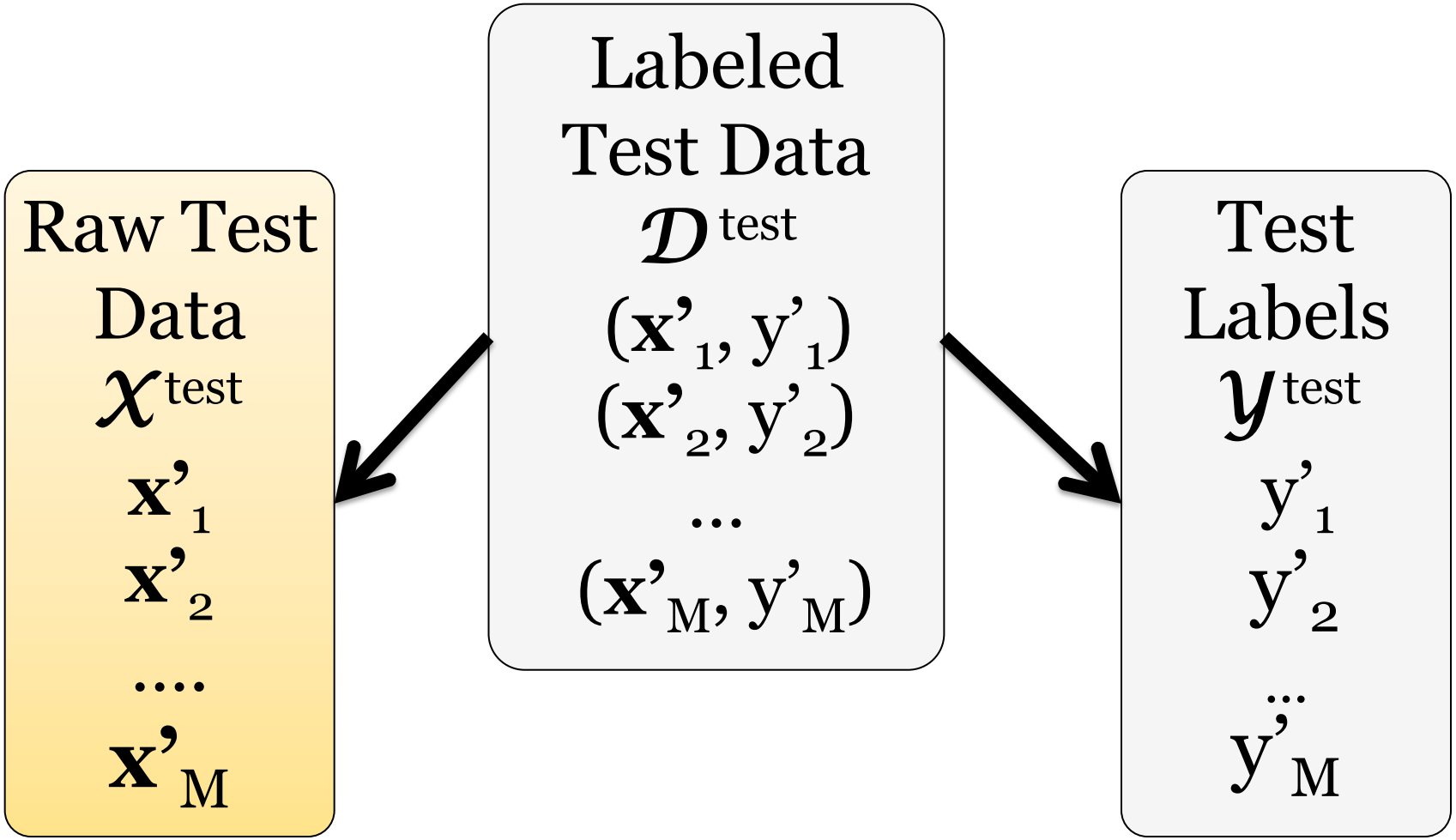
Supervised learning: Testing

Labeled
Test Data
$\mathcal{D}^{\text{test}}$

$(\mathbf{x'}_1, y'_1)$
$(\mathbf{x'}_2, y'_2)$
…
$(\mathbf{x'}_M, y'_M)$

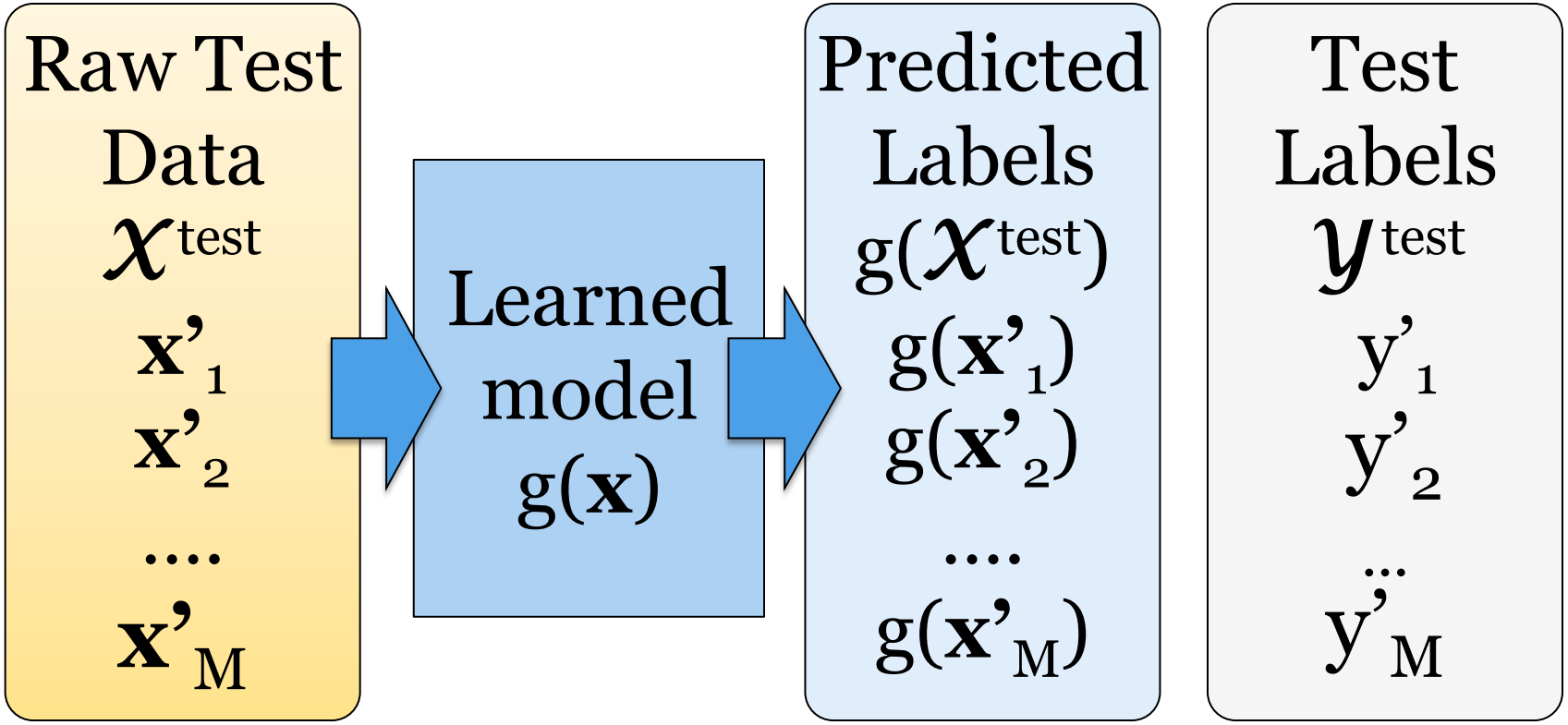
Reserve some labeled data for testing

Supervised learning: Testing



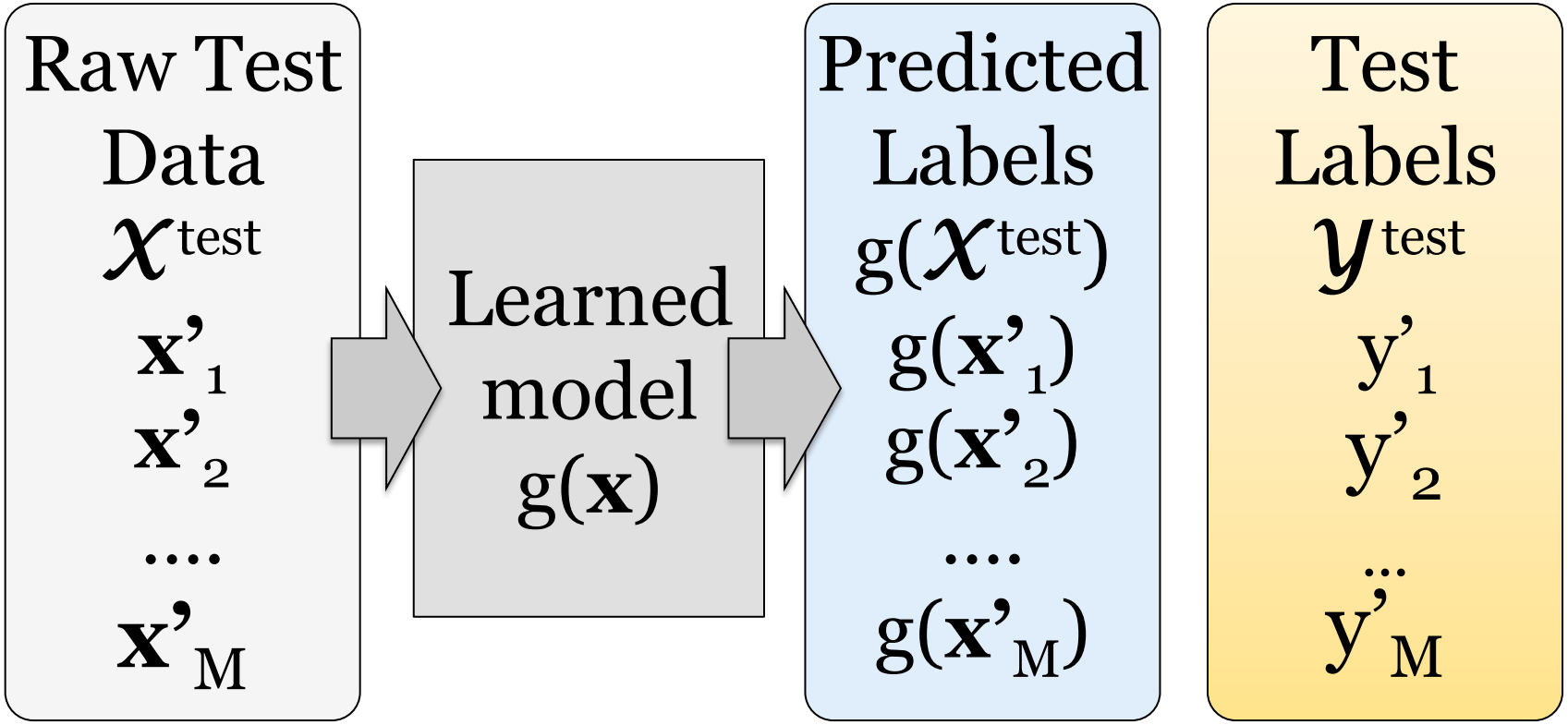
Supervised learning: Testing

Apply the model to the raw test data



Supervised learning: Testing

Evaluate the model by comparing the predicted labels against the test labels



The Badges game

The Badges game

+ Naoki Abe	- Eric Baum

Conference attendees to the 1994 Machine Learning conference were given name badges labeled with + or –.

What function was used to assign these labels?

Training data

+ Naoki Abe
- Myriam Abramson
+ David W. Aha
+ Kamal M. Ali
- Eric Allender
+ Dana Angluin
- Chidanand Apte
+ Minoru Asada
+ Lars Asker
+ Javed Aslam
+ Jose L. Balcazar
- Cristina Baroglio
+ Peter Bartlett
- Eric Baum
+ Welton Becket
- Shai Ben-David
+ George Berg
+ Neil Berkman
+ Malini Bhandaru
+ Bir Bhanu
+ Reinhard Blasig
- Avrim Blum
- Anselm Blumer
+ Justin Boyan
+ Carla E. Brodley
+ Nader Bshouty
- Wray Buntine
- Andrey Burago
+ Tom Bylander
+ Bill Byrne
- Claire Cardie
+ John Case
+ Jason Catlett
- Philip Chan
- Zhixiang Chen
- Chris Darken

Raw test data

Gerald F. DeJong
Chris Drummond
Yolanda Gil
Attilio Giordana
Jiarong Hong
J. R. Quinlan
Priscilla Rasmussen
Dan Roth
Yoram Singer
Lyle H. Ungar

Labeled test data

+ Gerald F. DeJong
- Chris Drummond
+ Yolanda Gil
- Attilio Giordana
+ Jiarong Hong
- J. R. Quinlan
- Priscilla Rasmussen
+ Dan Roth
+ Yoram Singer
- Lyle H. Ungar

How will we teach this class?

Lectures

Tuesdays and Thursdays
3:30 PM – 4:45 PM
Digital Computer Lab (Room 1320)

Slides will be on the website before class.

Lecture videos will be uploaded after class.

- I have no control over the quality of the recordings.
  In particular, I don't know when the sound is not being recorded while I'm teaching.
  Please let us know ASAP when there's a problem.

Contacting the CS446 staff

Professor:

Julia Hockenmaier juliahmr@illinois.edu

Teaching assistants:

Ryan Musa ramusa2@illinois.edu

Stephen Mayhew mayhew2@illinois.edu

Ruichen (Ray) Wang rwang11@illinois.edu

Anonymous feedback:

Via class website https://courses.engr.illinois.edu/cs446/

Office Hours (starting next week)

Julia Hockenmaier (3324 Siebel)

Thu, 5:00 PM – 6:00 PM

TA office hours (for on-campus and for on-line students) **will be announced next week.**

CS446 on the web

Check our **class website**:

Schedule, slides, videos, policies, anonymous feedback

http://courses.engr.illinois.edu/cs446/

Sign up, participate in our **Piazza forum**:

Announcements and discussions

https://piazza.com/illinois/spring2015/cs446/

Log on to **Compass**:

Submit assignments, get your grades

https://compass.illinois.edu

Assessment and Grading

If you take this class for 3 hours credit, your grade will be determined by your performance on

- Homework (33.3% of your grade)
- Midterm exam (33.3% of your grade)
- Final exam (33.3% of your grade)

Assessment and Grading

If you take this class for 4 hours credit, your grade will be determined by your performance on

- Homework (25% of your grade)
- Midterm exam (25% of your grade)
- Final exam (25% of your grade)
- Research project (25% of your grade)

Homework

There will be 6 assignments.

- We plan to release them on Thursdays in Weeks 2, 4, 6, 8, 10, and 12.
- Some, but not all require programming

  Probably some Matlab, some Java, some with a language of your choice
- You will have two weeks to complete them.

Homework: Submission

You need to use Compass to submit your solutions (http://compass2g.illinois.edu)

We do not accept any handwritten solutions.

- Reports have to be submitted as PDFs, typeset in LaTeX (templates provided)

Homework: Late Policy

Everybody is allowed a total of two late days for the semester.

If you have exhausted your contingent of late days, we will subtract 20% per late day.

We don't accept assignments more than three days after their due date.

Let us know if there are any special circumstances (family, health, etc.)

Homework: Collaboration

We encourage collaboration and discussion, but you need to submit your own work.

If you are asked to write your own code, do so.

Piazza: Use it to discuss problems and give (reasonable) hints. *But if you post complete solutions, you may fail the assignment.*

Homework: Plagiarism

We don't tolerate plagiarism.

- Cite all external sources (including external code) you have used
- We may compare your source code if we suspect plagiarism.
- Don't reuse old solutions from previous years.

Exams

Midterm exam: Thursday, March 5 *in class*

Final exam: Tuesday, May 5 *in class*

Let us know ASAP if you have a conflict
on those days.
Also let us know ASAP if you need special DRES accommodations.

Closed-book exams:

No books/cheat sheets/calculators/computers/phones

4th Credit Hour Projects

Perform an experimental research project that uses machine learning

We encourage you to work in pairs
(We don't allow larger groups)

Write a paper that describes your task, relevant background, and experiments

4th Credit Hour Projects

Milestone 1 (Week 5)

Have a partner, agreed on a task, submit proposal

Milestone 2 (Week 9)

Submit preliminary results and task description (including relevant background)

Milestone 3 (Week 13)

Submit more fleshed-out results and report

Milestone 4 (After the final exam)

Submit complete report, do brief oral presentation

Questions?

juliahmr@illinois.edu

http://courses.engr.illinois.edu/cs446/

http://piazza.com/illinois/spring2015/cs446