CS 383: Machine Learning

Prof Adam Poliak

Fall 2024

Outline

- Welcome + what is Machine Learning (ML)?
- Examples of ML
- Syllabus highlights
- ML terminology and notation
- First algorithm: K-nearest neighbors

Outline

- Welcome + what is Machine Learning (ML)?
- Examples of ML
- Syllabus highlights
- ML terminology and notation
- First algorithm: K-nearest neighbors

What is ML?

- 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 ML?

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

-Tom Dietterich

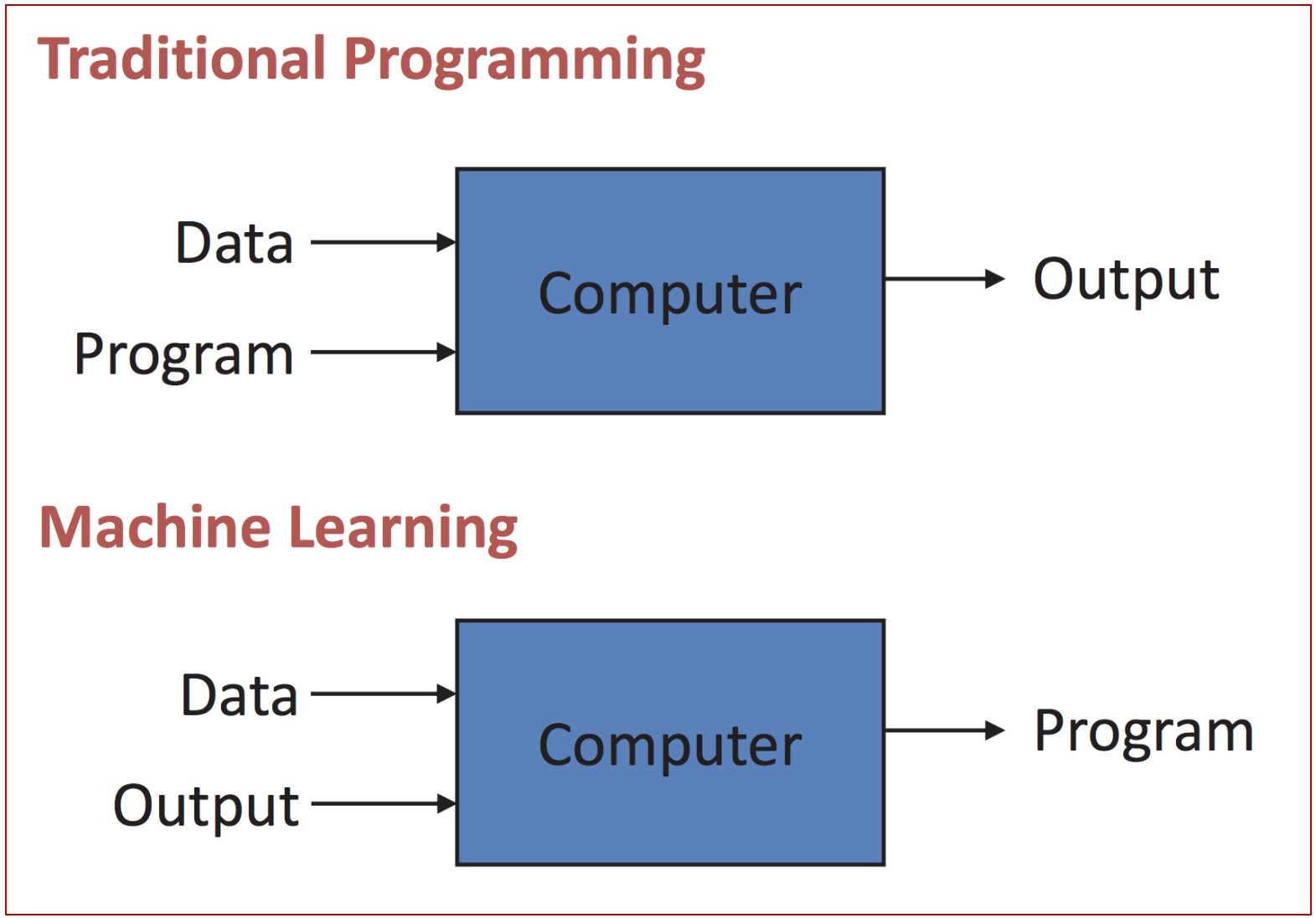
- “Machine Learning is about predicting the future based on the past.”

-Hal Duame III

- “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

Another definition of ML



Slide credit: Jessica Wu
Source: Pedro Domingos

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

Why do we want ML?

1) No human experts

   Example: predicting failure points for new machines

2) Human experts cannot explain expertise

   Example: predicting failure points for new machines

3) Phenomena change rapidly

   Example: predicting failure points for new machines

4) Customization for each user

   Example: program that adapts to each user's speech

Outline

- Welcome + what is Machine Learning (ML)?
- Examples of ML
- Syllabus highlights
- ML terminology and notation
- 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!

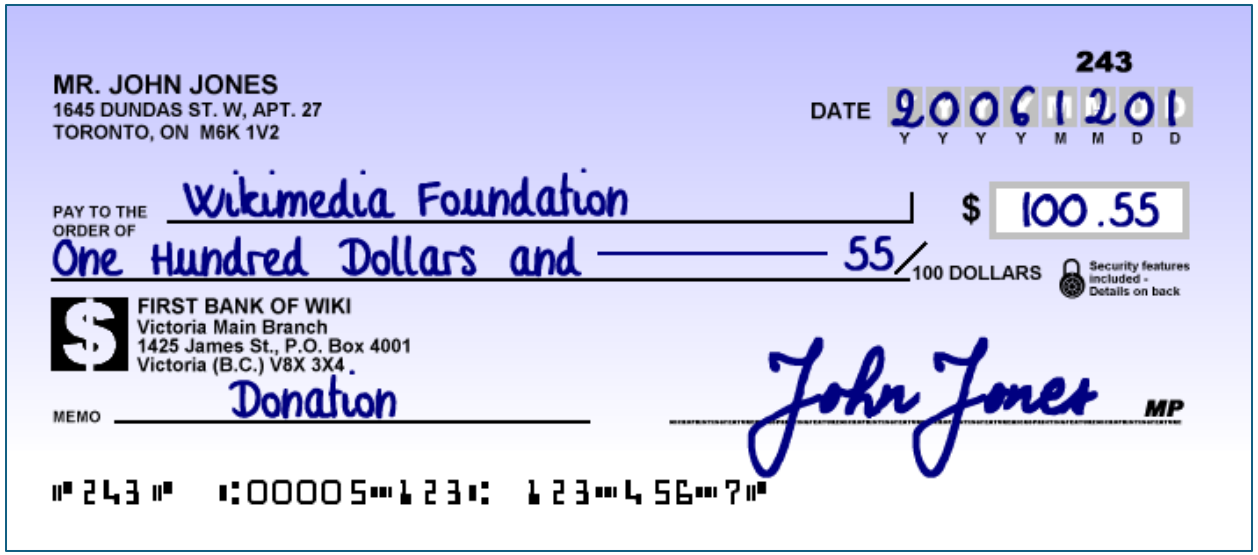
Deal of the week! Buy now!
Rolex w4tchs - $100
Med1cine (any kind) - $50
Also low cost M0rgages
available.

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

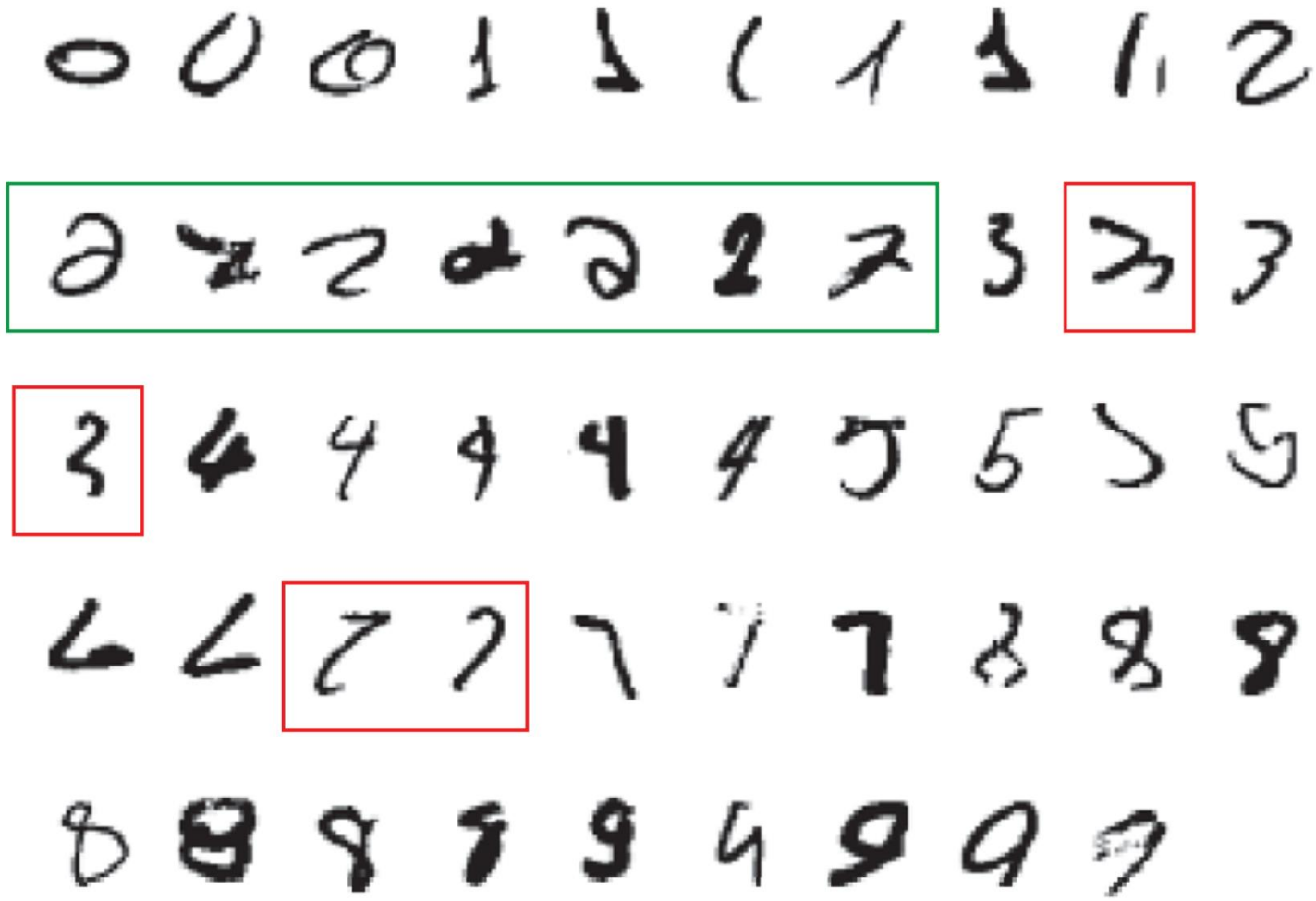
Classic Examples of ML

- Handwriting recognition (digits in a check)



A classic example of a task that requires machine learning:

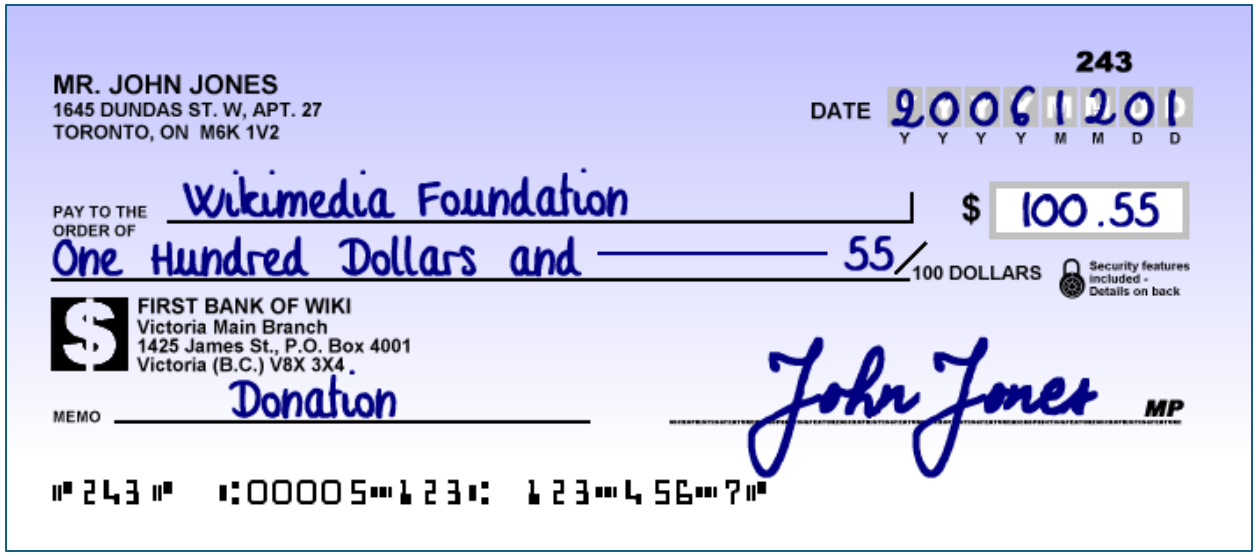
It is very hard to say what makes a 2



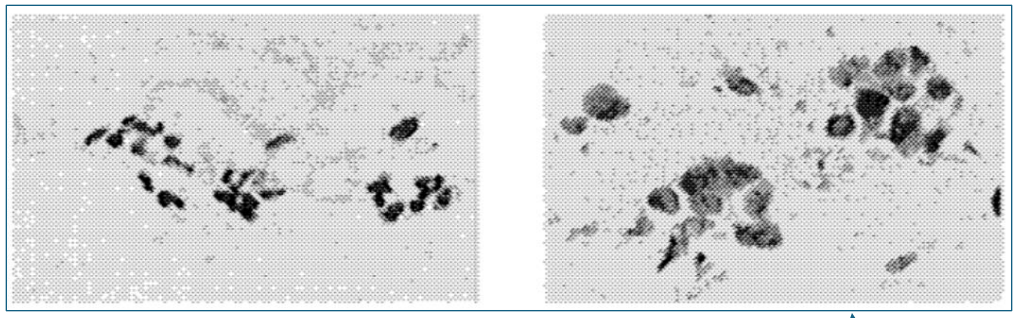
Slide credit: Jessica Wu
Source: Geoffrey Hinton

Classic Examples of ML

- Handwriting recognition (digits in a check)



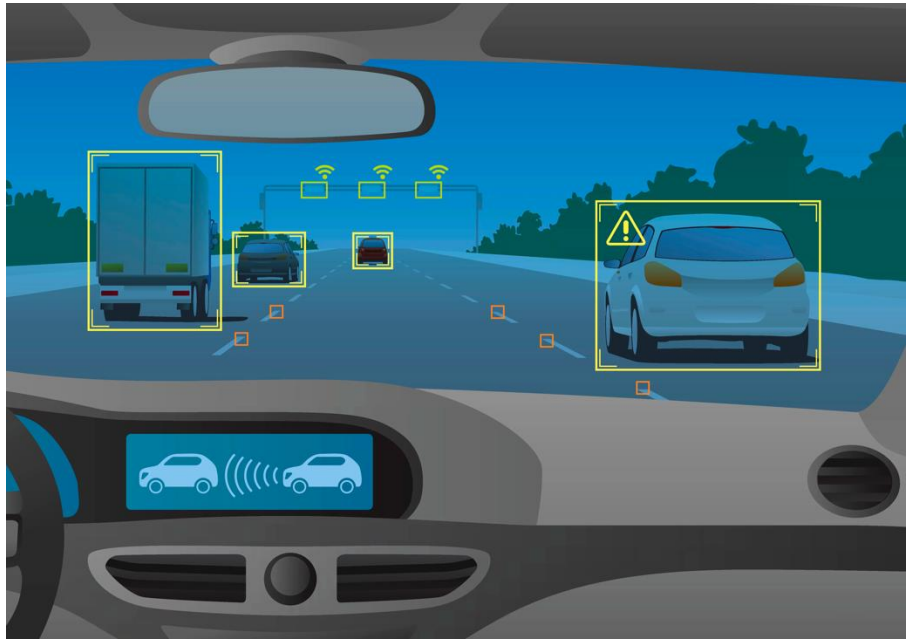
Classic Examples of ML



- Tumor detection (benign vs. malignant)

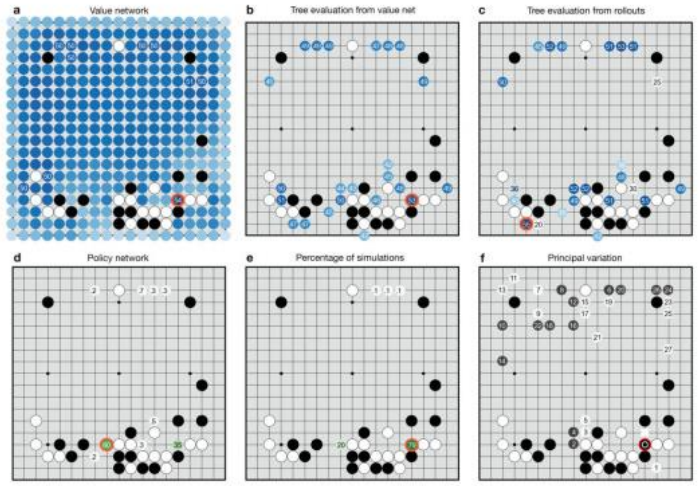
"On Breast Cancer Detection: An Application of Machine Learning Algorithms on the Wisconsin Diagnostic Dataset"

Modern Examples of ML



Self-driving cars are in our present and future

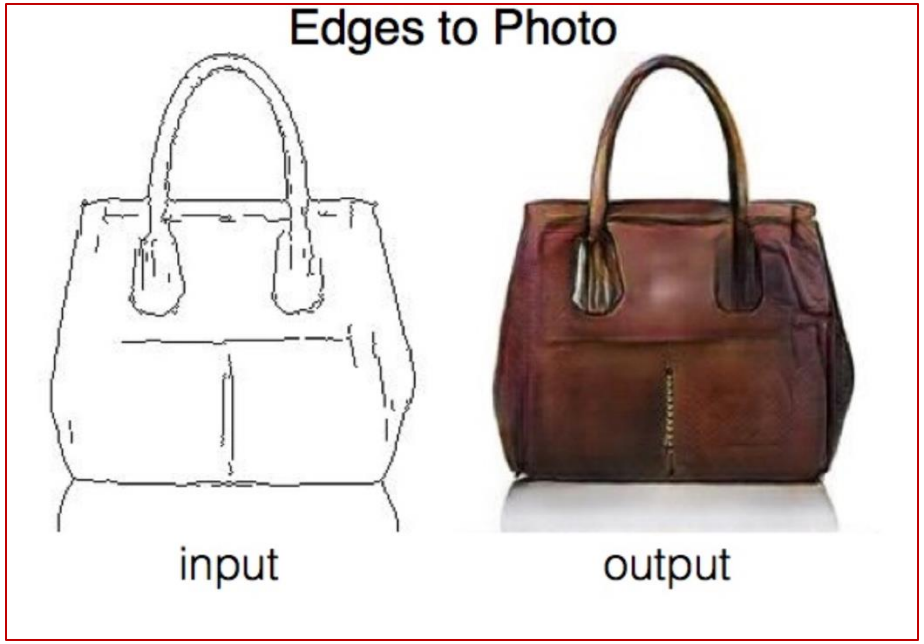
AlphaGo: moves humans never thought of



AlphaGo

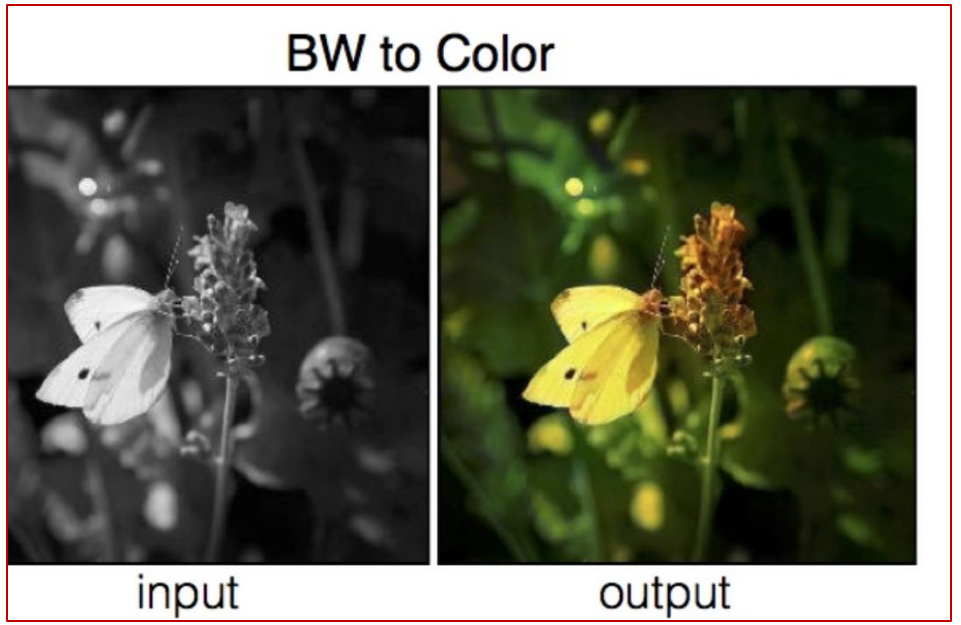
Images: Scientific American

Modern Examples of ML



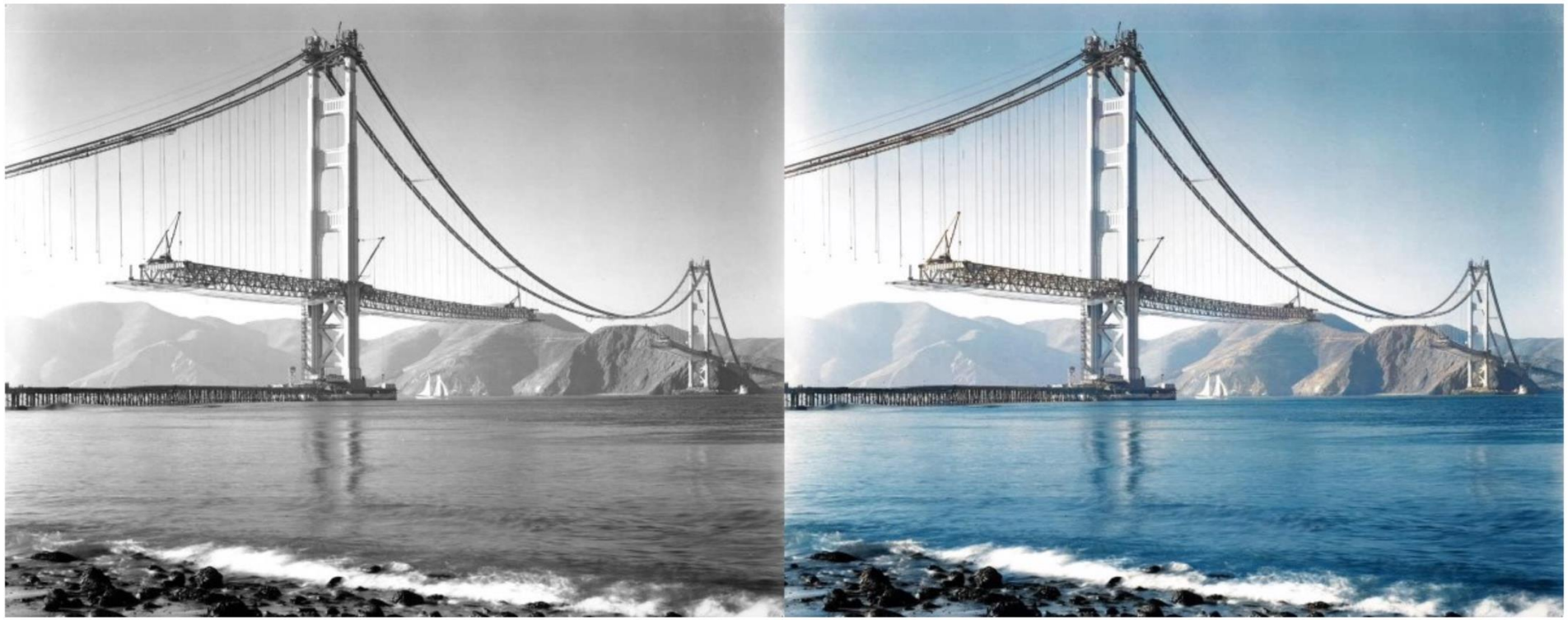
- Algorithms that learn how to create

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



Modern Examples of ML

"Building the Golden Gate Bridge" (est 1937)

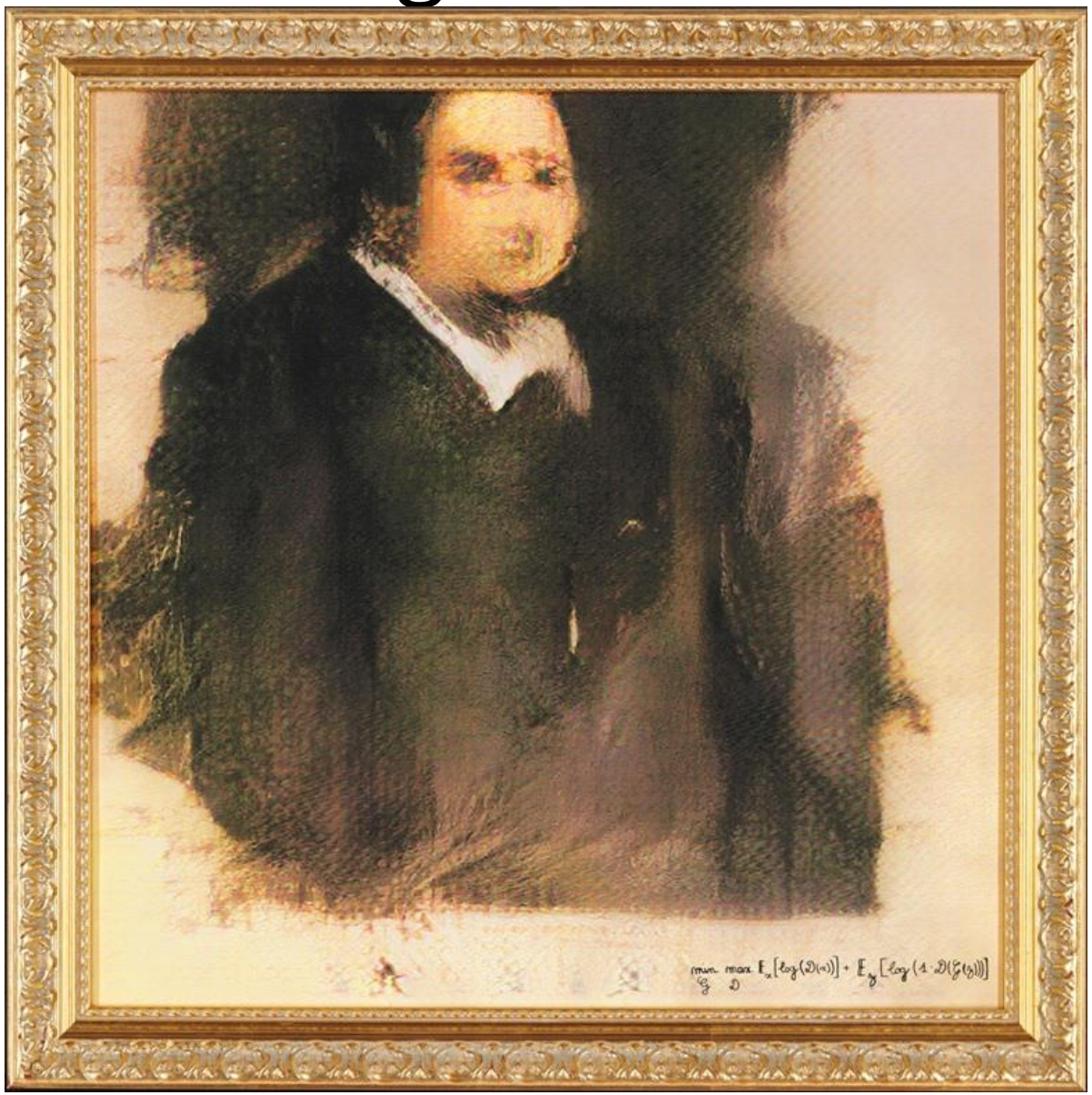


https://github.com/jantic/DeOldify

Face Generation Over Time

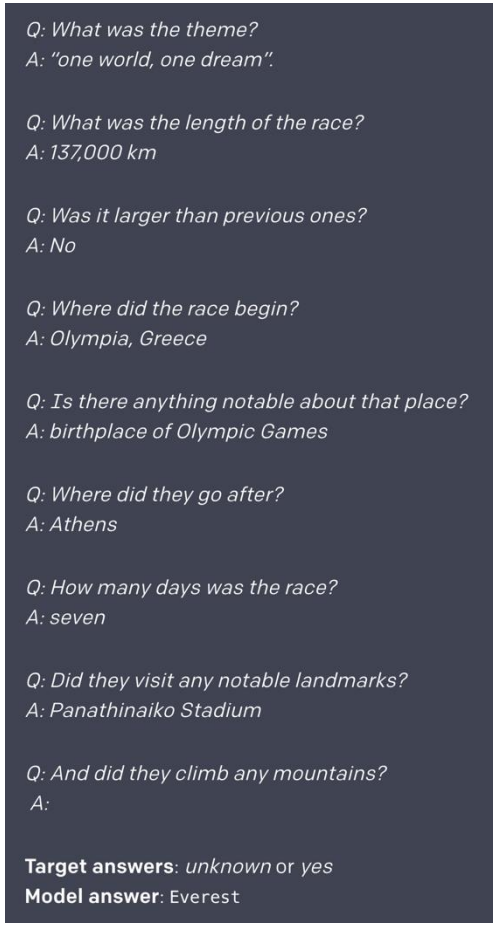
ML Generated Painting

Sold for almost half a million dollars



ML in Natural Language Processing

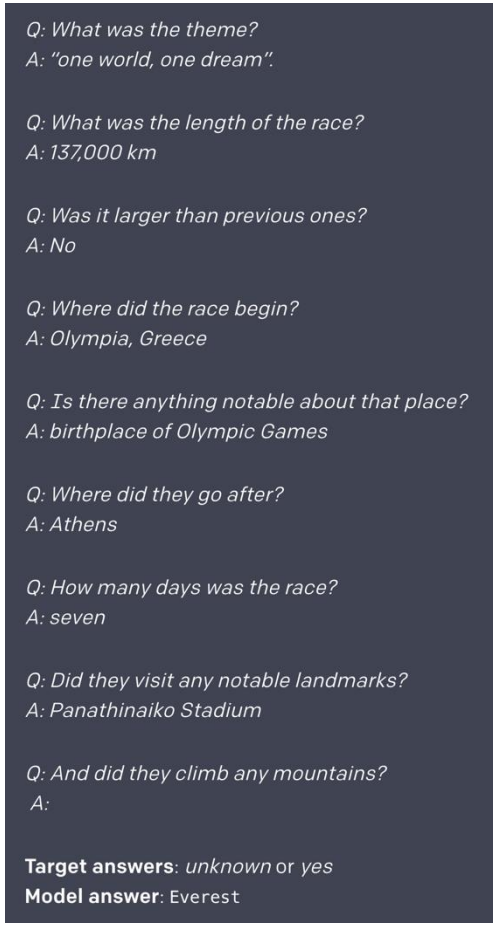
- Text Classification
- Language Modeling
- Speech Recognition
- Caption Generation
- Machine Translation
- Document Summarization
- Question Answering



https://openai.com/blog/better-language-models/

ML in Natural Language Processing

- Text Classification
- Language Modeling
- Speech Recognition
- Caption Generation
- Machine Translation
- Document Summarization
- Question Answering
- Dialogue Systems



https://openai.com/blog/better-language-models/

ML Online

Every day ML algorithms decide what people see and how their content is seen

- Search results
- Targeted ads
- Newsfeed content
- Facial recognition

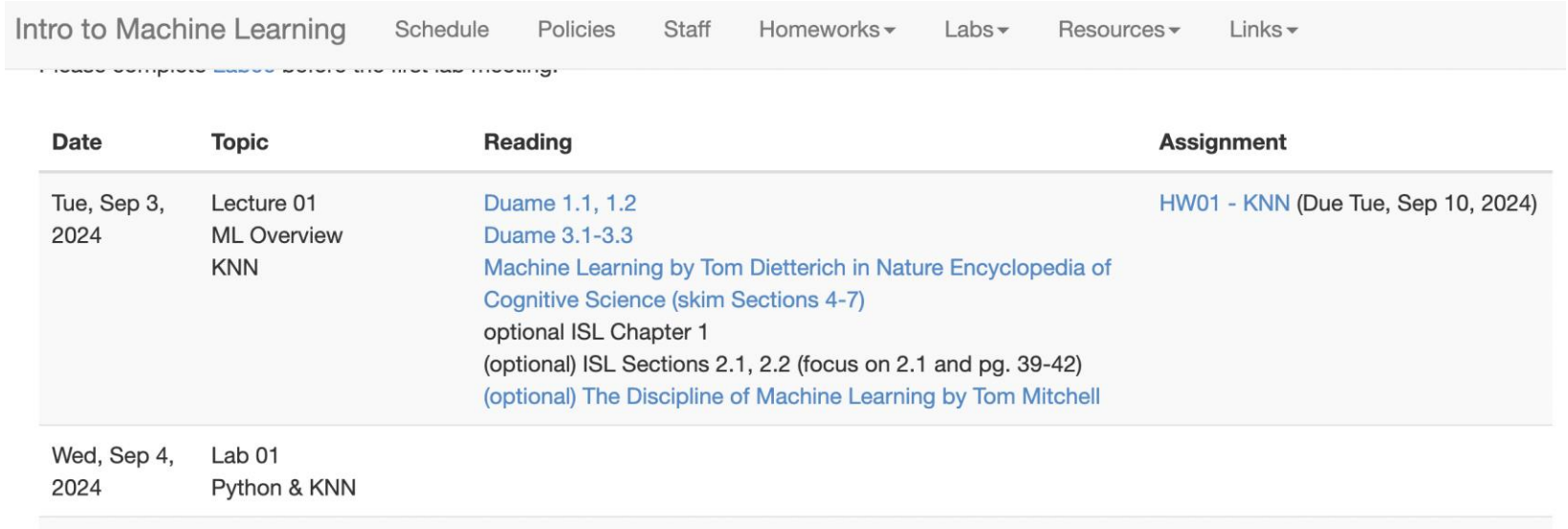
Example:

Outline

- Welcome + what is Machine Learning (ML)?
- Examples of ML
- Syllabus highlights
- ML terminology and notation
- First algorithm: K-nearest neighbors

Course Communication

https://cs.brynmawr.edu/cs383-ml/



Intro to Machine Learning Schedule Policies Staff Homeworks Labs Resources Links

Date	Topic	Reading	Assignment
Tue, Sep 3, 2024	Lecture 01 ML Overview KNN	Duame 1.1, 1.2 Duame 3.1-3.3 Machine Learning by Tom Dietterich in Nature Encyclopedia of Cognitive Science (skim Sections 4-7) optional ISL Chapter 1 (optional) ISL Sections 2.1, 2.2 (focus on 2.1 and pg. 39-42) (optional) The Discipline of Machine Learning by Tom Mitchell	HW01 - KNN (Due Tue, Sep 10, 2024)
Wed, Sep 4, 2024	Lab 01 Python & KNN		

Piazza vs Slack?

Learning Goal

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

Later on: using powerful libraries (i.e. sklearn, pytorch, etc)

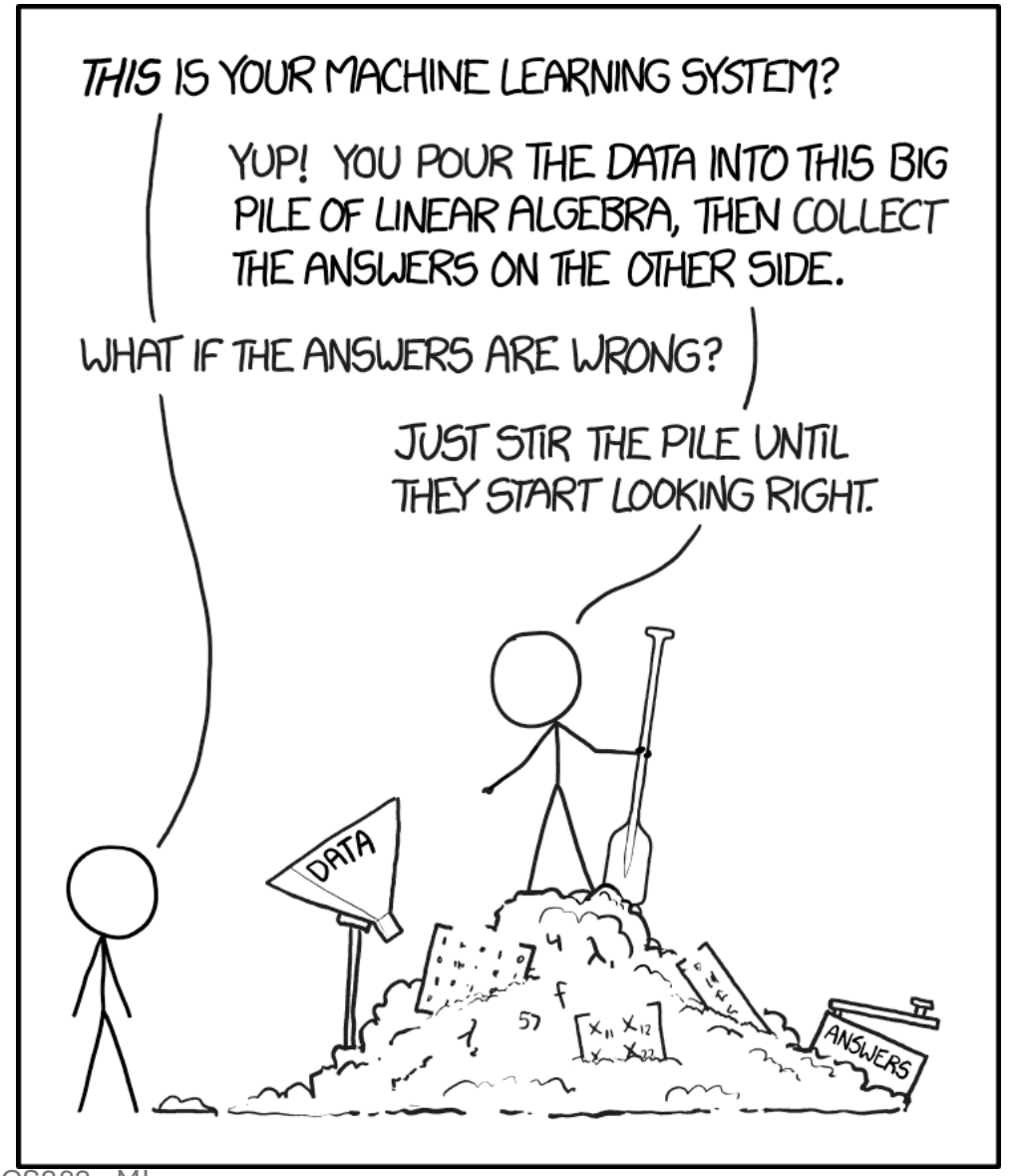
Throughout and during the project: hypothesis development, featurization, algorithm selection, interpretation of results, iteration, conclusions

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
- Neural networks
- LSTMs
- Transformers
- Unsupervised learning
- Dimensionality reduction
- Clustering
- Gaussian mixture models
- ML and ethics

There will be math!



Different Backgrounds

Prerequisites: Data Structures, Discrete Math, Linear Algebra

May or may not have statistics or probability

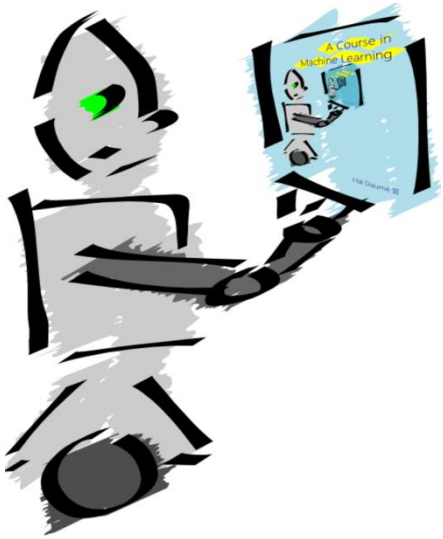
May or may not have taken artificial intelligence (AI), computer vision, bioinformatics, natural language processing (NLP)

Reading

Modern machine learning relies heavily on linear algebra, geometry, statistics, and probability

Our textbook provides a good introduction to terminology, how people think about ML, etc

Supplemental readings go deeper



A Course in Machine Learning

by Hal Daumé III

Course Components

Homeworks (roughly 8-10 total): 35%

Due Tuesday nights

Labs (weekly, some week will be lecture): 5%

2 Midterms: 40%

Final project: 15% (includes an oral presentation and “lab notebook”)

Participation (includes reading quizzes): 5%

Expectations

Come to class (Tu/Th) and lab (W), **actively participate**

If you are sick, do not come to class!

Complete the weekly reading *before lecture* (for this week, just read **1.1-1.2**)

Come to office hours (this week: Wedn. 1:00-2:30pm, Thurs. 11:45am-12:30pm)

Park 200C

Post questions on Piazza/Slack

Syllabus Notes

(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 are partner work
4. Homeworks are individual work
5. You will get **2 late days** during the semester
6. Extensions beyond these two days must be arranged with your class dean
7. Email: allow 24 hours for a response (more during weekends)
8. Piazza/Slack: should be used for all content/logistics questions

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/discuss

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

Working well with your lab partner during lab

- Switching who is at the keyboard
- Discussing details instead of just trying to get to the end of the lab

Asking and answering questions on Piazza/Slack

- 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

Outline

- Welcome + what is Machine Learning (ML)?
- Examples of ML
- Syllabus highlights
- ML terminology and notation
- First algorithm: K-nearest neighbors

Machine Learning Terminology

Alice takes ML for a semester, but the exam is on History of Pottery
All exam questions are exactly the same as homework questions

Neither is a good judge of Alice's learning!

"*Generalization* is perhaps the most central concept in machine learning." –Duame

Machine Learning Terminology

"*Generalization* is perhaps the most central concept in machine learning." –Duame

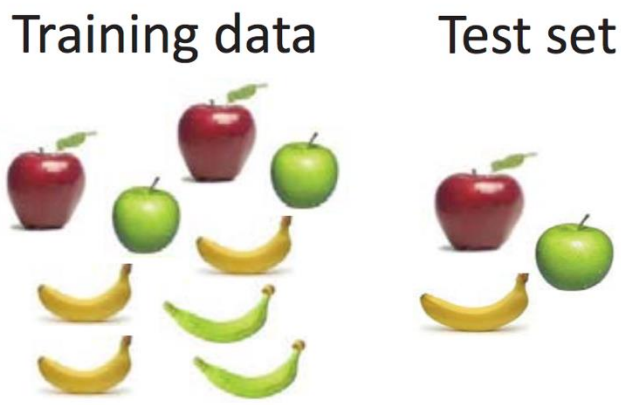


Image: modified from Jessica Wu
Source based on: David Kauchak

Machine Learning Terminology

"*Generalization* is perhaps the most central concept in machine learning." –Duame

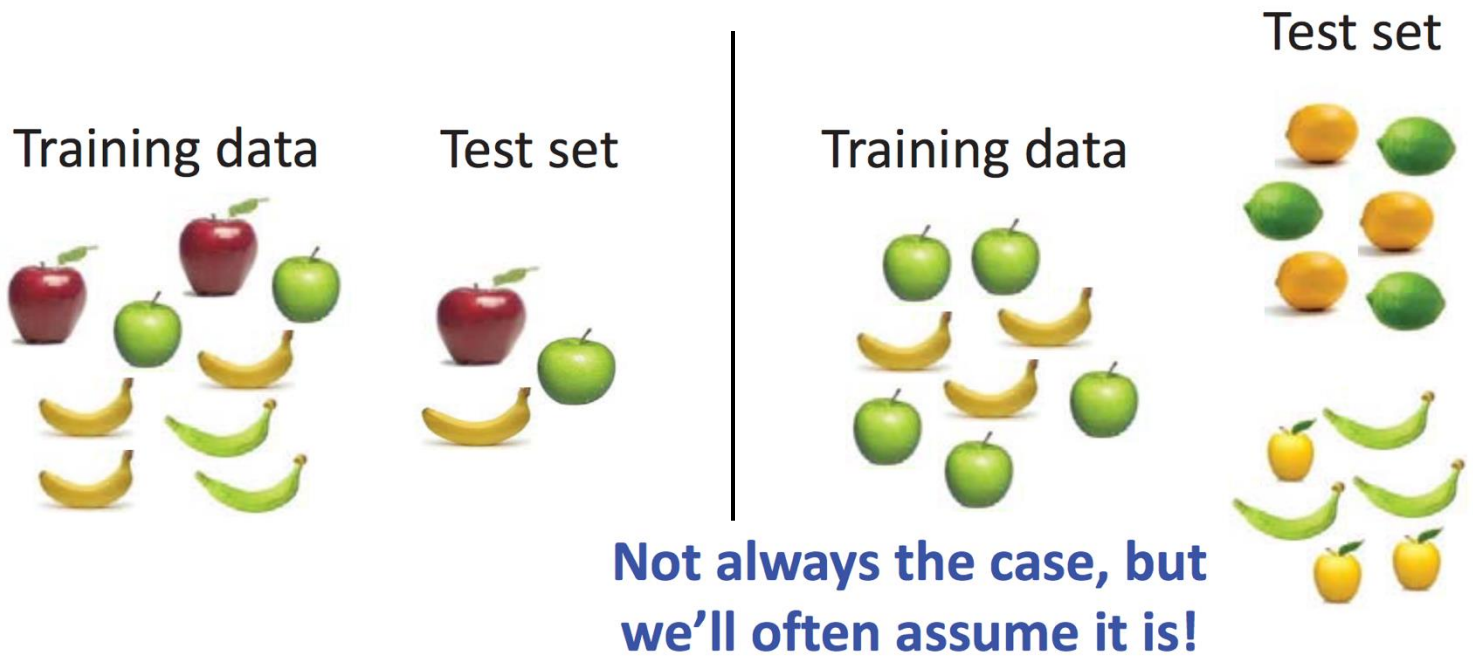


Image: modified from Jessica Wu
Source based on: David Kauchak

Machine Learning Terminology

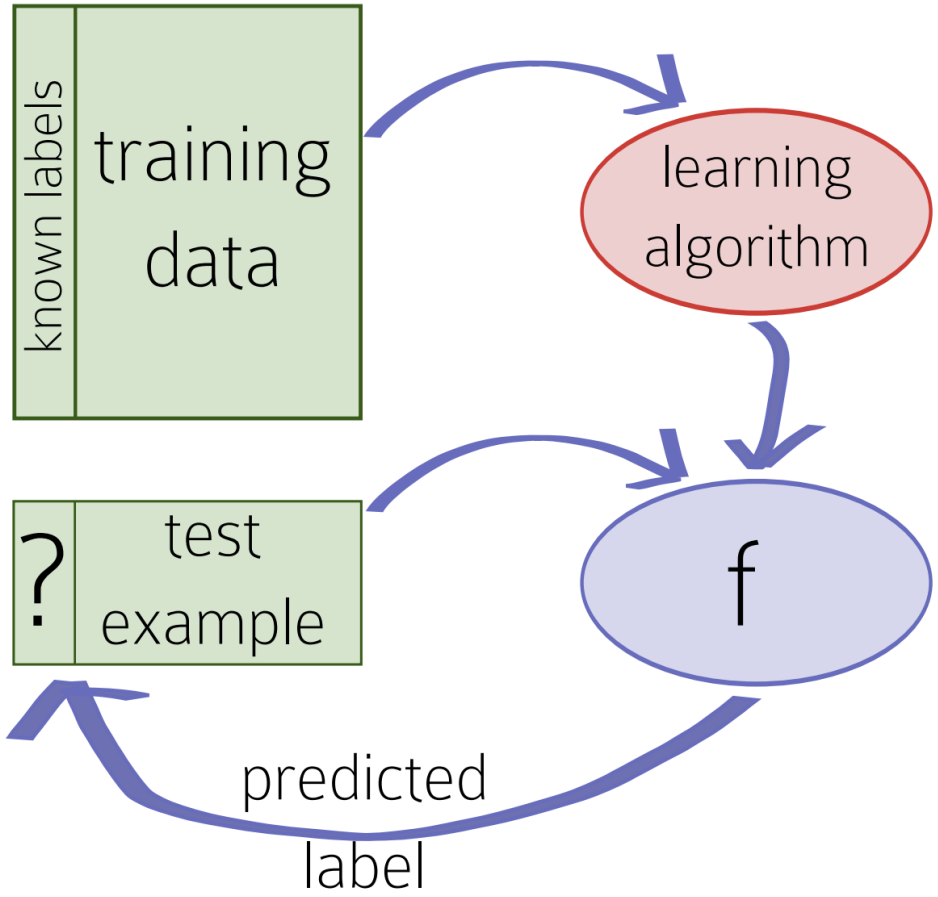


Figure 1.1, Duame

Machine Learning Terminology

Training: usually involves the program learning from many *examples* (in a 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

Must never look at the test data!

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

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

Regression:

output or *response variable* is *continuous*

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

Classifications:

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

Example: is a credit card transaction legitimate or fraudulent?

Classification Example: Bagel vs Dog

How can we *distinguish* between similar objects?

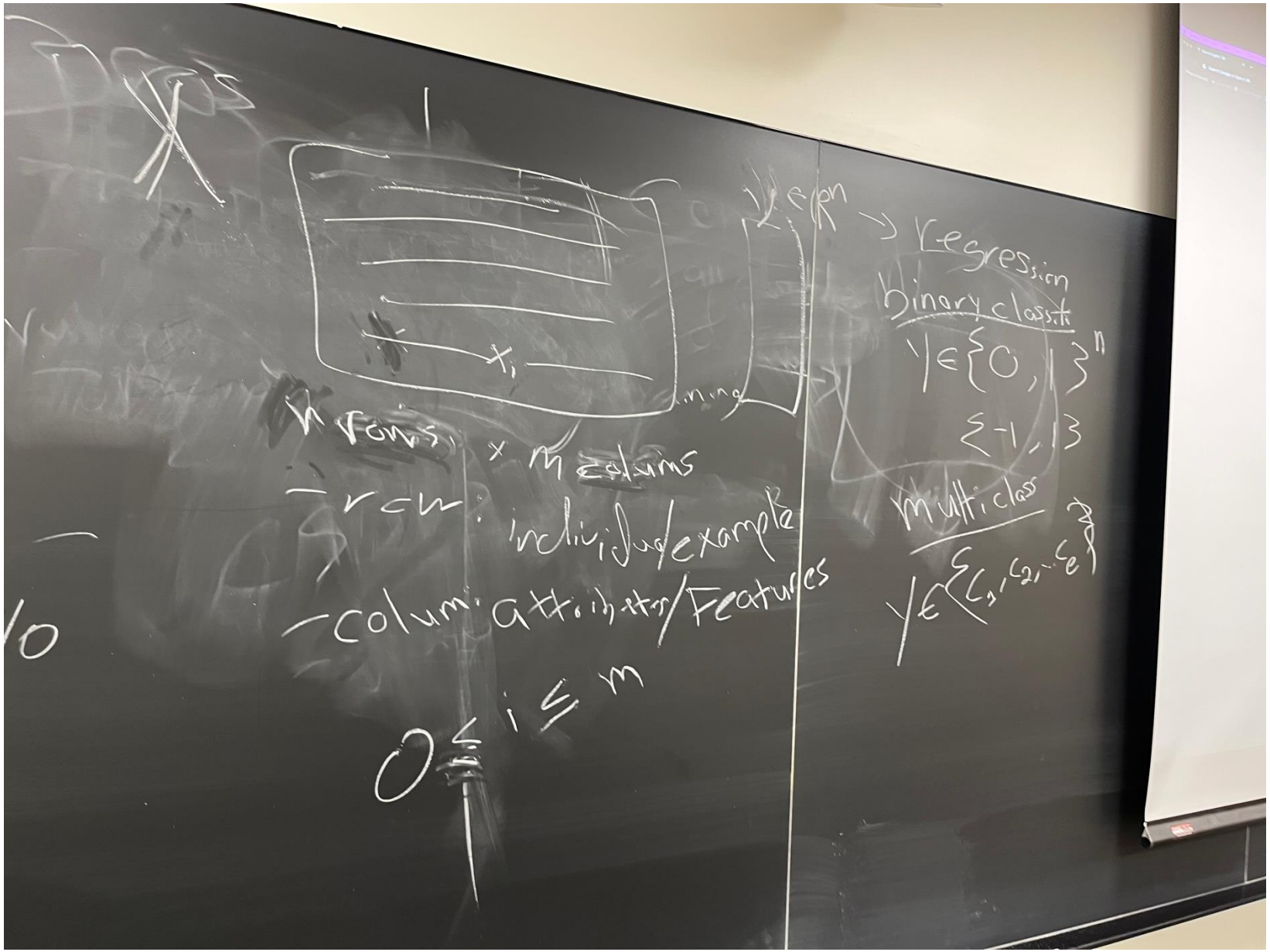
Outline

- Welcome + what is Machine Learning (ML)?
- Examples of ML
- Syllabus highlights
- ML terminology and notation
- First algorithm: K-nearest neighbors

Input/Output

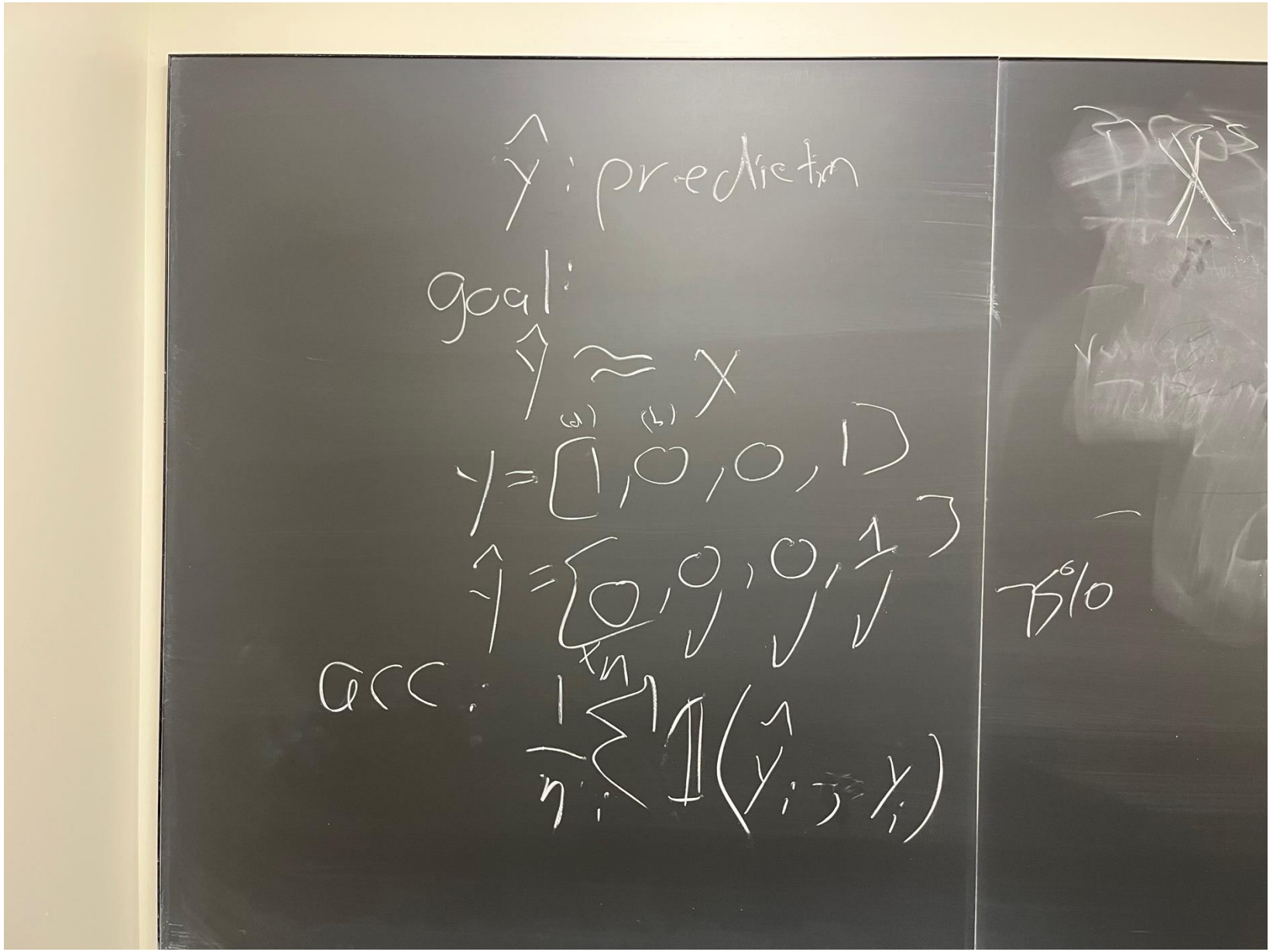
Feature matrix

Types of Predictions



Evaluation

Accuracy



K-nearest neighbors

Problem: Given a collection of labeled examples, determine the class of a new unlabeled example

Solution: assign the same class as the ***most similar*** examples

Key idea: whats ***most similar***?

Todo for Thursday

Reading: Duame **1.1-1.2** (3 pages!)

Short **reading quiz** on these pages on Thurs

Continue to 3.1-3.3 if you have time

Fill out **Handout 1** and bring back for introductions!

HW01 due Tuesday night

TODO for Thursday

Reading: Duame **1.1-1.2** (3 pages!)

Short **reading quiz** on these pages on Thurs

Continue to 3.1-3.3 if you have time

Fill out **course survey** and bring back for introductions!