CS 66: Machine Learning

Prof. Sara Mathieson

Spring 2019

SWARTHMORE

Admin

- **EVERYONE**: sign in (separate sheets for registered and waitlist)
- **Waitlist**: come to either lab (1:15pm or 3pm)
- **Registered**: must come to your assigned lab section unless you have found someone to switch with you (email me ASAP if that's the case)
- We are looking for 1 person to switch from B to A

Outline for January 23

- Welcome + what is Machine Learning (ML)?
- Examples of ML
- Syllabus highlights
- ML terminology
- Notation for this class
- First algorithm: K-nearest neighbors

Outline for January 23

- Welcome + what is Machine Learning (ML)?
- Examples of ML
- Syllabus highlights
- ML terminology
- Notation for this class
- First algorithm: K-nearest neighbors

Discuss with a Partner

- Introduce yourselves
- Come up with your own definition of "Machine Learning"
- How is Machine Learning different from or similar to:
  - Statistics
  - Data mining
  - Psychology of learning

What is Machine Learning?

- “Machine Learning is the study of methods for programming computers to learn.”

-Tom Dietterich

What is Machine Learning?

- “Machine Learning is the study of methods for programming computers to learn.”

  -Tom Dietterich

- “Machine Learning seeks to answer the question: `How can we build computer systems that automatically improve with experience, and what are the fundamental laws that govern all learning processes?’ ”

  -Tom Mitchell

Why would we want ML?

Why would we want ML?

1) No human experts

Example: predicting failure points for new machines

Why would we want ML?

1) No human experts

Example: predicting failure points for new machines

2) Human experts cannot explain expertise

Example: cannot explain exactly what a handwritten “2” looks like

Why would we want ML?

1) No human experts

   Example: predicting failure points for new machines

2) Human experts cannot explain expertise

   Example: cannot explain exactly what a handwritten “2” looks like

3) Phenomena change rapidly

   Example: predicting the stock market

Why would we want ML?

1) No human experts

Example: predicting failure points for new machines

2) Human experts cannot explain expertise

Example: cannot explain exactly what a handwritten “2” looks like

3) Phenomena change rapidly

Example: predicting the stock market

4) Customization for each user

Example: program that adapts to each user’s speech

Adapted from “Machine Learning” by Tom Dietterich

ML and related fields

- **Statistics:** understanding phenomenon that generated the data

ML and related fields

- **Statistics:** understanding phenomenon that generated the data

- **Data Mining:** find patterns in data that are understandable to humans

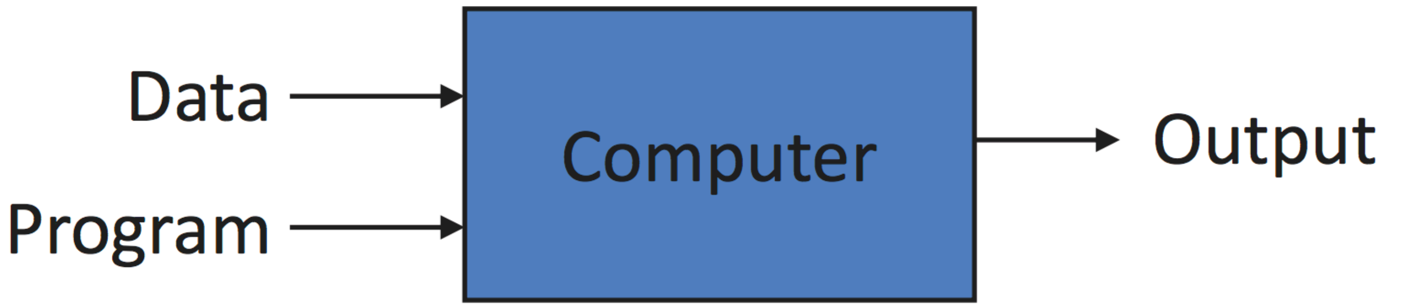
ML and related fields

- **Statistics:** understanding phenomenon that generated the data
- **Data Mining:** find patterns in data that are understandable to humans
- **Psychology of learning:** understand the mechanisms behind how humans learn

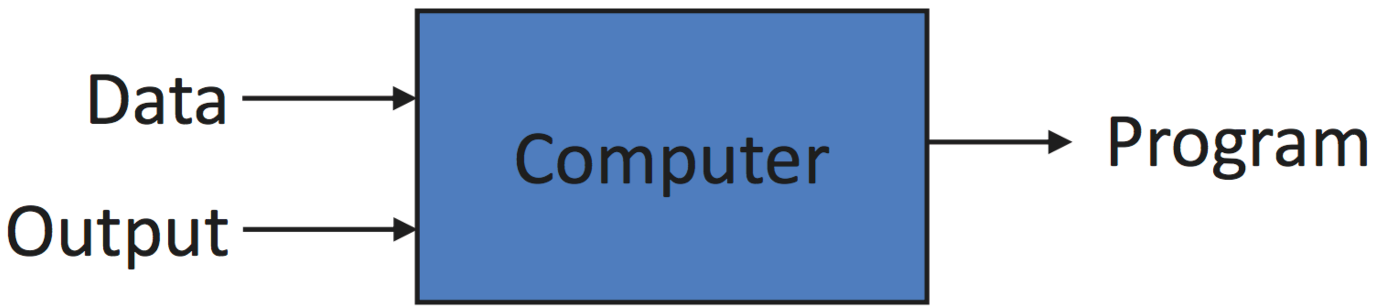
Adapted from “Machine Learning” by Tom Dietterich

One more definition of ML

Traditional Programming



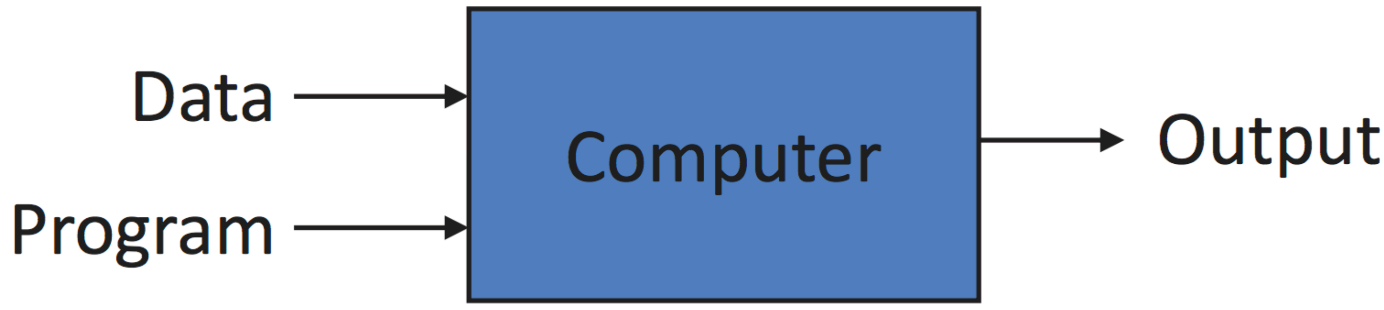
Machine Learning



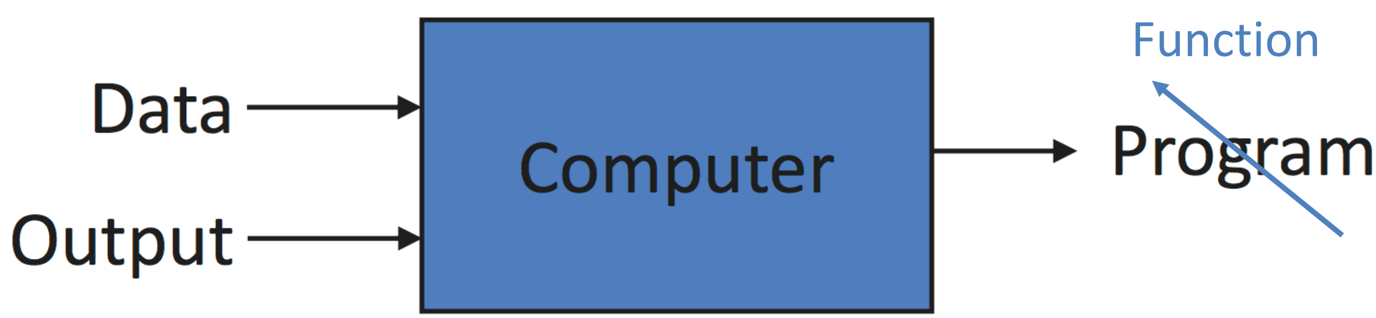
Slide credit: Jessica Wu
Source: Pedro Domingos

One more definition of ML

Traditional Programming



Machine Learning



Slide credit: Jessica Wu
Source: Pedro Domingos

Outline for January 23

- Welcome + what is Machine Learning (ML)?
- Examples of ML
- Syllabus highlights
- ML terminology
- Notation for this class
- First algorithm: K-nearest neighbors

Classic examples of ML

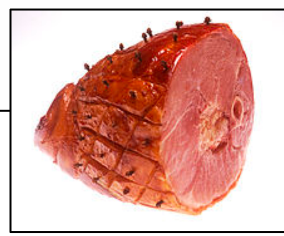
- Email filtering (spam vs. not-spam)

From: cheapsales@buystufffromme.com
To: ang@cs.stanford.edu
Subject: Buy now!

Deal of the week! Buy now!
Rolex w4tchs - $100
Med1cine (any kind) - $50
Also low cost M0rgages
available.

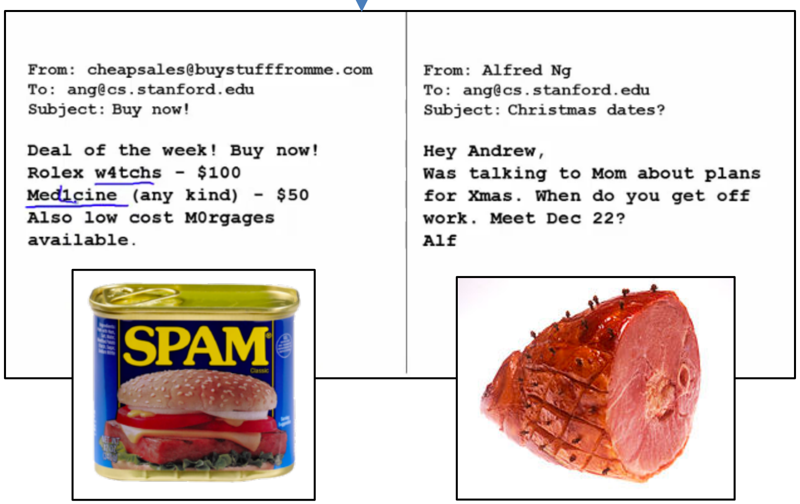
From: Alfred Ng
To: ang@cs.stanford.edu
Subject: Christmas dates?

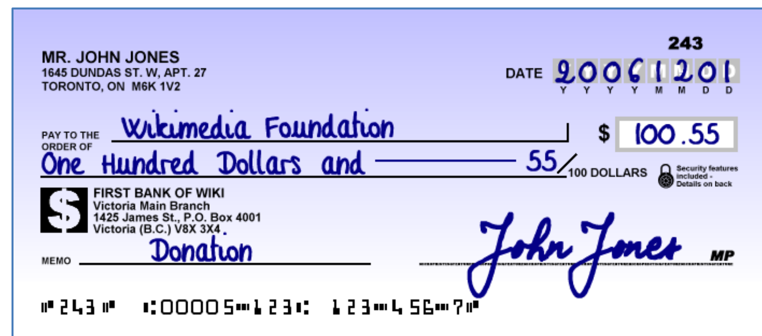
Hey Andrew,
Was talking to Mom about plans
for Xmas. When do you get off
work. Meet Dec 22?
Alf



Classic examples of ML

- Email filtering (spam vs. not-spam)

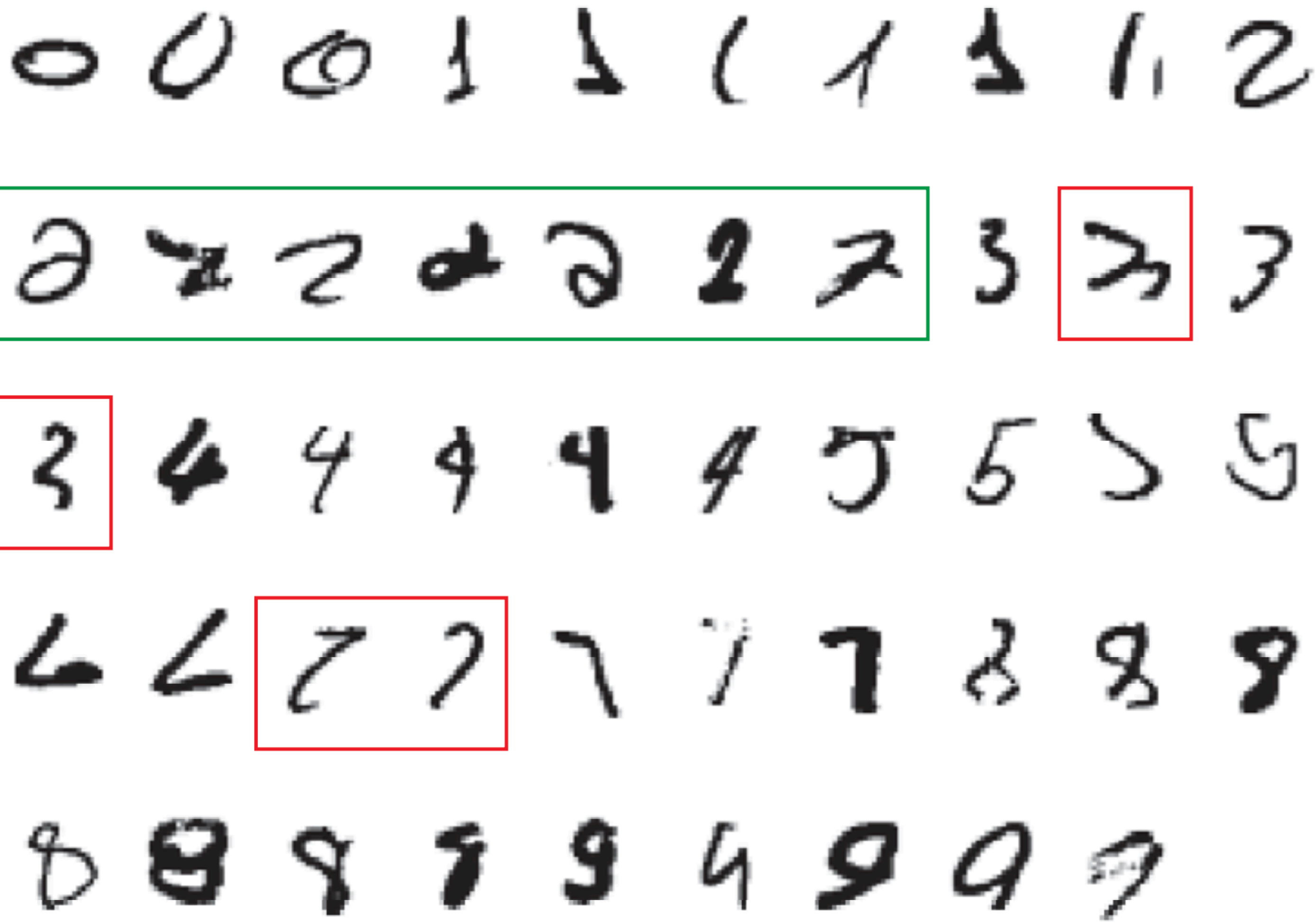




- Handwriting recognition (digits in a check)

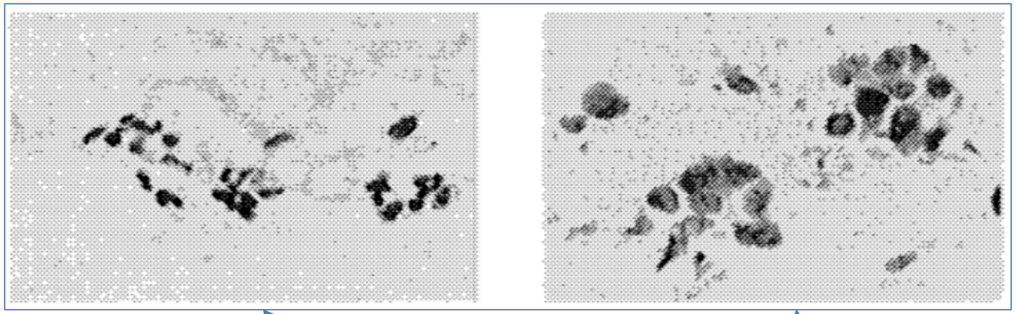
Images: Wikipedia, Matouš Havlena

A classic example of a task that requires machine learning: It is very hard to say what makes a 2



Slide credit: Jessica Wu
Source: Geoffrey Hinton

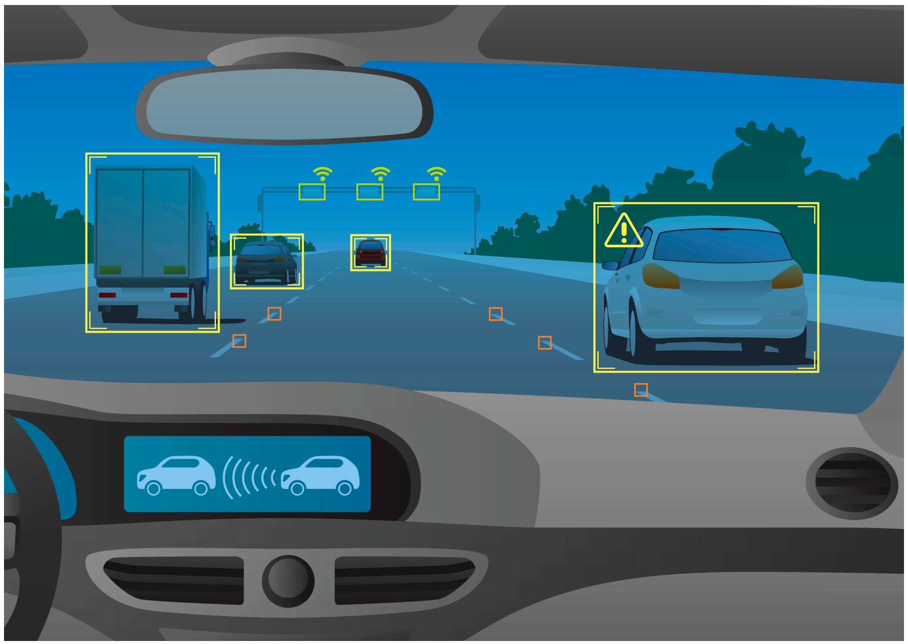
Classic examples of ML



- Tumor detection (benign vs. malignant)

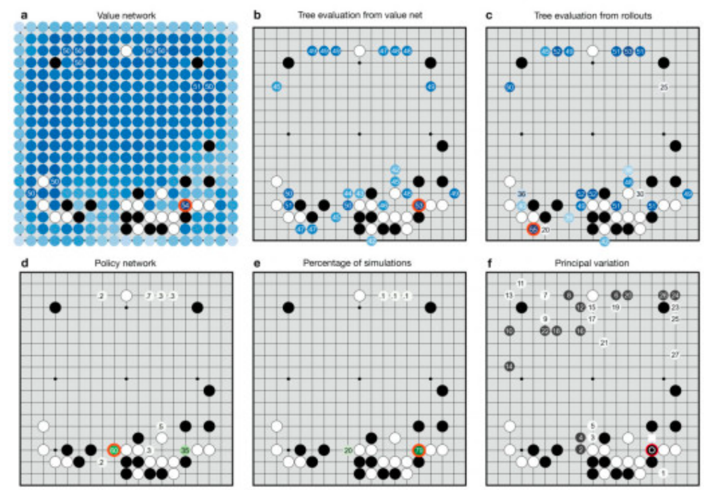
"On Breast Cancer Detection: An Application of Machine Learning Algorithms on the Wisconsin Diagnostic Dataset"

Modern examples of ML



Self-driving cars are in our present and future

AlphaGo: plays humans never thought of



AlphaGo

Images: Scientific American

Modern examples of ML

- Speech-to-text
- Machine translation
- Image caption generation
- Image generation from caption
- ….

Modern examples of ML

- Algorithms that learn how to create

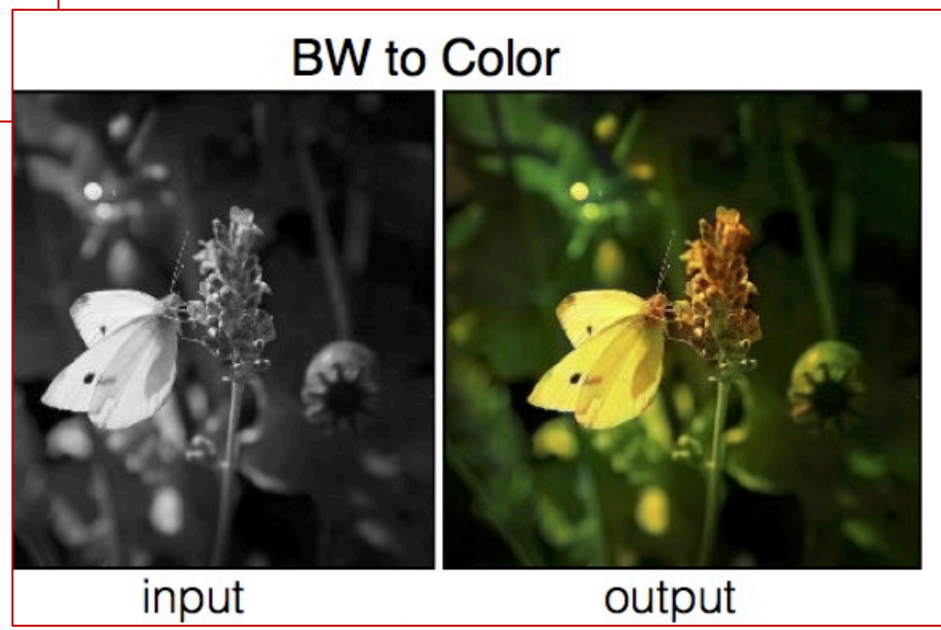


Image-to-Image Translation with Conditional Adversarial Nets (Nov 2016)

Outline for January 23

- Welcome + what is Machine Learning (ML)?
- Examples of ML
- Syllabus highlights

  https://www.cs.swarthmore.edu/~smathieson/teaching/s19/
- ML terminology
- Notation for this class
- First algorithm: K-nearest neighbors

Learning Goals

- First part of the semester: focus on understanding and implementing algorithms

Learning Goals

- First part of the semester: focus on understanding and implementing algorithms

- Later on: using powerful libraries (i.e. TensorFlow, Keras, etc)

Learning Goals

- First part of the semester: focus on understanding and implementing algorithms
- Later on: using powerful libraries (i.e. TensorFlow, Keras, etc)
- Throughout: hypothesis development, algorithm selection, interpretation of results, iteration, conclusions

Learning Goals

- First part of the semester: focus on understanding and implementing algorithms
- Later on: using powerful libraries (i.e. TensorFlow, Keras, etc)
- Throughout: hypothesis development, algorithm selection, interpretation of results, iteration, conclusions
- Language: Python3, will use numpy/scipy throughout

Topics (tentative)

- ML terminology and notation
- K-nearest neighbors
- Decision Trees
- Linear regression
- Logistic regression
- Naïve Bayes
- Ensemble methods
- Support vector machines
- Unsupervised learning
- Dimensionality reduction
- Clustering
- Neural networks
- CNNs
- GANs
- Gaussian mixture models
- Special topics in deep learning
- ML and ethics

Different Backgrounds

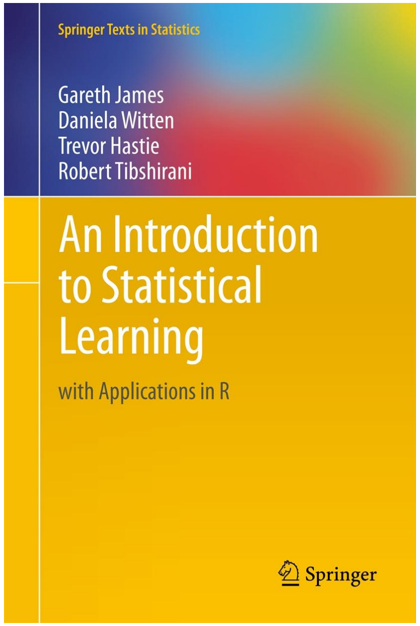
- Prerequisite: CS35
- CS66 could be first or last upper-level course
- May or may not have linear algebra, statistics, probability, etc
- May or may not have taken AI, computer vision, bioinformatics, NLP, etc

There will be math!

- Modern machine learning relies heavily on statistics and probability
- Our textbook is more stats-oriented, but we will not be coding in R
- Supplementary readings with CS focus

There will be math!

- Modern machine learning relies heavily on statistics and probability
- Our textbook is more stats-oriented, but we will not be coding in R
- Supplementary readings with CS focus



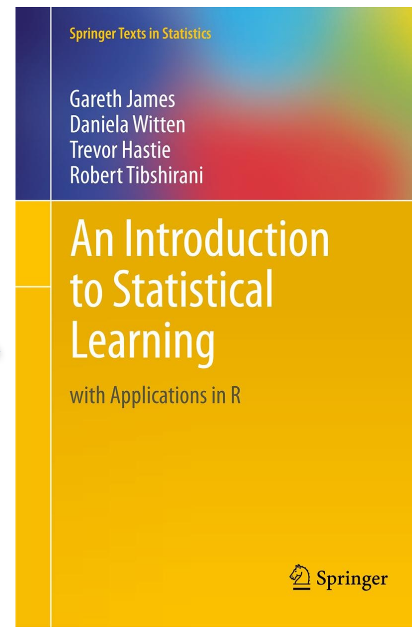
An Introduction to Statistical Learning

with Applications in R

Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani

There will be math!

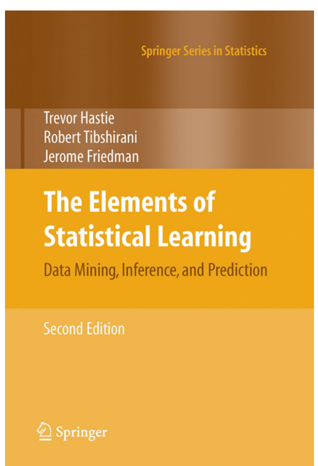
- Modern machine learning relies heavily on statistics and probability
- Our textbook is more stats-oriented, but we will not be coding in R
- Supplementary readings with CS focus



THE ELEMENTS OF STATISTICAL LEARNING

Trevor Hastie, Robert Tibshirani, and Jerome Friedman

Graduate level



An Introduction to Statistical Learning

with Applications in R

Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani

Course Components

- Labs (roughly 8-9 total): 35%
- Midterms (2 in-lab, week 6 and week 13): 40% (20% each)
- Final project: 15% (includes an oral presentation and “lab notebook”)
- Participation: 10%

My Expectations

- Come to class (M/W/F) and lab (W), and actively participate during both
  - Email me if you will be absent from class
  - If you are sick, do not come to class!
- Complete the weekly reading *before* lab (except this week)
- Come to office hours (Mon. 12:30-2pm, Fri. 1-3pm)
  - Occasional meetings outside of office hours are okay (keep the class size in mind)
- Post questions on Piazza

My Expectations

- Come to class (M/W/F) and lab (W), and actively participate during both
  - Email me if you will be absent from class
  - If you are sick, do not come to class!
- Complete the weekly reading *before* lab (except this week)
- Come to office hours (Mon. 12:30-2pm, Fri. 1-3pm)
  - Occasional meetings outside of office hours are okay (keep the class size in mind)
- Post questions on Piazza

WEEK	DAY	ANNOUNCEMENTS	TOPIC & READING	LABS
1	Jan 21	MLK Day - NO CLASS	**Introduction to Machine Learning** • Machine learning terminology • Notation • K-nearest neighbors Reading: • **Machine Learning** by Tom Dietterich in *Nature Encyclopedia of Cognitive Science* (skim Sections 4-7) • ISL: Sections 2.1, 2.2 (focus on 2.1 and pg. 39-42) • (*optional*) ISL: Chapter 1 • (*optional*) **The Discipline of Machine Learning** by Tom Mitchell	**Lab 1**: K-nearest neighbors
	Jan 23			
	Jan 25			

Syllabus Highlights

(Note: you are responsible for reading the entire syllabus on the course webpage)

Syllabus Highlights

(Note: you are responsible for reading the entire syllabus on the course webpage)

1. Notes and slides will be posted *after* class on the course webpage

Syllabus Highlights

(Note: you are responsible for reading the entire syllabus on the course webpage)

1. Notes and slides will be posted *after* class on the course webpage
2. Lab is **mandatory** (attendance will be taken)

Syllabus Highlights

(Note: you are responsible for reading the entire syllabus on the course webpage)

1. Notes and slides will be posted *after* class on the course webpage
2. Lab is **mandatory** (attendance will be taken)
3. Labs will be done in **pairs** (randomly assigned for the first lab)

Syllabus Highlights

(Note: you are responsible for reading the entire syllabus on the course webpage)

1. Notes and slides will be posted *after* class on the course webpage
2. Lab is **mandatory** (attendance will be taken)
3. Labs will be done in **pairs** (randomly assigned for the first lab)
4. You will get **2 late days** during the semester (counts for both partners)

Syllabus Highlights

(Note: you are responsible for reading the entire syllabus on the course webpage)

1. Notes and slides will be posted *after* class on the course webpage
2. Lab is **mandatory** (attendance will be taken)
3. Labs will be done in **pairs** (randomly assigned for the first lab
4. You will get **2 late days** during the semester (counts for both partners)
5. Extensions beyond these two days must be arranged with your class dean, **no exceptions**

Syllabus Highlights

(Note: you are responsible for reading the entire syllabus on the course webpage)

1. Notes and slides will be posted *after* class on the course webpage
2. Lab is **mandatory** (attendance will be taken)
3. Labs will be done in **pairs** (randomly assigned for the first lab
4. You will get **2 late days** during the semester (counts for both partners)
5. Extensions beyond these two days must be arranged with your class dean, **no exceptions**
6. Email: allow 24 hours for a response (more during weekends)

Syllabus Highlights

(Note: you are responsible for reading the entire syllabus on the course webpage)

1. Notes and slides will be posted *after* class on the course webpage
2. Lab is **mandatory** (attendance will be taken)
3. Labs will be done in **pairs** (randomly assigned for the first lab
4. You will get **2 late days** during the semester (counts for both partners)
5. Extensions beyond these two days must be arranged with your class dean, **no exceptions**
6. Email: allow 24 hours for a response (more during weekends)
7. Piazza: should be used for all content/logistics questions

Participation

What counts as participation?

- Asking and answering questions in class (very important!)
  - Raise your hand (because some people are more/less comfortable shouting out answers)
  - Will call on groups, but only after giving you a few minutes to think out loud in pairs

Participation

What counts as participation?

- Asking and answering questions in class (very important!)
  - Raise your hand (because some people are more/less comfortable shouting out answers)
  - Will call on groups, but only after giving you a few minutes to think out loud in pairs

- Actively participating in in-class activities (group work, handouts, etc)

Participation

What counts as participation?

- Asking and answering questions in class (very important!)
  - Raise your hand (because some people are more/less comfortable shouting out answers)
  - Will call on groups, but only after giving you a few minutes to think out loud in pairs
- Actively participating in in-class activities (group work, handouts, etc)
- Working well with your lab partner during lab
  - Switching “driver” and “navigator” frequently (at least every 30 min)
  - Discussing details instead of just trying to get to the end of the lab

Participation

What counts as participation?

- Asking and answering questions in class (very important!)
  - Raise your hand (because some people are more/less comfortable shouting out answers)
  - Will call on groups, but only after giving you a few minutes to think out loud in pairs
- Actively participating in in-class activities (group work, handouts, etc)
- Working well with your lab partner during lab
  - Switching “driver” and “navigator” frequently (at least every 30 min)
  - Discussing details instead of just trying to get to the end of the lab
- Asking and answering questions on Piazza
  - Avoid long blocks of code and giving away answers
  - Only non-anonymous posts count toward participation grade

Participation

What counts as participation?

- Asking and answering questions in class (very important!)
  - Raise your hand (because some people are more/less comfortable shouting out answers)
  - Will call on groups, but only after giving you a few minutes to think out loud in pairs

- Actively participating in in-class activities (group work, handouts, etc)

- Working well with your lab partner during lab
  - Switching "driver" and "navigator" frequently (at least every 30 min)
  - Discussing details instead of just trying to get to the end of the lab

- Asking and answering questions on Piazza
  - Avoid long blocks of code and giving away answers
  - Only non-anonymous posts count toward participation grade

- Attending office hours

Participation

What counts as participation?

- Asking and answering questions in class (very important!)
  - Raise your hand (because some people are more/less comfortable shouting out answers)
  - Will call on groups, but only after giving you a few minutes to think out loud in pairs
- Actively participating in in-class activities (group work, handouts, etc)
- Working well with your lab partner during lab
  - Switching “driver” and “navigator” frequently (at least every 30 min)
  - Discussing details instead of just trying to get to the end of the lab
- Asking and answering questions on Piazza
  - Avoid long blocks of code and giving away answers
  - Only non-anonymous posts count toward participation grade
- Attending office hours

Sometimes participation goes too far…

- Try to avoid dominating class discussion, office hours, Piazza, pair-programming, etc

Academic Integrity

Discussing ideas and approaches to problems with others on a general level is fine (in fact, we encourage you to discuss general strategies with each other), but you should never read anyone else's code or let anyone else read your code.

- No code from online
- No code from students who took this course previously

Class Deans

Karen Henry '87

ASSISTANT DEAN/FRESHMAN CLASS & DIRECTOR OF FIRST IN FAMILY/FIRST GEN., LOW INCOME STUDENT INITIATIVES

khenry1@swarthmore.edu
(610) 328-8169
Office: Parrish 130

Profile ›

Thomas Alexander

INTERIM ASST. DEAN /SOPHOMORE CLASS, INTERIM DIRECTOR OF THE INTERCULTURAL CENTER

Intercultural Center

talexan3@swarthmore.edu
(610) 328-7360

Profile ›

Dion Lewis

ASSISTANT DEAN/JUNIOR CLASS, DIRECTOR OF THE BLACK CULTURAL CENTER

Black Cultural Center
Black Cultural Center

dlewis1@swarthmore.edu
(610) 328-8455
Black Cultural Center Main
Parrish Hall 122

Profile ›

Michelle D. Ray

ASSISTANT DEAN/SENIOR CLASS, DIRECTOR OF CASE MANAGEMENT, DIRECTOR OF THE DIVERSITY PEER ADVISORS

Dean's Office

mray2@swarthmore.edu
(610) 690-5299
Parrish Hall 116

Profile ›

Disability Services

https://www.swarthmore.edu/academic-advising-support/welcome-to-student-disability-services

Registering with the Student Disability Service

Please contact the Student Disability Services staff at studentdisabilityservices@swarthmore.edu, call Monica Vance at 610-328-7358, or Jenna Rose at 610-690-5538 to arrange an intake appointment. We are happy to hold initial appointments for incoming students by phone. If at all possible, please submit documentation of your disability in advance so that we can review it prior to talking with you. We recommend that you contact us as early as possible since some accommodations (e.g., electronic books, interpreters, etc.) can take a several weeks to arrange. We want to be sure that your needs are met in time for classes.

Outline for January 23

- Welcome + what is Machine Learning (ML)?
- Examples of ML
- Syllabus highlights
- ML terminology
- Notation for this class
- First algorithm: K-nearest neighbors

Machine learning terminology

- One common style of machine learning is *classification*

- Goal: separate examples into two or more *classes* or *categories* (*discrete* setting)

Example: the classic bagel v. dog challenge

How can we *distinguish* between similar objects?

Images: veriy.com

Machine learning terminology

- Another common ML task is *regression*

- In this case the output or *response variable* is *continuous*

Machine learning terminology

- Another common ML task is *regression*

- In this case the output or *response variable* is *continuous*

Example: modeling **house price** as a function of **size, location, year built**, etc

Machine learning terminology

- *Supervised learning:* we have information about the output or response variable
  – (can be easier for the computer to learn the function between input and output)

- *Unsupervised learning:* data is unlabeled (no output/class information)
  – Note: there may not be an output to learn

Machine learning terminology

- *Training*: usually involves the program learning from many *examples* (in the supervised setting we know the “answer” or *label* and are using this to learn)

- *Testing*: program predicts output/label for new examples without using their labels

Machine learning terminology

- *Training*: usually involves the program learning from many *examples* (in the supervised setting we know the "answer" or *label* and are using this to learn)

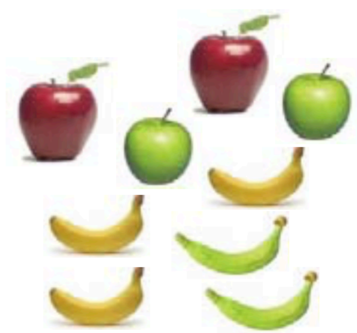
- *Testing*: program predicts output/label for new examples without using their labels

Caveat: not all ML problems decompose into training and testing!

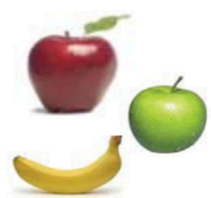
Learning System

- Learning is about ***generalizing*** from training data
- What does this ***assume*** about training and test set?

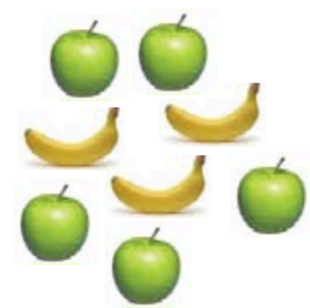
Training data



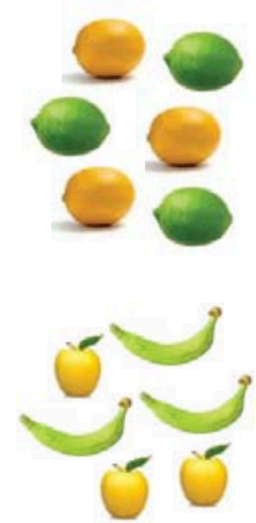
Test set



Training data



Test set



Not always the case, but we'll often assume it is!

Slide credit: Jessica Wu
Source based on: David Kauchak

Outline for January 23

- Welcome + what is Machine Learning (ML)?
- Examples of ML
- Syllabus highlights
- ML terminology
- Notation for this class
- First algorithm: K-nearest neighbors

ML Notation

$$X = \underbrace{\begin{bmatrix} \\ \text{—}\; x_i \;\text{—} \\ \\ \end{bmatrix}}_{p\,=\,\#\text{ features}} \left.\vphantom{\begin{bmatrix} \\ \\ \\ \end{bmatrix}}\right\}\; n = \#\text{ data points}$$

input/
observations

y

(size, location, year built)

x = representative example

X_{train} = training data (75-80%)

X_{test} = test data (20-25%)

-5-80%)

25%)

y = label/output/class

$l = \#$

discrete: $y \in \{0, 1, 2, \ldots, l-1\}$,

in

$$Y = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

continuous: $y \in \mathbb{R}^n$

real

example: house price

price + Δ

example

$$y = \underbrace{\begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}}_{\text{truth}}, \quad \hat{y} = \underbrace{\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}}_{\text{prediction}}$$

$$\text{accuracy} = \frac{1}{5}(1 + 0 + 1 + 1 + 0)$$

$$= \boxed{\frac{3}{5}}$$