ECE 4424/5424:

Machine Learning / Advanced Machine Learning

Dhruv Batra

Virginia Tech

ECE 4424/5424:
Machine Learning / Advanced Machine Learning

Stefan Lee
Virginia Tech

ECE 4424/5424: Introduction to Machine Learning

with more work and creative research expected from graduate students.

Stefan Lee
Virginia Tech

Today

- What is Machine Learning?
- Why would I want to study Machine Learning?
- How will this class operate?
- HW0 Goes Out

- “If you were a current computer science student what area would you start studying heavily?”
  - Answer: Machine Learning.
  - “The ultimate is computers that learn”
  - Bill Gates, Reddit AMA

- “Machine learning is the next Internet”
  - Tony Tether, Director, DARPA

- “Machine learning is today’s discontinuity”
  - Jerry Yang, CEO, Yahoo

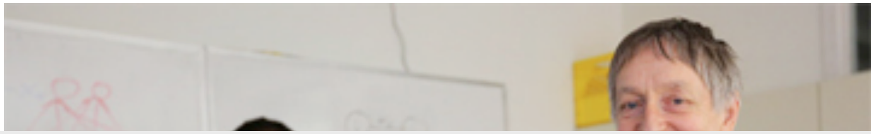
Google snaps up object recognition startup DNNr

Google has ac
Toronto, who

by Josh Lowensohn

2 / 0

Google has acqui
research compan



« Search needs a shake-up

Songbirds use grammar rules »

Machine Learning Startup Acquired by ai-one

Press Release

For Immediate Release: August 4, 2011

San Diego artificial intelligence startup acquired by leading
…ning SDKs as market for advanced

…oday that it acquired Auto-Semantics, a local start-up …es to corporate IT departments. The acquisition is the …nd acquisitions by ai-one that consolidates its …ing market for machine learning technologies.

Topic: Cloud

Follow via:

Microsoft acquires legal-focused machine-learning vendor Equivio

Summary: *Microsoft has purchased Equivio, maker of a machine-learning platform for the legal industry, for an undisclosed amount.*

By Mary Jo Foley for All About Microsoft | January 20, 2015 -- 16:24 GMT (08:24 PST)

Follow @maryjofoley 95.6K followers *Get the ZDNet Cloud newsletter now*

Microsoft has purchased Equivio, an eDiscovery/compliance vendor with a specialization in text analysis, for an undisclosed amount.

Microsoft officials announced the acquisition of the Israeli company -- its first acquisition of 2015 using more of its offshore cash -- on January 20.

Update: The Wall Street Journal reported back in October last year that Microsoft planned to buy Equivio for $200 million.

Update No. 2: A Microsoft spokesperson said the $200 million estimate was inflated and incorrect, but declined to provide a different figure.

FOUNDED
2011

OVERVIEW
DeepMind is a cutting edge artificial intelligence company. We combine the best techniques from machine learning and systems neuroscience to build powerful general-purpose learning algorithms. Founded by Demis Hassabis, Shane Legg and Mustafa Suleyman, the company is based in London and supported by some of the most iconic technology entrepreneurs and investors of the past decade. Our first commercial ...

What is ~~Machine~~ Learning?

- "the acquisition of knowledge or skills through experience, study, or by being taught."

What is Machine Learning?

- [Arthur Samuel, 1959]
  - Field of study that gives computers the ability to learn without being explicitly programmed

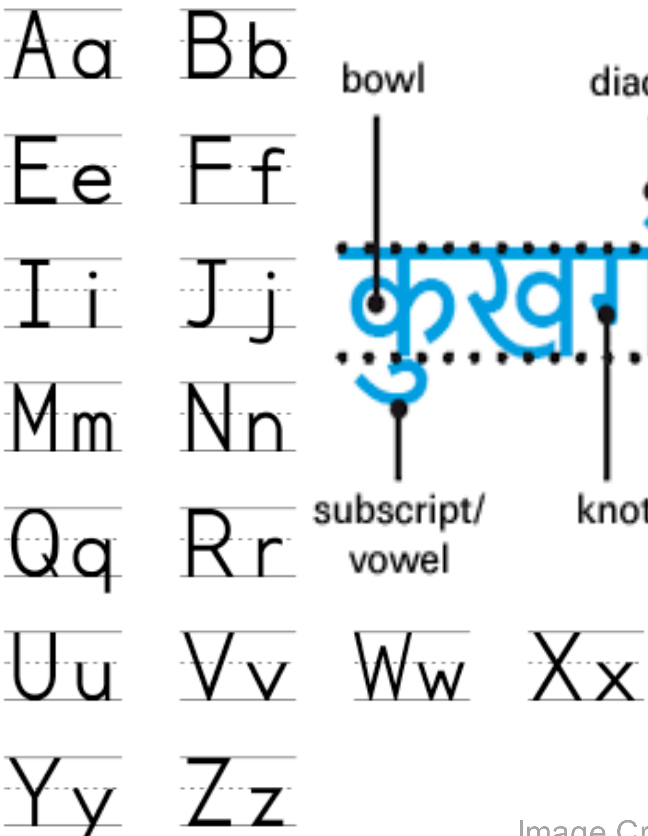
- [Kevin Murphy] algorithms that
  - automatically detect patterns in data
  - use the uncovered patterns to predict future data or other outcomes of interest

- [Tom Mitchell] algorithms that
  - improve their performance (P)
  - at some task (T)
  - with experience (E)

What is Machine Learning?

- Let’s say you want to solve Character Recognition

- Hard way: Understand handwriting/characters



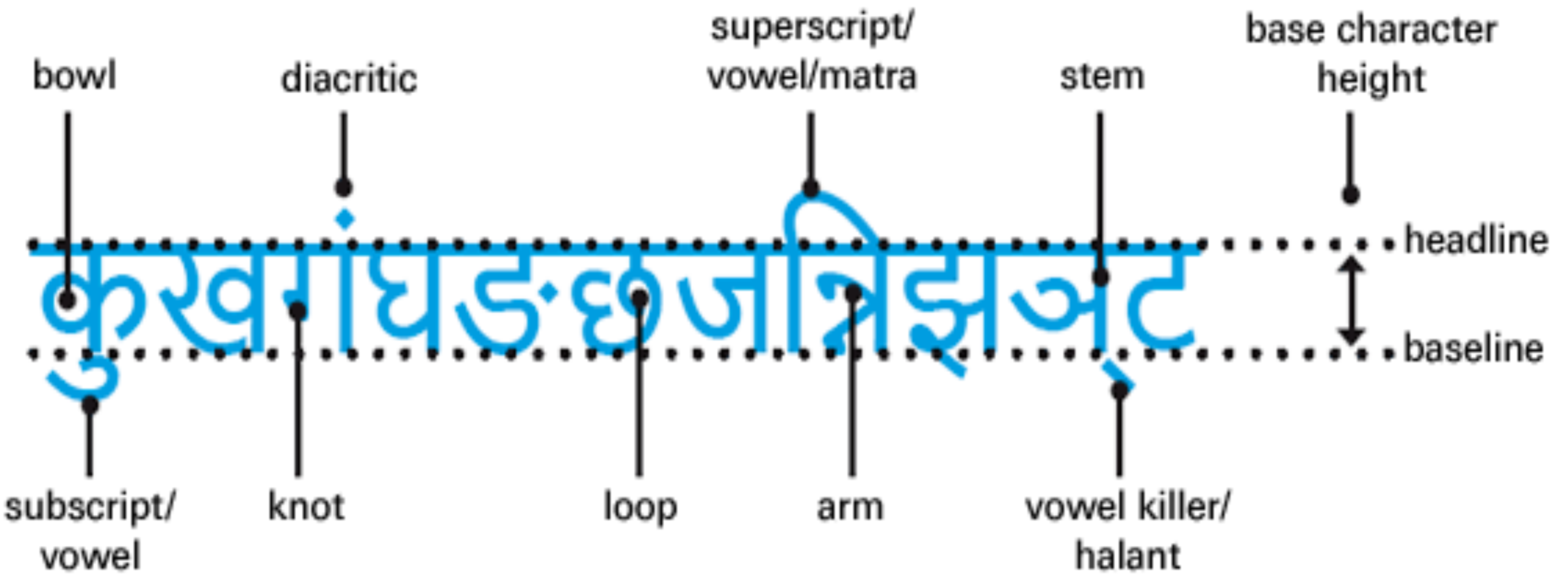
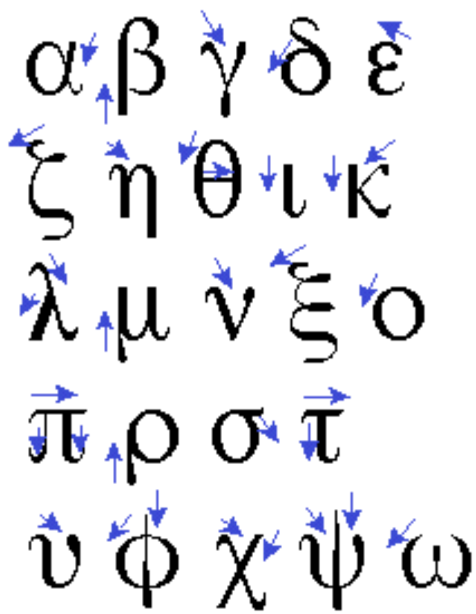


Image Credit: http://www.linotype.com/6896/devanagari.html

What is Machine Learning?

- Let's say you want to solve Character Recognition

- Hard way: Understand handwriting/characters
  - Latin
  - Devanagri
  - Symbols: http://detexify.kirelabs.org/classify.html



What is Machine Learning?

- Let’s say you want to solve Character Recognition
- Hard way: Understand handwriting/characters
- Lazy way: Throw data!

Example: Netflix Challenge

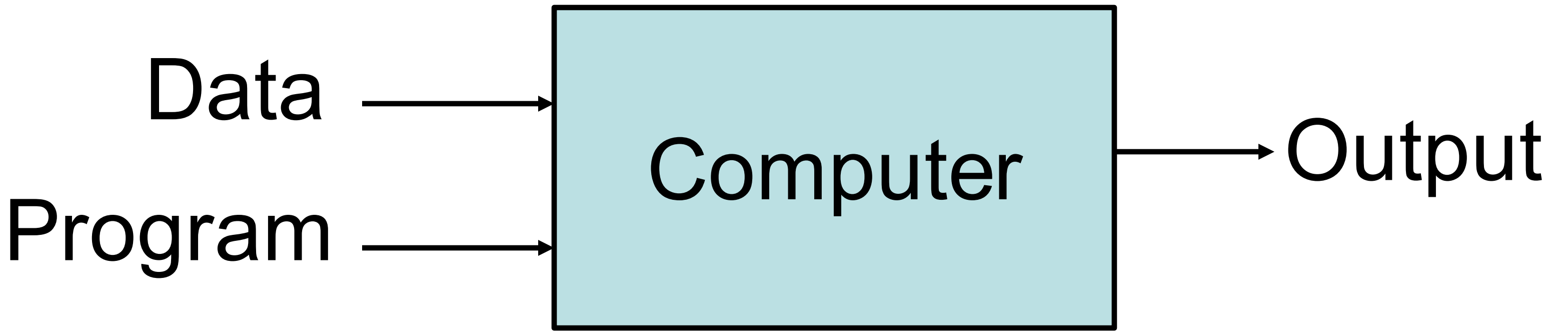
- Goal: Predict how a viewer will rate a movie
- 10% improvement = 1 million dollars

Example: Netflix Challenge

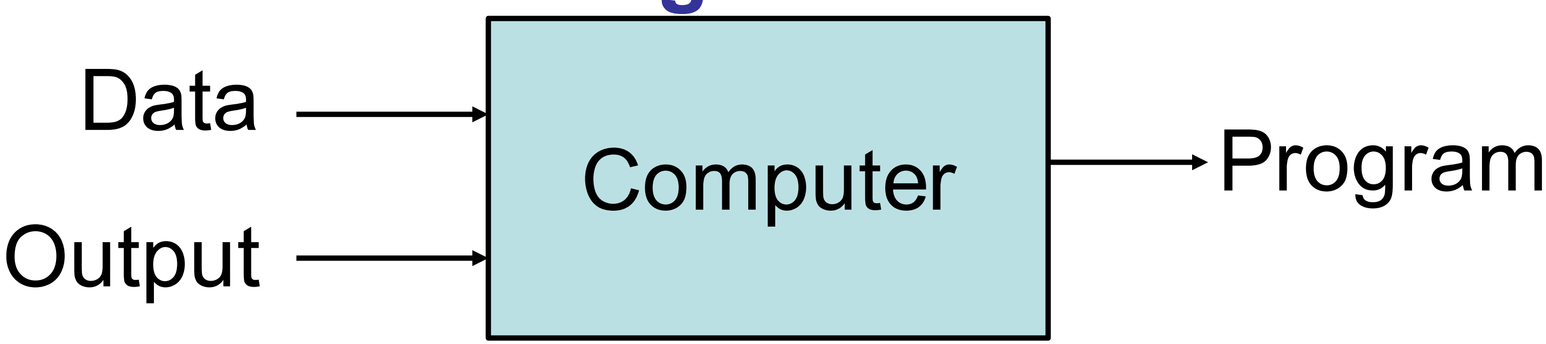
- Goal: Predict how a viewer will rate a movie
- 10% improvement = 1 million dollars
- Essence of Machine Learning:
  - A pattern exists
  - We cannot pin it down mathematically
  - We have data on it

A Convenient Oversimplification

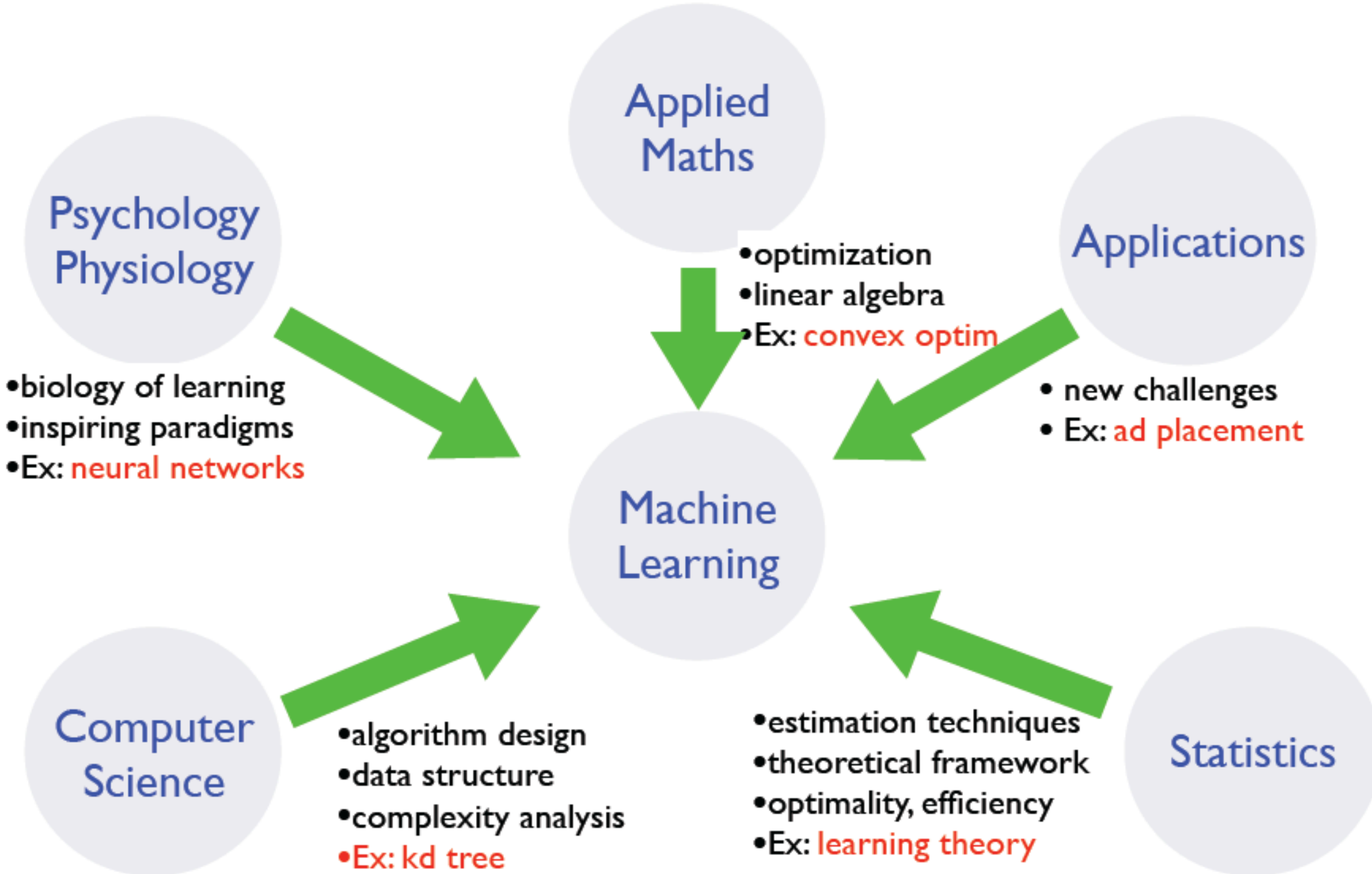
- **Traditional Programming**



- **Machine Learning**



Where does ML fit in?



Why Study Machine Learning?
Engineering Better Computing Systems

- Develop systems
  - too difficult/expensive to construct manually
  - because they require specific detailed skills/knowledge
  - ***knowledge engineering bottleneck***
- Develop systems
  - that adapt and customize themselves to individual users.
  - Personalized news or mail filter
  - ***Personalized tutoring***
- Discover new knowledge from large databases
  - Medical text mining (e.g. migraines to calcium channel blockers to magnesium)
  - ***data mining***

Slide Credit: Ray Mooney

Why Study Machine Learning?
Cognitive Science

- Computational studies of learning may help us understand learning in humans
  - and other biological organisms.
  - Hebbian neural learning
    - “Neurons that fire together, wire together.”

Slide Credit: Ray Mooney

Why Study Machine Learning?
The Time is Ripe

- Data
  - Large amounts of on-line data available.

- Computing
  - Large amounts of computational resources available.

- Algorithms
  - Many basic effective and efficient algorithms available.
    - Finally!

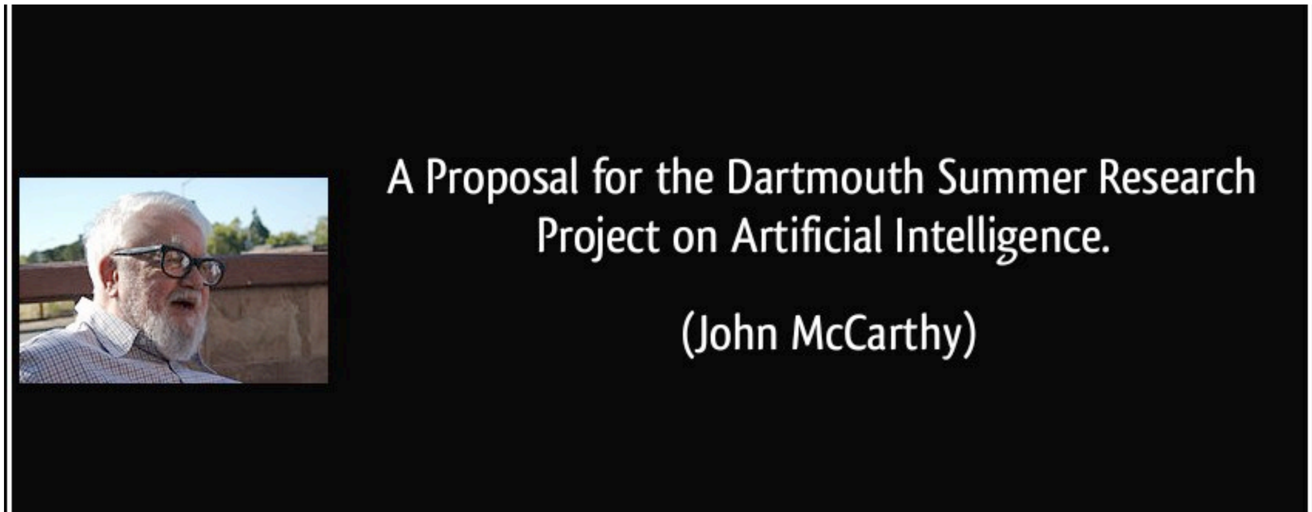
Slide Credit: Ray Mooney

Why Study Machine Learning?

- If you are a Scientist

- If you are an Engineer / Entrepreneur
  - Get lots of data
  - Machine Learning
  - ???
  - Profit!

A Brief History of AI



A Brief History of AI

- “We propose that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire.”

- We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.”

AI Predictions: Experts

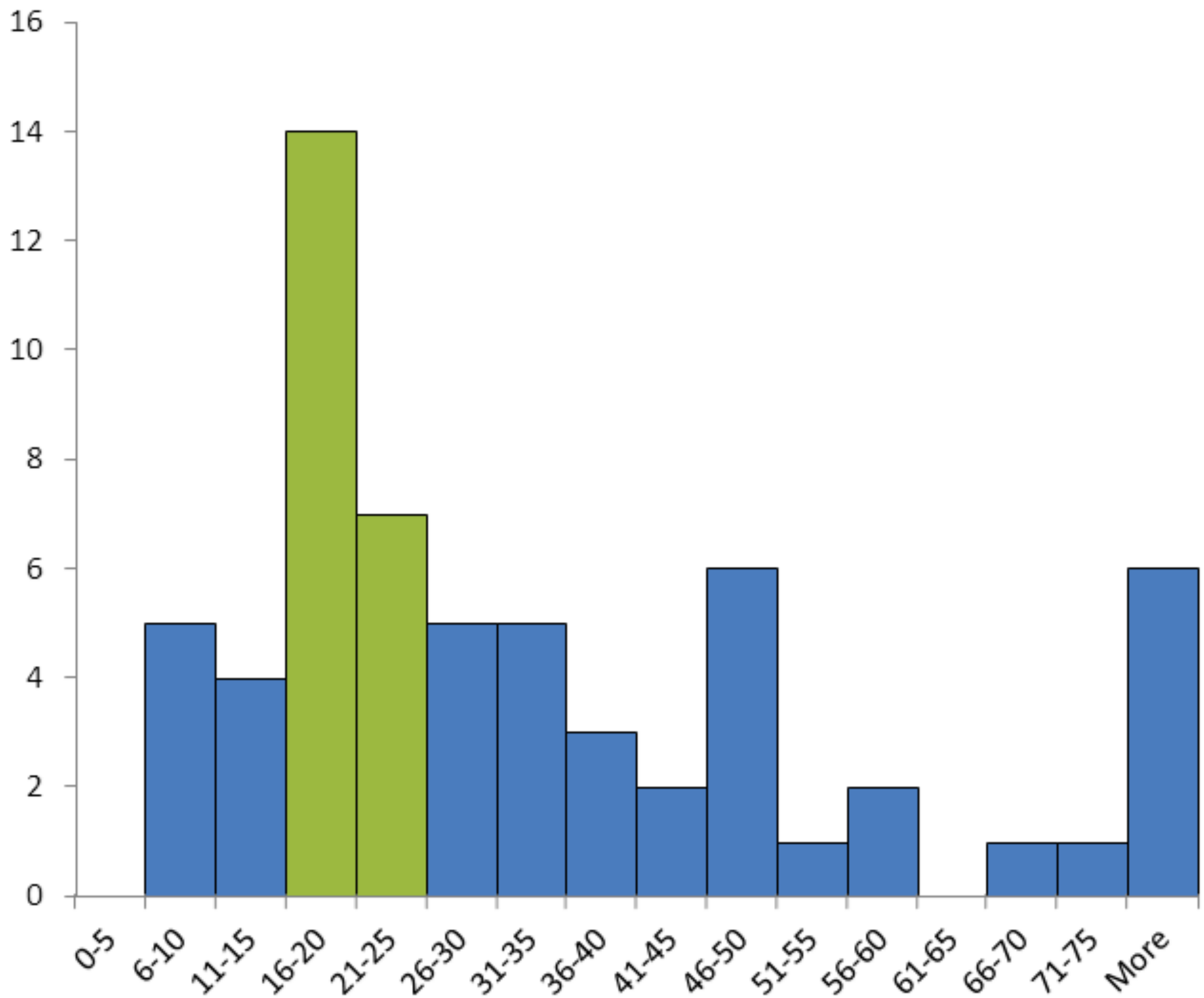


Image Credit: http://intelligence.org/files/PredictingAI.pdf

AI Predictions: Non-Experts

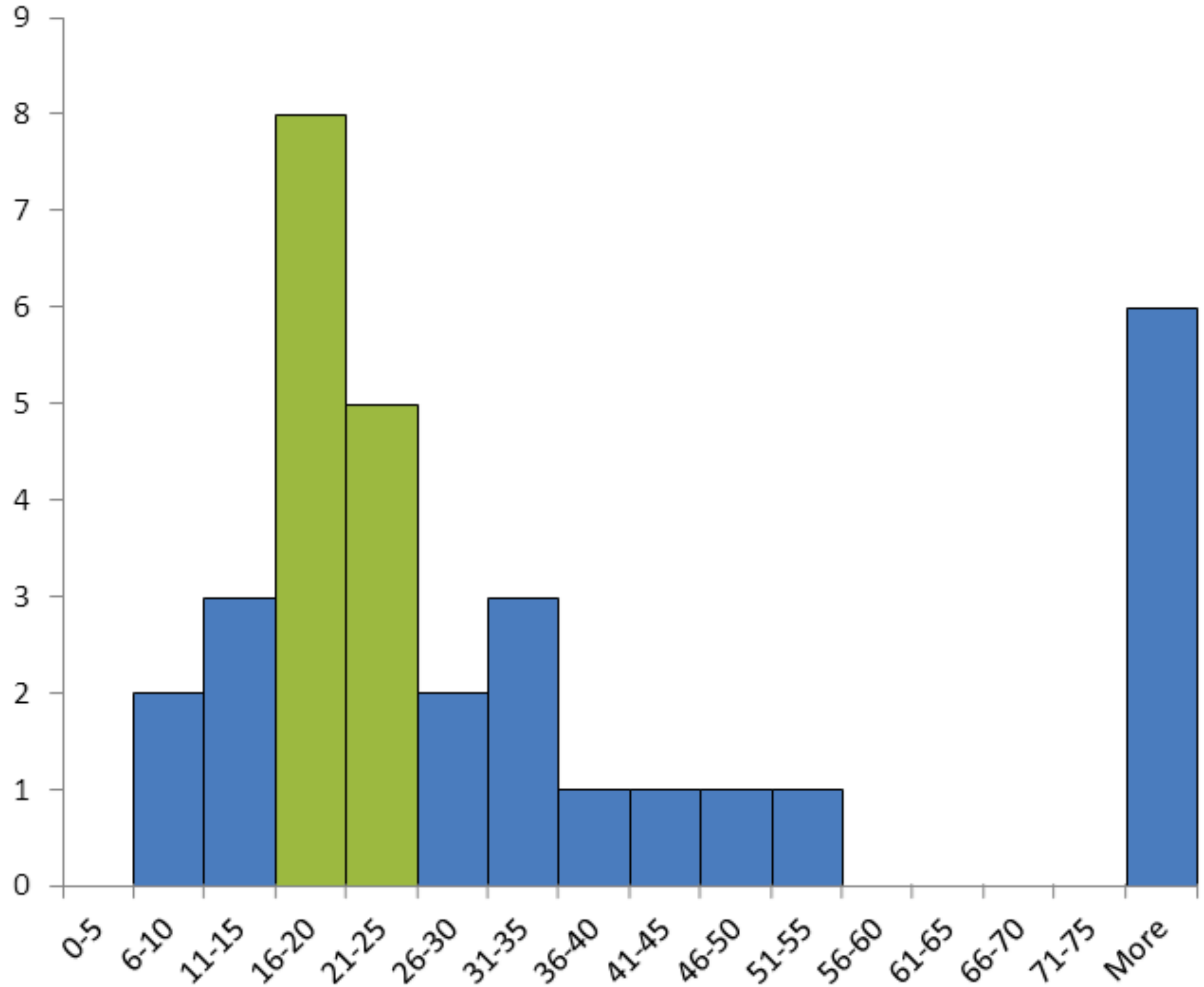


Image Credit: http://intelligence.org/files/PredictingAI.pdf

AI Predictions: Failed

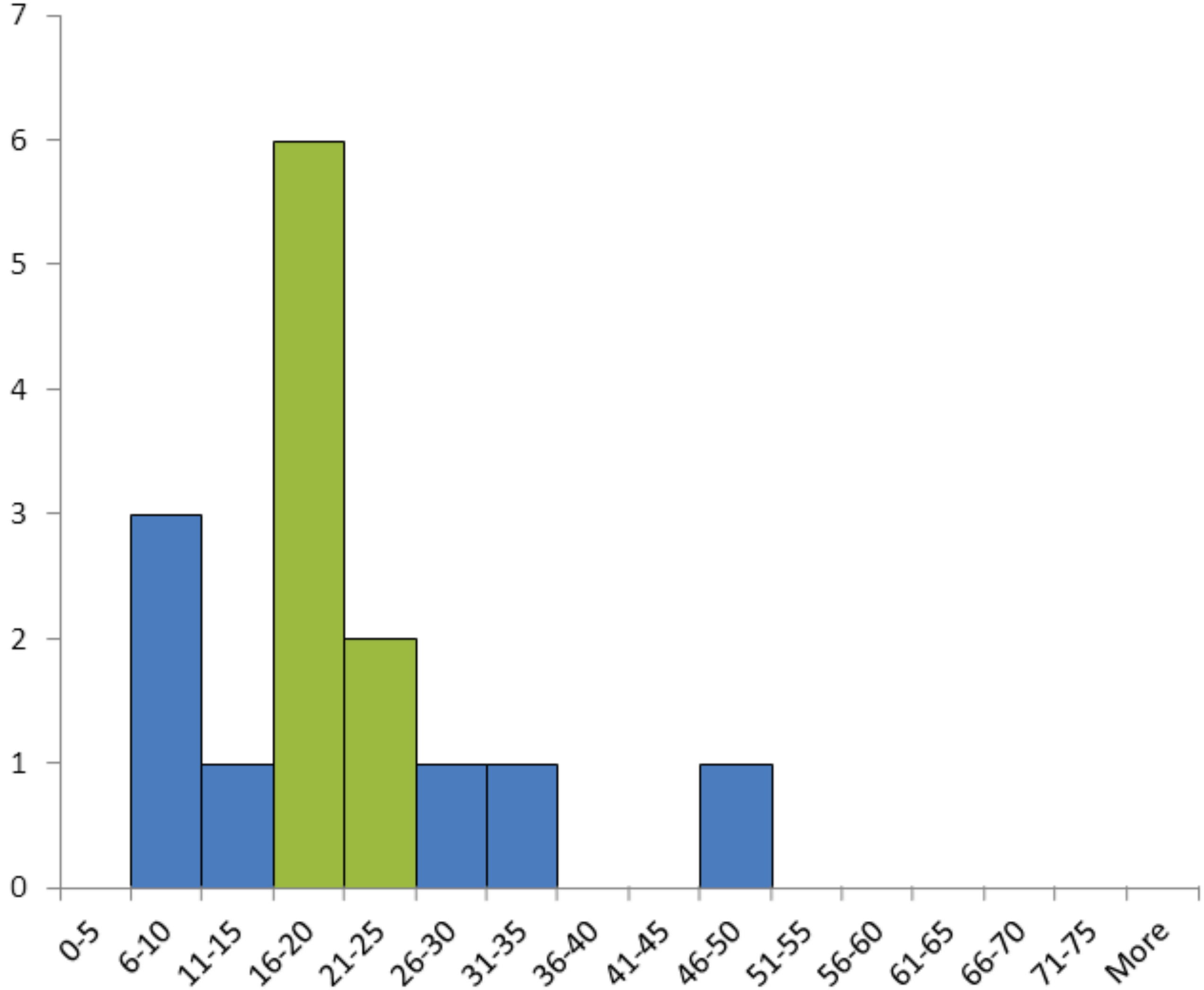
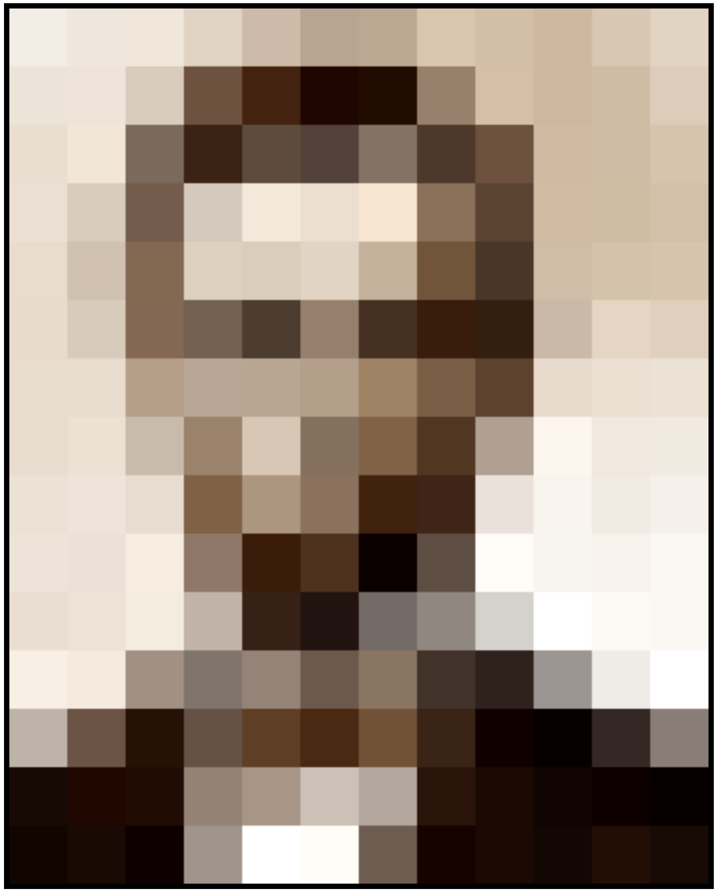


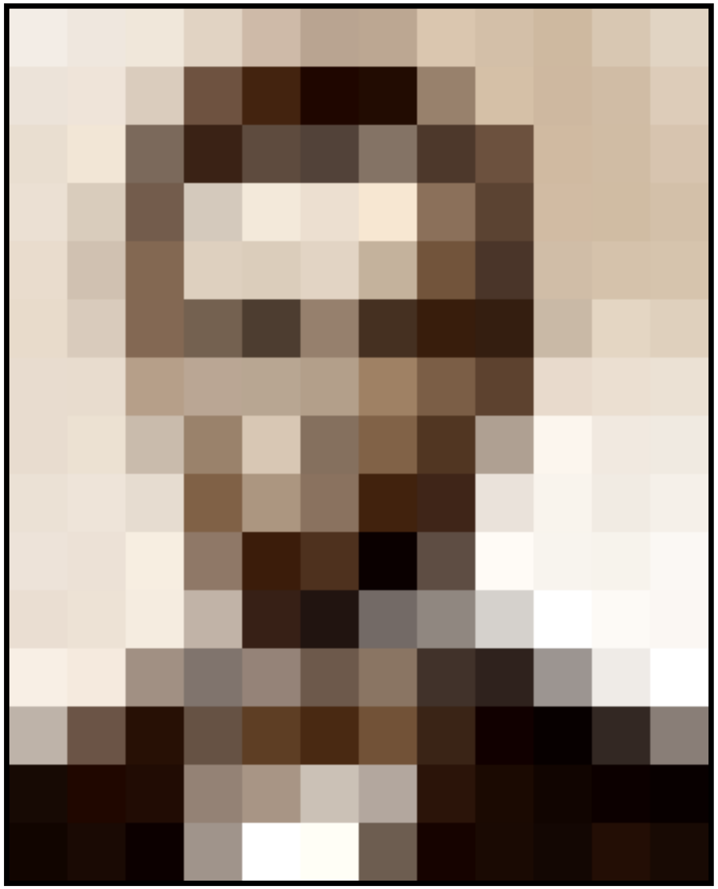
Image Credit: http://intelligence.org/files/PredictingAI.pdf

Why is AI hard?

What humans see



What computers see



“I saw her duck”

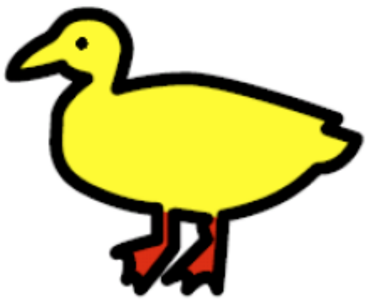
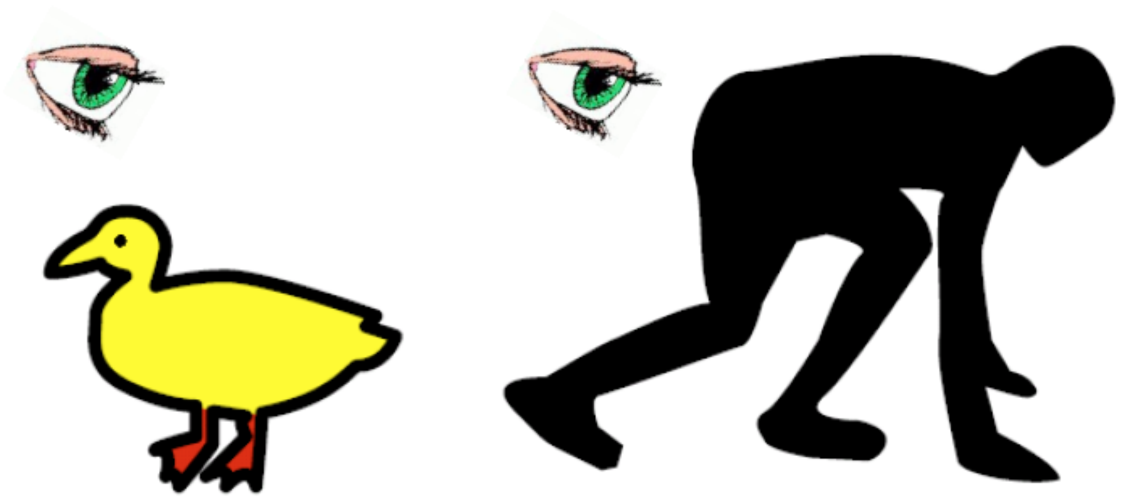


Image Credit: Liang Huang

“I saw her duck”



"I saw her duck"

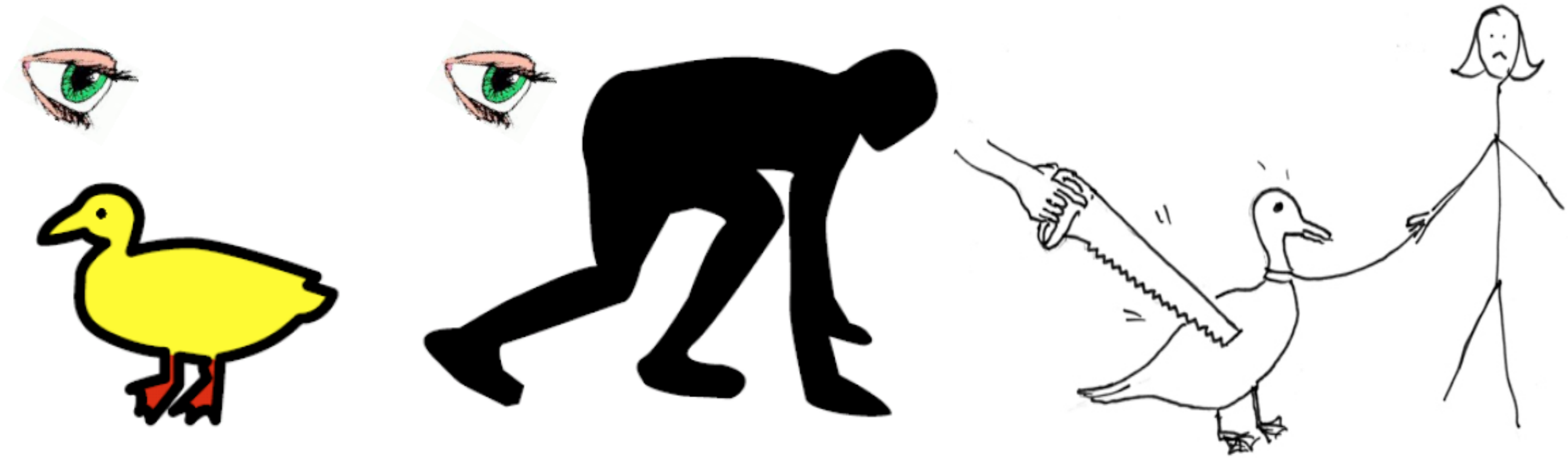


Image Credit: Liang Huang

“I saw her duck with a telescope…”

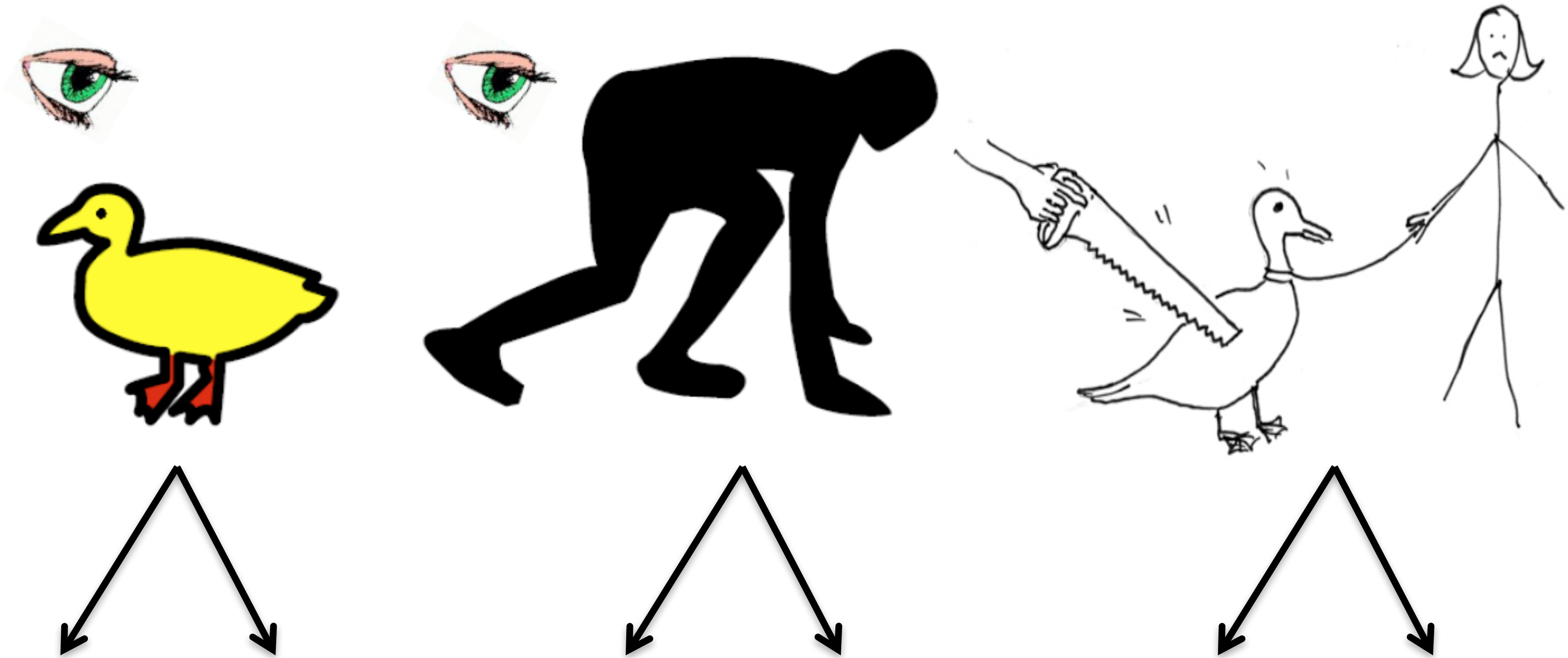


Image Credit: Liang Huang

We've come a long way…

- What is Jeopardy? Watson
  - https://youtu.be/qO1i7-Qx00k?t=72
  - https://youtu.be/_429UIzN1JM?t=36
- Alpha Go

- Future: Automated operator, doctor assistant, finance

Why are things working today?

- More compute power
- More data
- Better algorithms /models

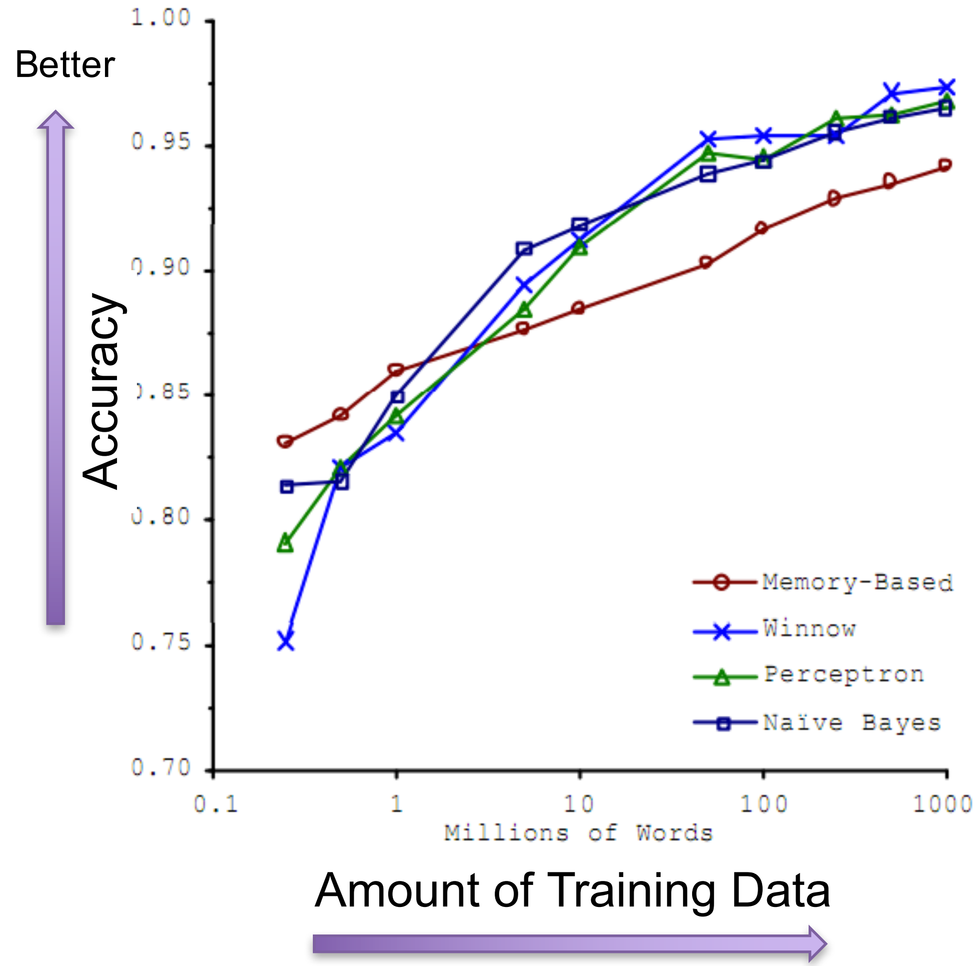


Figure Credit: Banko & Brill, 2011

ML in a (tiny)Nutshell

- Tens of thousands of machine learning algorithms
  - Hundreds new every year

- Decades of ML research oversimplified:
  - All of Machine Learning:
  - Learn a mapping from input to output f: X $\rightarrow$ Y
  - X: emails, Y: {spam, notspam}

ML in a Nutshell

- Input: x (images, text, emails…)
- Output: y (spam or non-spam…)
- (Unknown) Target Function
  - f: X → Y (the "true" mapping / reality)
- Data
  - $(x_1,y_1), (x_2,y_2), \ldots, (x_N,y_N)$
- Model / Hypothesis Class
  - g: X → Y
  - $y = g(x) = \text{sign}(w^T x)$

ML in a Nutshell

- Every machine learning algorithm has three components:
  - Representation / Model Class
  - Evaluation / Objective Function
  - Optimization

Representation / Model Class

- Decision trees
- Sets of rules / Logic programs
- Instances
- Graphical models (Bayes/Markov nets)
- Neural networks
- Support vector machines
- Model ensembles
- Etc.

Evaluation / Objective Function

- Accuracy
- Precision and recall
- Squared error
- Likelihood
- Posterior probability
- Cost / Utility
- Margin
- Entropy
- K-L divergence
- Etc.

Optimization

- Discrete/Combinatorial optimization
  - greedy search
  - Graph algorithms (cuts, flows, etc)

- Continuous optimization
  - Convex/Non-convex optimization
  - Linear programming

Types of Learning

- Supervised learning
  - Training data includes desired outputs
- Unsupervised learning
  - Training data does not include desired outputs
- Weakly or Semi-supervised learning
  - Training data includes a few desired outputs
- Reinforcement learning
  - Rewards from sequence of actions

Spam vs Regular Email

Sebring, Tracy
To: Batra, Dhruv
ECE 4424 proposal

January 21, 2015 2:53 PM
Hide Details

CUSP has approved ECE 4424 with the
copy of the proposal with these items ad
Thanks!!!
Tracy

VS

nadia bamba
To: undisclosed recipients: ;
Reply-To: nadia bamba
From Miss Nadia BamBa,

January 19, 2015 5:57 AM
Hide Details

From Miss Nadia BamBa,

Greeting, Permit me to inform you of my desire of going into business relationship with you. I am Nadia BamBa the only Daughter of late Mr and Mrs James BamBa, My father was a director of cocoa merchant in Abidjan, the economic capital of Ivory Coast before he was poisoned to death by his business associates on one of their outing to discus on a business deal. When my mother died on the 21st October 2002, my father took me very special because i am motherless.

Before the death of my father in a private hospital here in Abidjan, He secretly called me on his bedside and told me that he had a sum of $6, 8000.000(SIX Million EIGHT HUNDRED THOUSAND), Dollars) left in a suspense account in a Bank here in Abidjan, that he used my name as his first Daughter for the next of kin in deposit of the fund.

He also explained to me that it was because of this wealth and some huge amount of money That his business associates supposed to balance him from the deal they had that he was poisoned by his business associates, that I should seek for a God fearing foreign partner in a country of my choice where I will transfer this money and use it for investment purposes, (such as real estate Or Hotel management).please i am honourably seeking your assistance in the following ways.

1) To provide a Bank account where this money would be transferred to.
2) To serve as the guardian of this Money since I am a girl of 19 years old.
3)Your private phone number's and your family background' s that we can know each order more.

Moreover i am willing to offer you 15% of the total sum as compensation for effort input after the successful transfer of this fund to your designated account overseas,

Anticipating to hear from you soon.
Thanks and God Bless.
Best regards.

Intuition

- Spam Emails
  - a lot of words like
    - “money”
    - “free”
    - “bank account”
    - “viagara” … in a single email

- Regular Emails
  - word usage pattern is more spread out

Simple Strategy: Let us count!

This is X

I know you will be very much interested, kindly provide me with the details below.

First Name............
Surname................
Address..............
City.......................
State/Province.......
Country.................
Telephone No........
Occupation ……….
Date of Birth (date/m/yr) ………..
Copy of International Passport Or ID card………

$$\begin{pmatrix} \text{free} & 100 \\ \text{money} & 2 \\ \vdots & \vdots \\ \text{account} & 2 \\ \vdots & \vdots \end{pmatrix}$$

From: Ross Girshick
Subject: **Re: hey**
Date: January 17, 2013 7:48:18 PM EST
To: Dhruv Batra

Hi Dhruv,

sorry for the high latency. I just got back from Singapore last night ar

$$\begin{pmatrix} \text{free} & 1 \\ \text{money} & 1 \\ \vdots & \vdots \\ \text{account} & 2 \\ \vdots & \vdots \end{pmatrix}$$

Final Procedure

$$\begin{pmatrix} \text{free} & 100 \\ \text{money} & 2 \\ \vdots & \vdots \\ \text{account} & 2 \\ \vdots & \vdots \end{pmatrix}$$

Why these words?

$$\begin{pmatrix} 100 \times 0.2 \\ 2 \times 0.3 \\ \vdots \\ 2 \times 0.3 \\ \vdots \end{pmatrix} = 3.2$$

$$\begin{pmatrix} 100 \times 0.01 \\ 2 \times 0.02 \\ \vdots \\ 2 \times 0.01 \\ \vdots \end{pmatrix} = 1.03$$

Why linear combination?

Where do the weights come from?

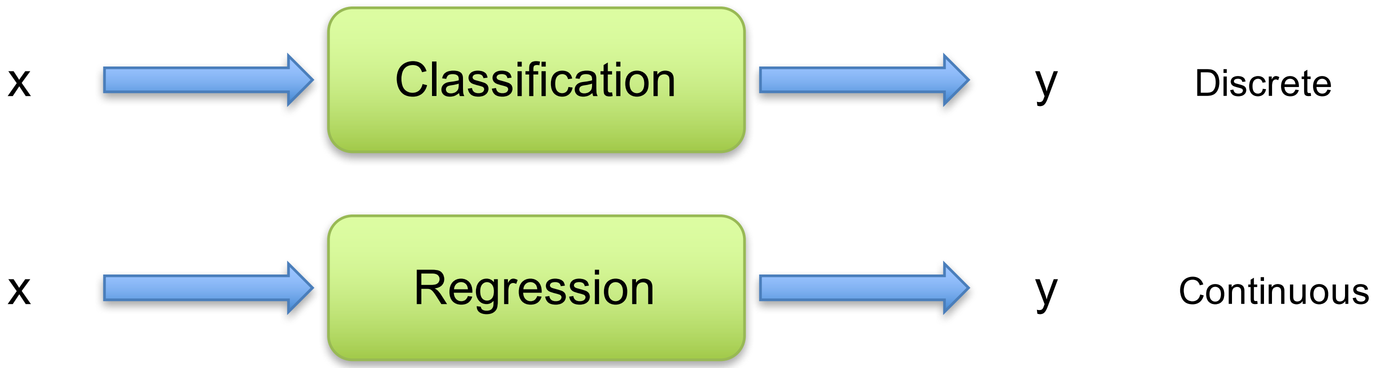
Confidence / performance guarantee?

Types of Learning

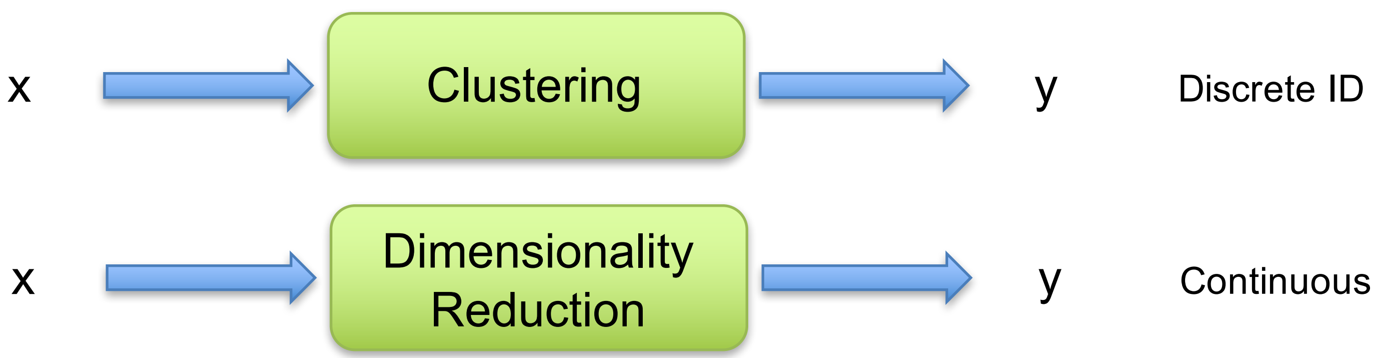
- Supervised learning
  - Training data includes desired outputs
- Unsupervised learning
  - Training data does not include desired outputs
- Weakly or Semi-supervised learning
  - Training data includes a few desired outputs
- Reinforcement learning
  - Rewards from sequence of actions

Tasks

Supervised Learning



Unsupervised Learning



Supervised Learning

Classification

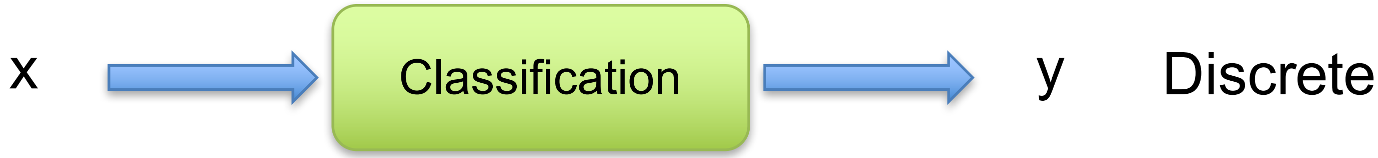


Image Classification

- Im2tags; Im2text
- http://deeplearning.cs.toronto.edu/

Pizza

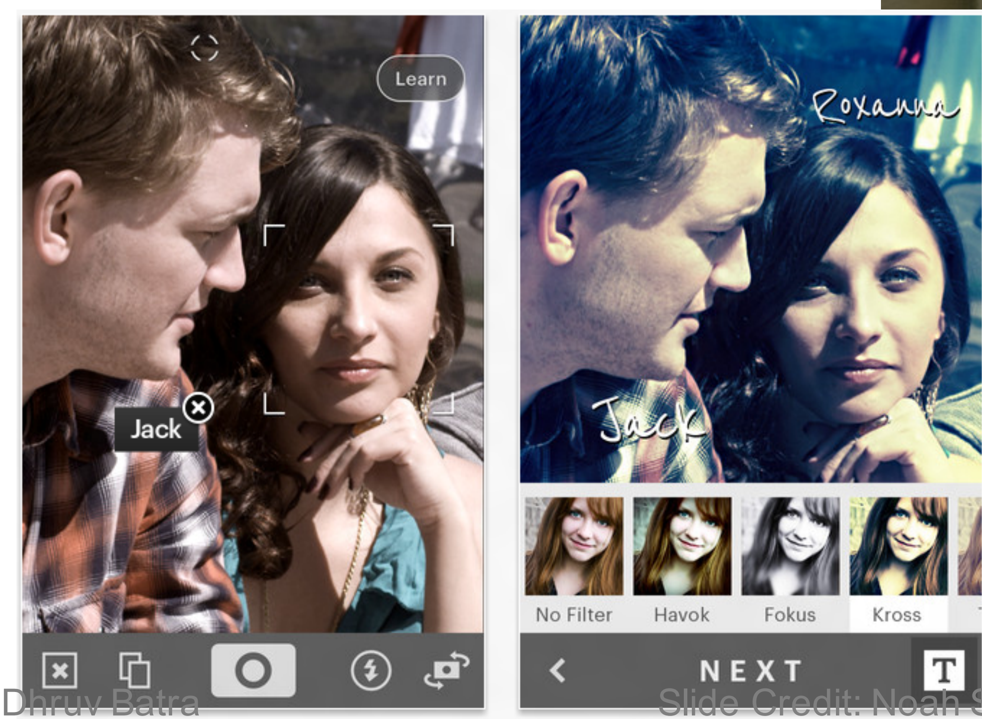
Wine

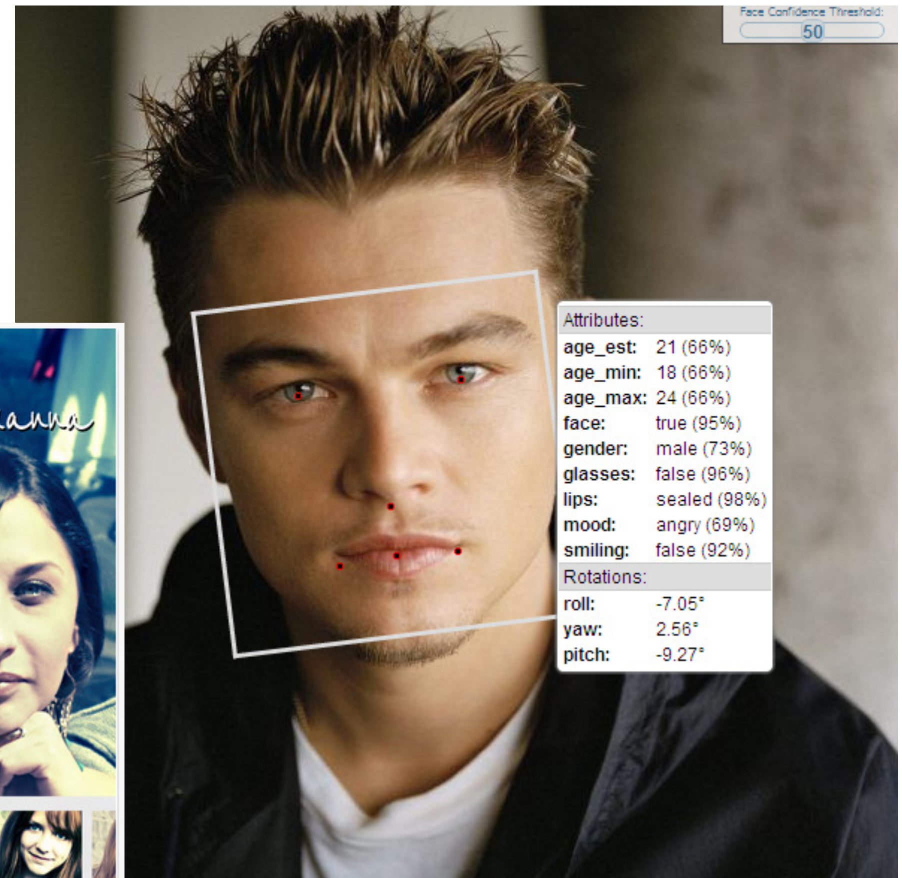
Stove

Face Recognition

face.com

http://developers.face.com/tools/





Machine Translation

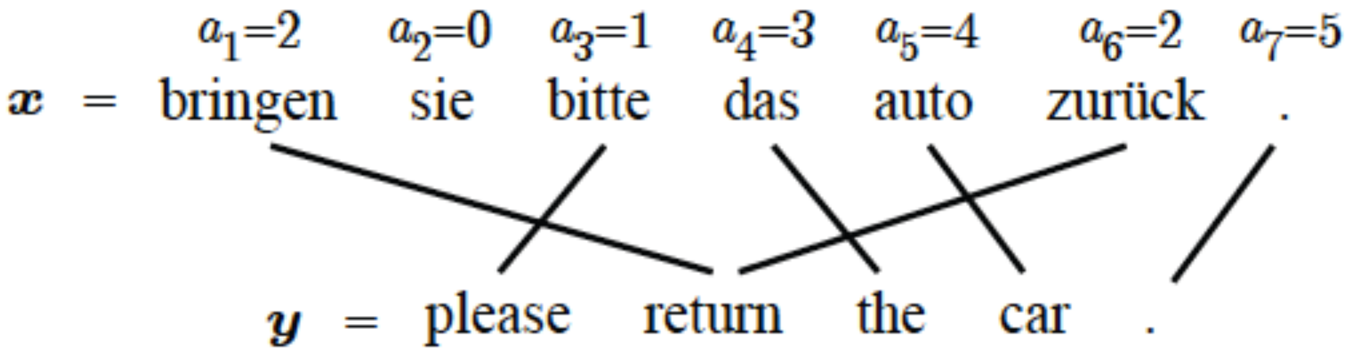
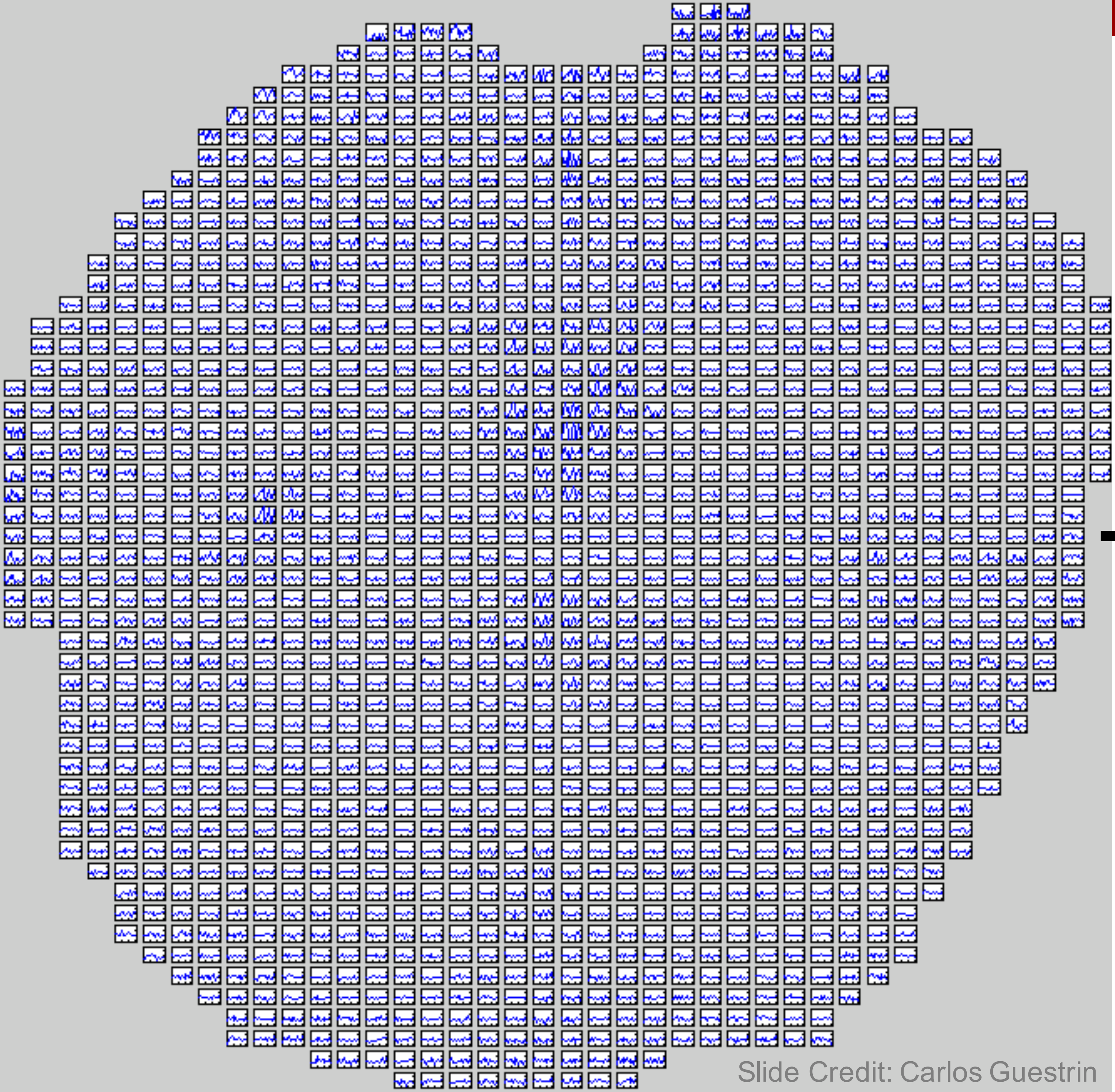


Figure Credit: Kevin Gimpel

Speech Recognition

Speech Recognition

- Rick Rashid speaks Mandarin
  - http://youtu.be/Nu-nlQqFCKg?t=7m30s



Reading a noun (vs verb)

[Rustandi et al., 2005]

Slide Credit: Carlos Guestrin

Seeing is worse than believing

- [Barbu et al. ECCV14]

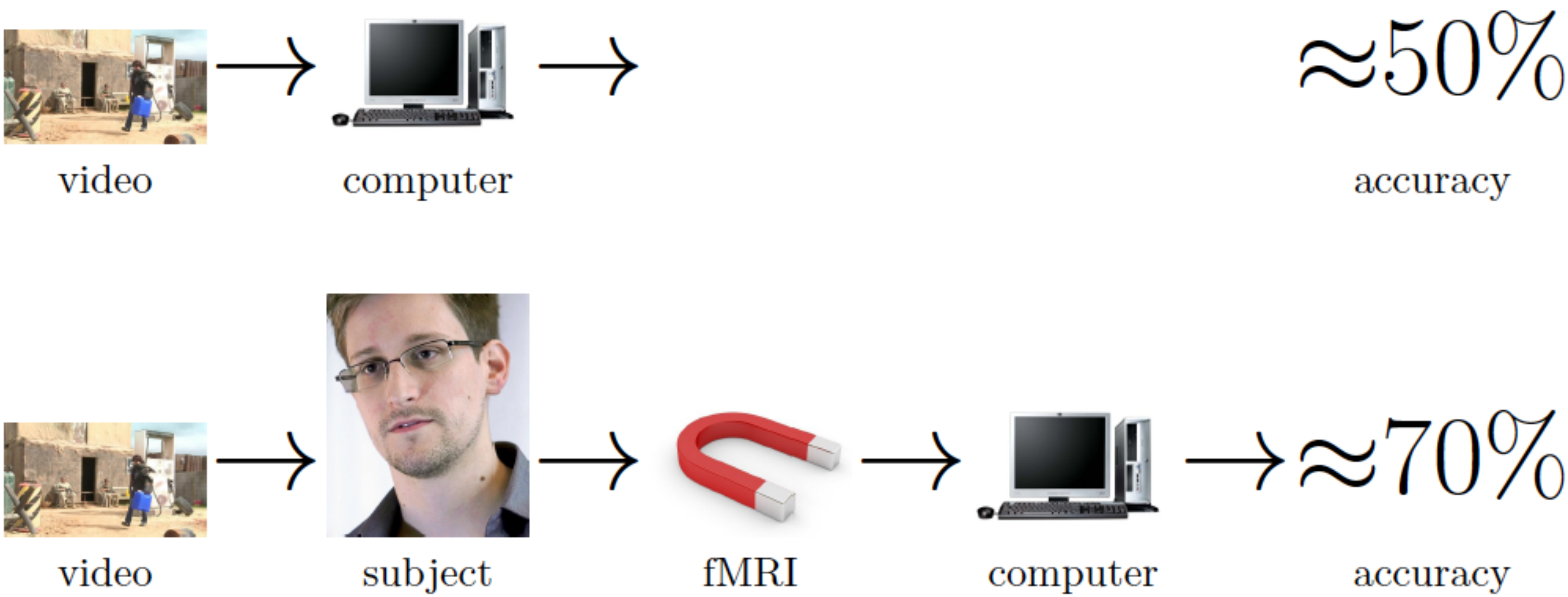
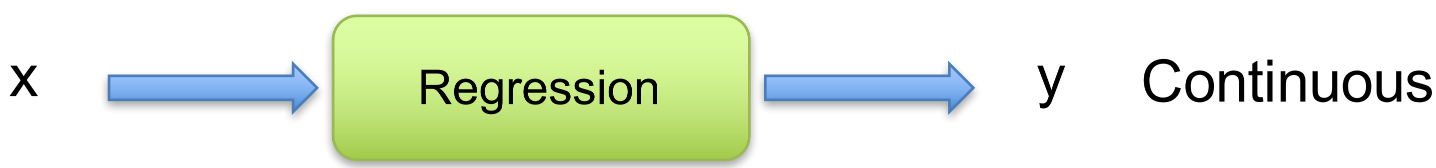


Image Credit: Barbu et al.

Supervised Learning

Regression



Stock market

Weather prediction

Temperature

27°C

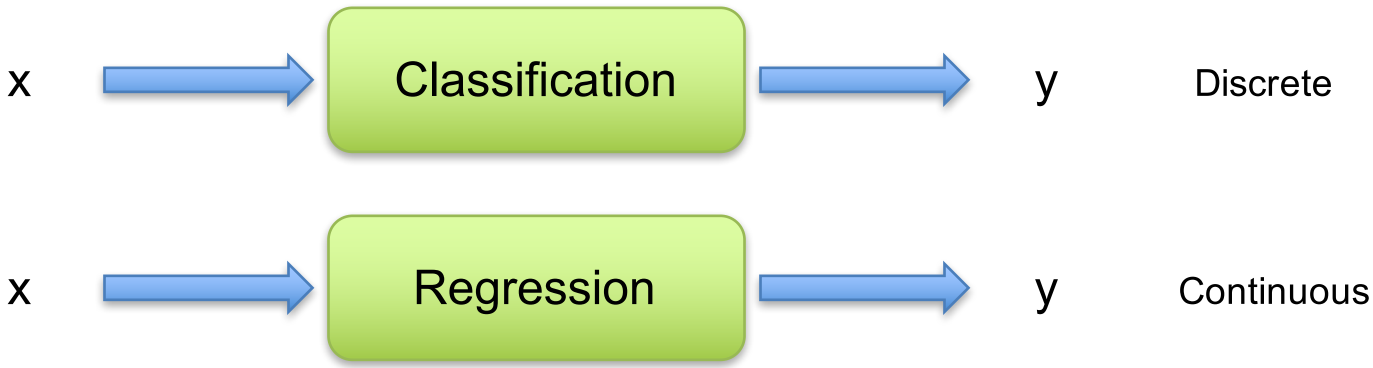
Pose Estimation

Pose Estimation

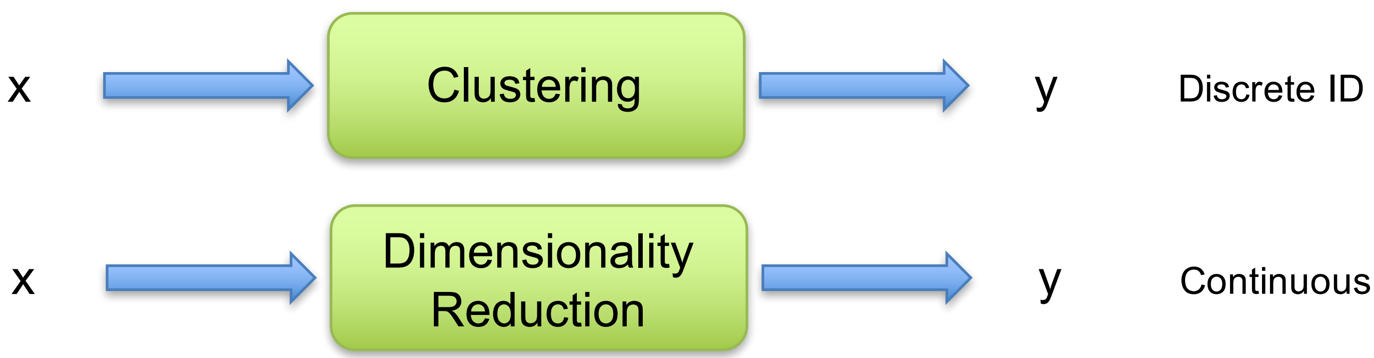
- 2010: (Project Natal) Kinect
  - http://www.youtube.com/watch?v=r5-zZDSsgFg
- 2012: Kinect One
  - http://youtu.be/Hi5kMNfgDS4?t=28s
- 2013: Leap Motion
  - http://youtu.be/gby6hGZb3ww

Tasks

Supervised Learning



Unsupervised Learning



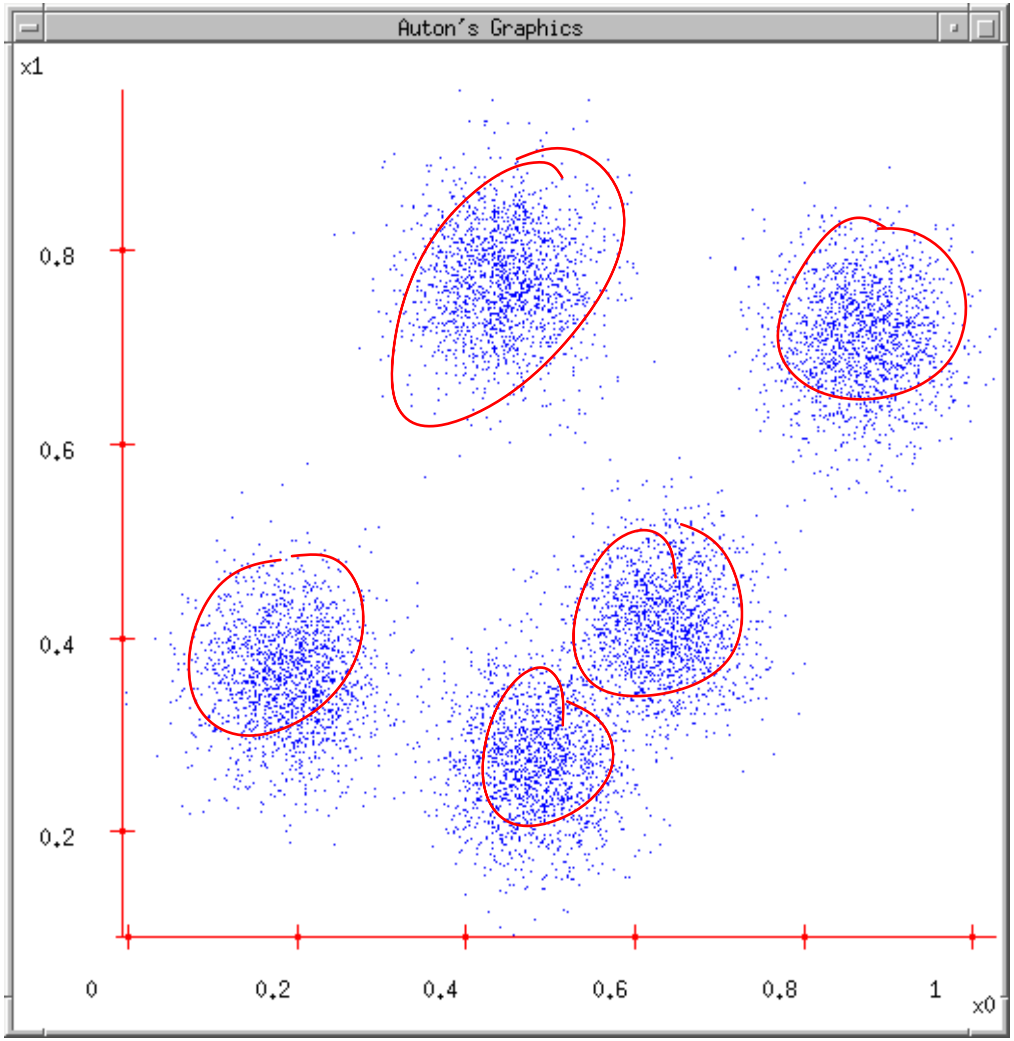
Unsupervised Learning

Clustering

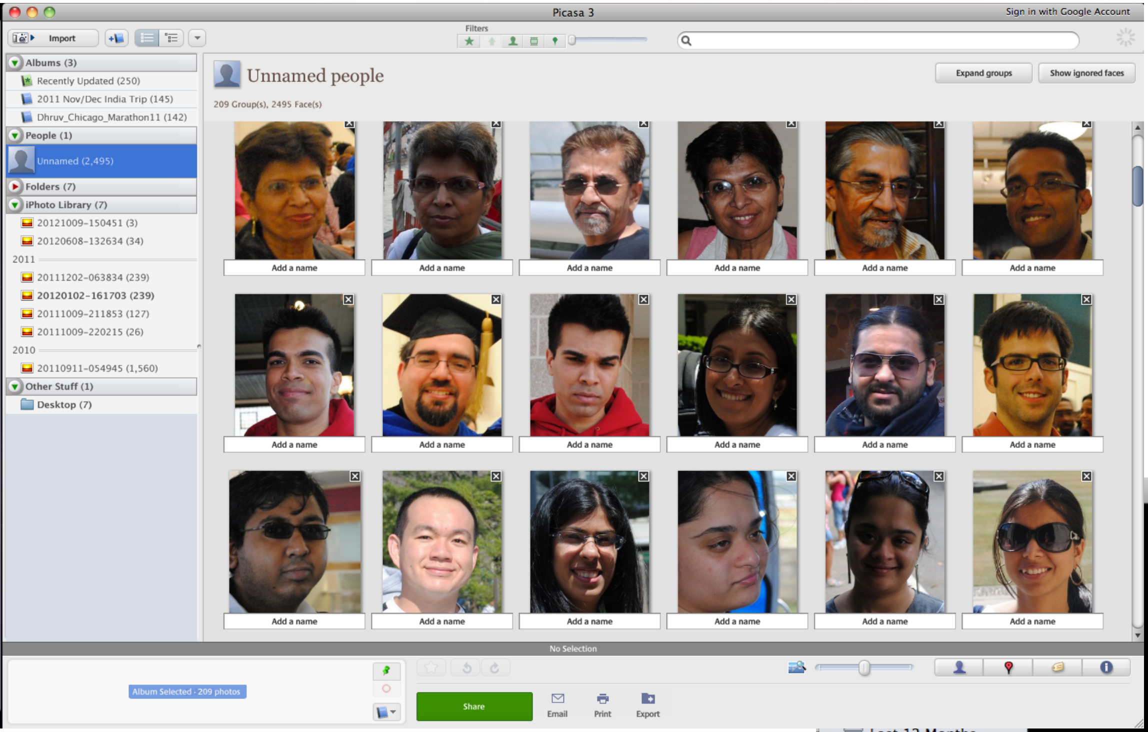
Unsupervised Learning

Y not provided

Clustering Data: Group similar things



Face Clustering



Picassa

iPhoto

Embedding

Visualizing x

Unsupervised Learning

Dimensionality Reduction / Embedding

x → Clustering → y Continuous

Unsupervised Learning

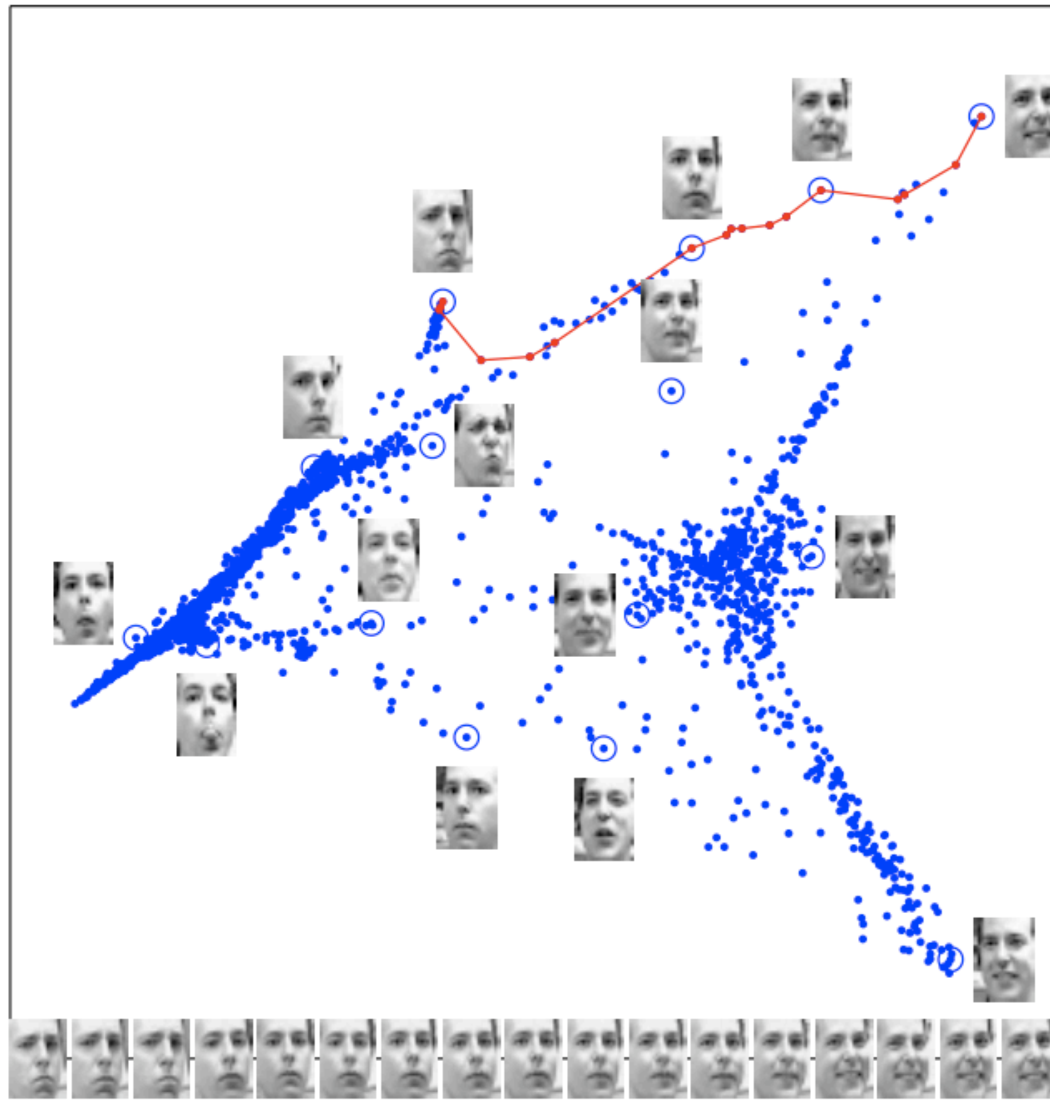
Y not provided

Embedding images

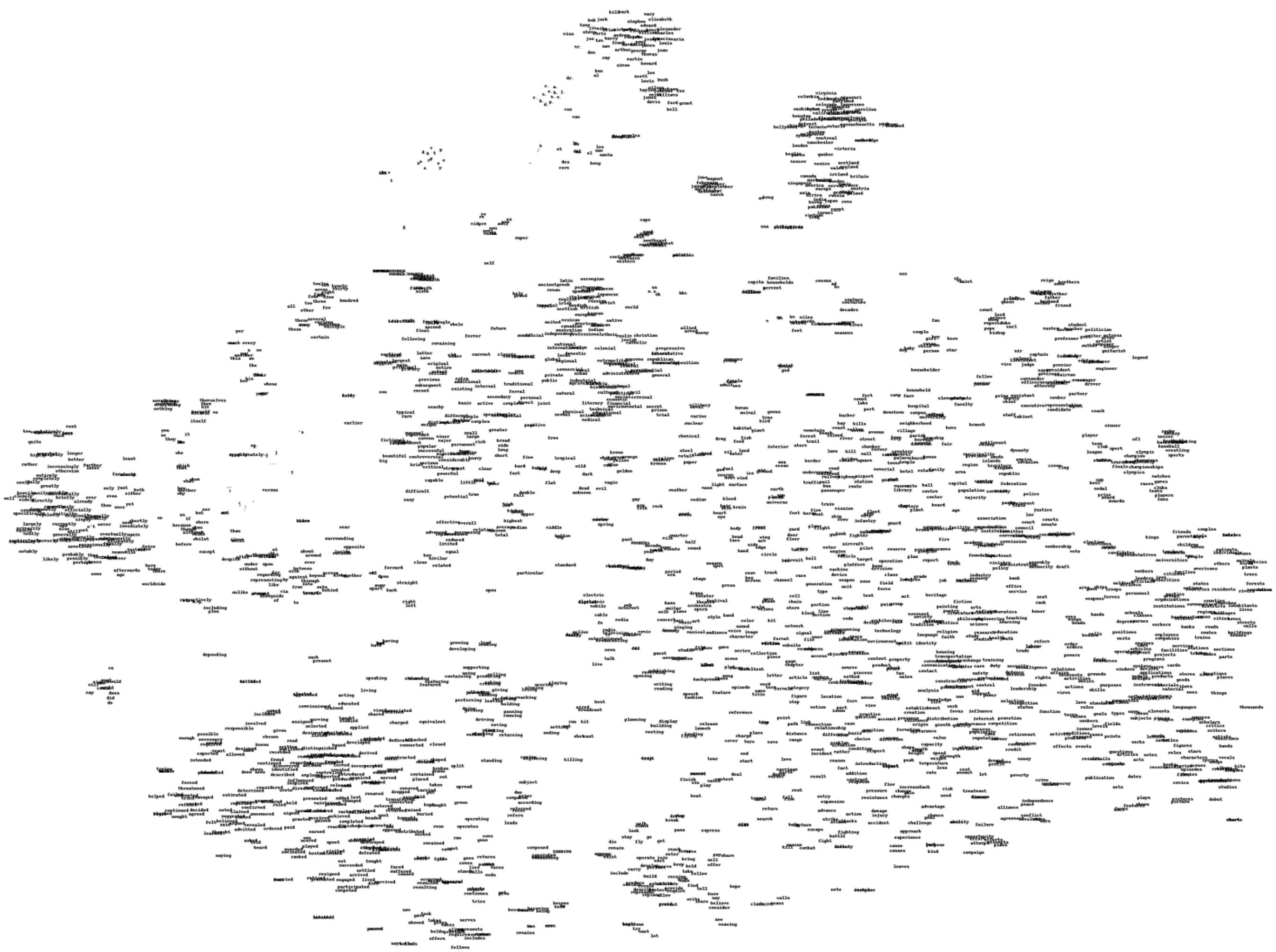
Images have thousands or millions of pixels.

Can we give each image a coordinate,

such that similar images are near each other?



Embedding words



ThisPlusThat.me (now closed)

Image Credit: http://insightdatascience.com/blog/thisplusthat_a_search_engine_that_lets_you_add_words_as_vectors.html

ThisPlusThat.me (now closed)

Image Credit: http://insightdatascience.com/blog/thisplusthat_a_search_engine_that_lets_you_add_words_as_vectors.html

Reinforcement Learning

x → Reinforcement Learning → y   Actions

Learning from feedback

Reinforcement Learning: Learning to act

- There is only one “supervised” signal at the end of the game.
- But you need to make a move at every step
- RL deals with “credit assignment”

Learning to act

- Reinforcement learning
- An agent
  - Makes sensor observations
  - Must select action
  - Receives rewards
    - positive for “good” states
    - negative for “bad” states

- Towel Folding
  - http://youtu.be/gy5g33S0Gzo

Course Information

- Instructor: Stefan Lee
  - steflee@vt.edu
  - Office Hours: Fridays 3-5PM
  - Location: Whittemore 468

- TA: Aroma Mahendru
  - Office Hours: TBA

Syllabus

- Basics of Statistical Learning
  - Loss functions, MLE, MAP, Bayesian estimation, bias-variance tradeoff, overfitting, regularization, cross-validation
- Supervised Learning
  - Nearest Neighbour, Naïve Bayes, Logistic Regression, Support Vector Machines, Kernels, Neural Networks, Decision Trees
  - Ensemble Methods: Bagging, Boosting
- Unsupervised Learning
  - Clustering: k-means, Gaussian mixture models, EM
  - Dimensionality reduction: PCA, SVD, LDA
- Advanced Topics
  - Weakly-supervised and semi-supervised learning
  - Reinforcement learning
  - Deep Neural Networks (CNNS, RNNS)
  - Probabilistic Graphical Models: Bayes Nets, HMM
  - Applications to Vision, Natural Language Processing

Syllabus

- You will learn about the methods you heard about
- But we are not teaching “how to use a toolbox”
- You will understand algorithms, theory, applications, and implementations
- **It’s going to be FUN and HARD WORK ☺**

Prerequisites

- Probability and Statistics
  - Distributions, densities, Moments, typical distributions
- Calculus and Linear Algebra
  - Matrix multiplication, eigenvalues, positive semi-definiteness, multivariate derivates…
- Algorithms
  - Dynamic programming, basic data structures, complexity (NP-hardness)…
- Programming
  - Matlab for HWs. Your language of choice for project.
  - NO CODING / COMPILATION SUPPORT
- **Ability to deal with abstract mathematical concepts**
- We provide some background, but the class will be fast paced

Textbook

- No required book.
  - We will assign readings from online/free books, papers, etc

- Reference Books:
  - [On Library Reserve]
    Machine Learning: A Probabilistic Perspective
    Kevin Murphy

  - [Free PDF from author's webpage]
    Bayesian reasoning and machine learning
    David Barber

  - Pattern Recognition and Machine Learning
    Chris Bishop

Grading

- 4 Homeworks (40%)
  - First one goes out Sept 1st
    - Start early, Start early, Start early, Start early, Start early, Start early, Start early, Start early, Start early, Start early

- Final Project (25%)
  - Proposals due Sept 21st. More details to follow.
  - Projects done individually, or groups of two students

- Midterm (10%)
  - Date 10/6th tentative date, in class

- Final (20%)
  - TBD

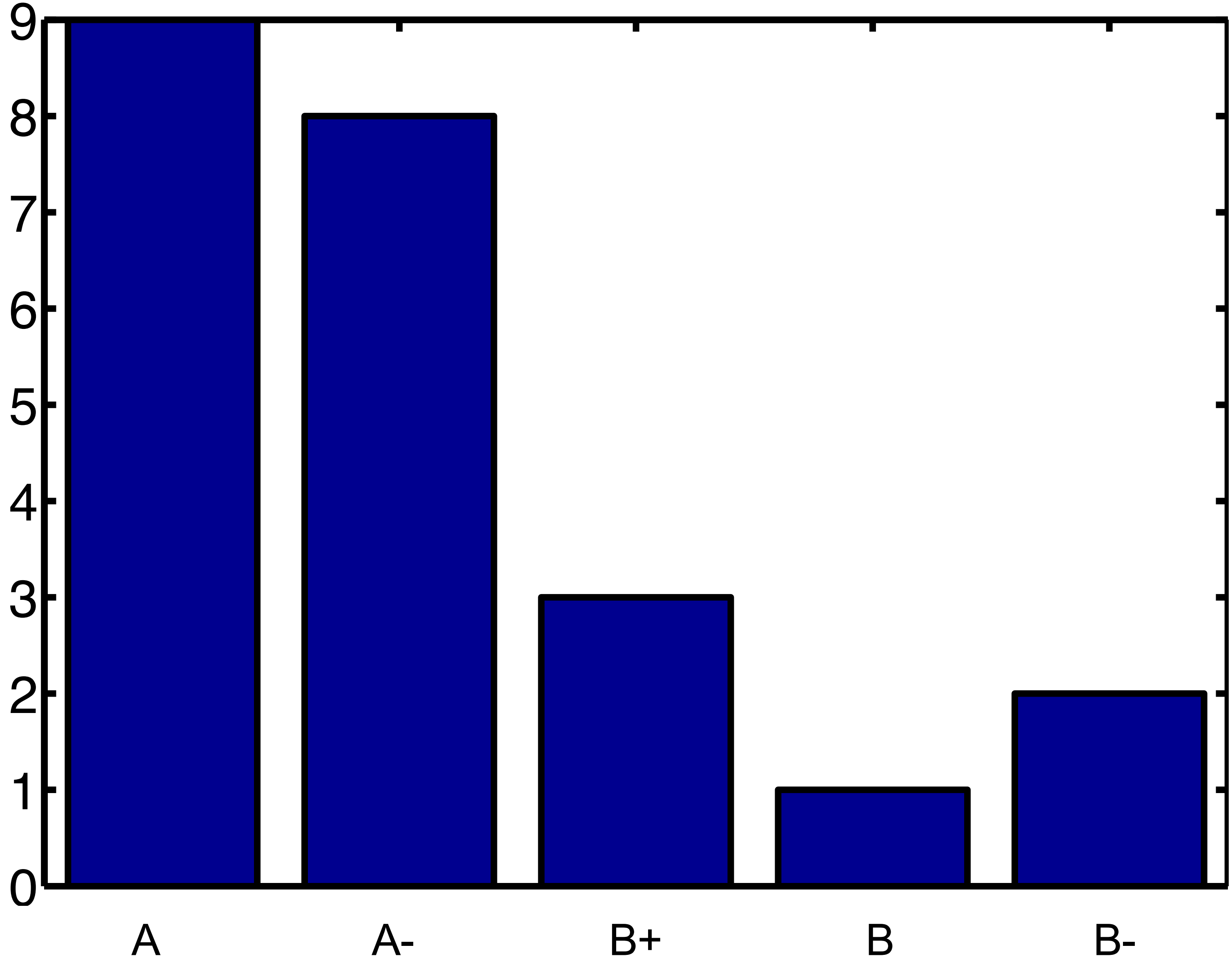
- Class Participation (5%)
  - Contribute to class discussions on Scholar
  - Ask questions, answer questions

Re-grading Policy

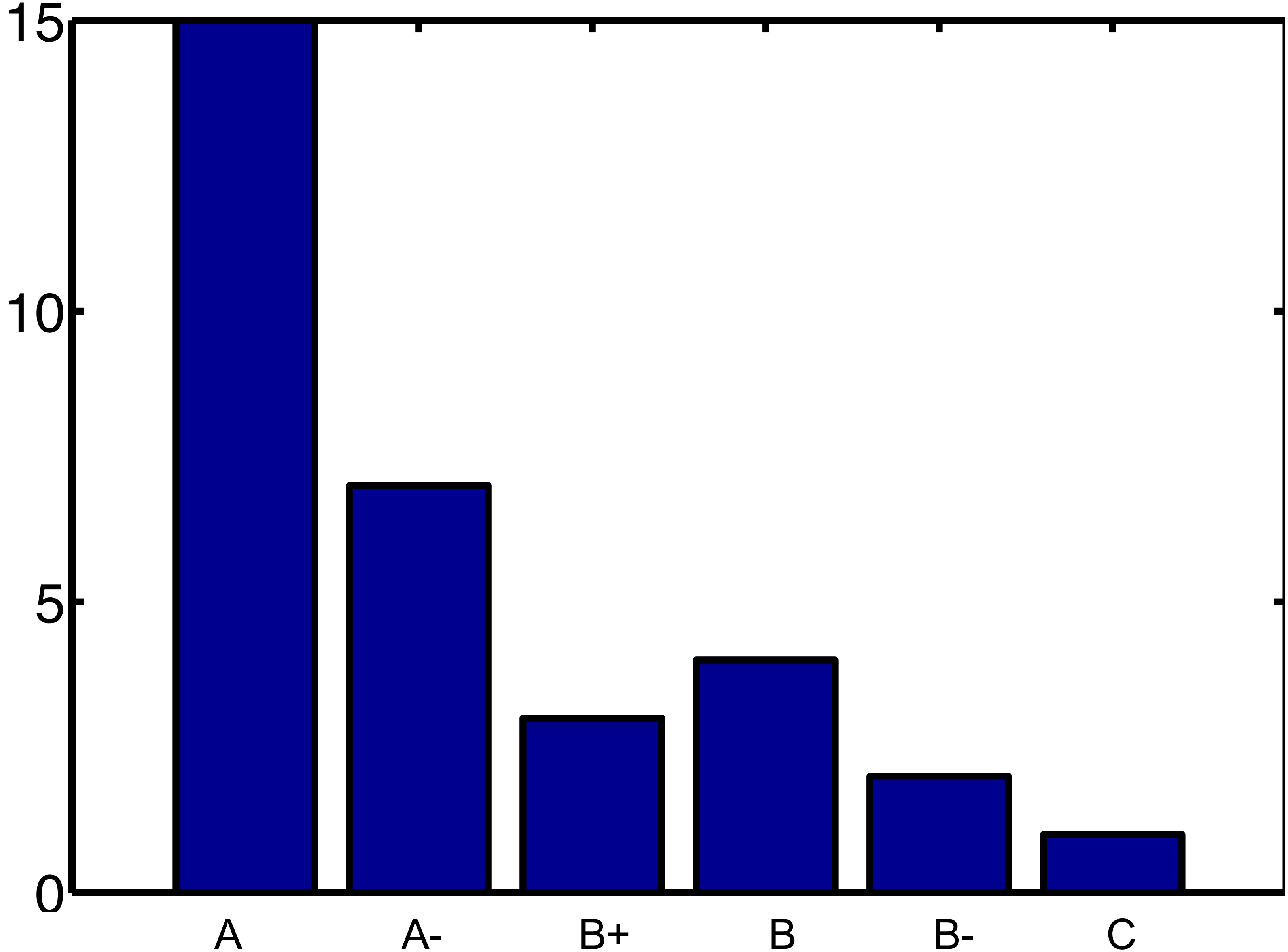
- Homework assignments and midterm
  - **Within 1 week** of receiving grades: see me
  - No change after that.

- Reasons are not accepted for re-grading
  - I cannot graduate if my GPA is low or if I fail this class.
  - I need to upgrade my grade to maintain/boost my GPA.
  - This is the last course I have to take before I graduate.
  - I have a deadline before the homework/project/midterm.
  - I have done well in other courses / I am a great programmer/theoretician

Spring 2013 Grades



Fall 2013 Grades



Homeworks

- Homeworks are difficult, start early!
  - Due in ~2 weeks via Scholar (Assignments tool)
  - Theory + Implementation
  - Kaggle Competitions:

- "Free" Late Days
  - 5 late days for the semester
    - Use for HW, project proposal/report
    - Cannot use for HW0, midterm or final exam, or poster session

  - After free late days are used up:
    - 25% penalty for each late day

HW0

- Out today; due Thursday (8/25) by 11:55pm
  - Available on Scholar now

- Grading
  - Does not count towards grade.
  - Will be graded Pass/Fail.
  - <=75% means that you might not be prepared for the class

- Topics
  - Probability
  - Linear Algebra
  - Calculus
  - Ability to Prove

Project

- Goal
  - Chance to explore Machine Learning
  - Can combine with other classes
    - get permission from both instructors; delineate different parts
  - Extra credit for shooting for a publication

- Main categories
  - Application/Survey
    - Compare a bunch of existing algorithms on a new application domain of your interest
  - Formulation/Development
    - Formulate a new model or algorithm for a new or old problem
  - Theory
    - Theoretically analyze an existing algorithm

Project

- For graduate students [5424G]
  - Encouraged to apply ML to your research (aerospace, mechanical, UAVs, computational biology…)
  - Must be done this semester. No double counting.

- For undergraduate students [4424]
  - Chance to implement something
  - No research necessary. Can be an implementation/comparison project.
  - E.g. write an iphone app (predict activity from GPS/gyro data).

- Support
  – We will give a list of ideas, points to dataset/algorithms/code
  – Mentor teams and give feedback.

Spring 2013 Projects

- Poster/Demo Session

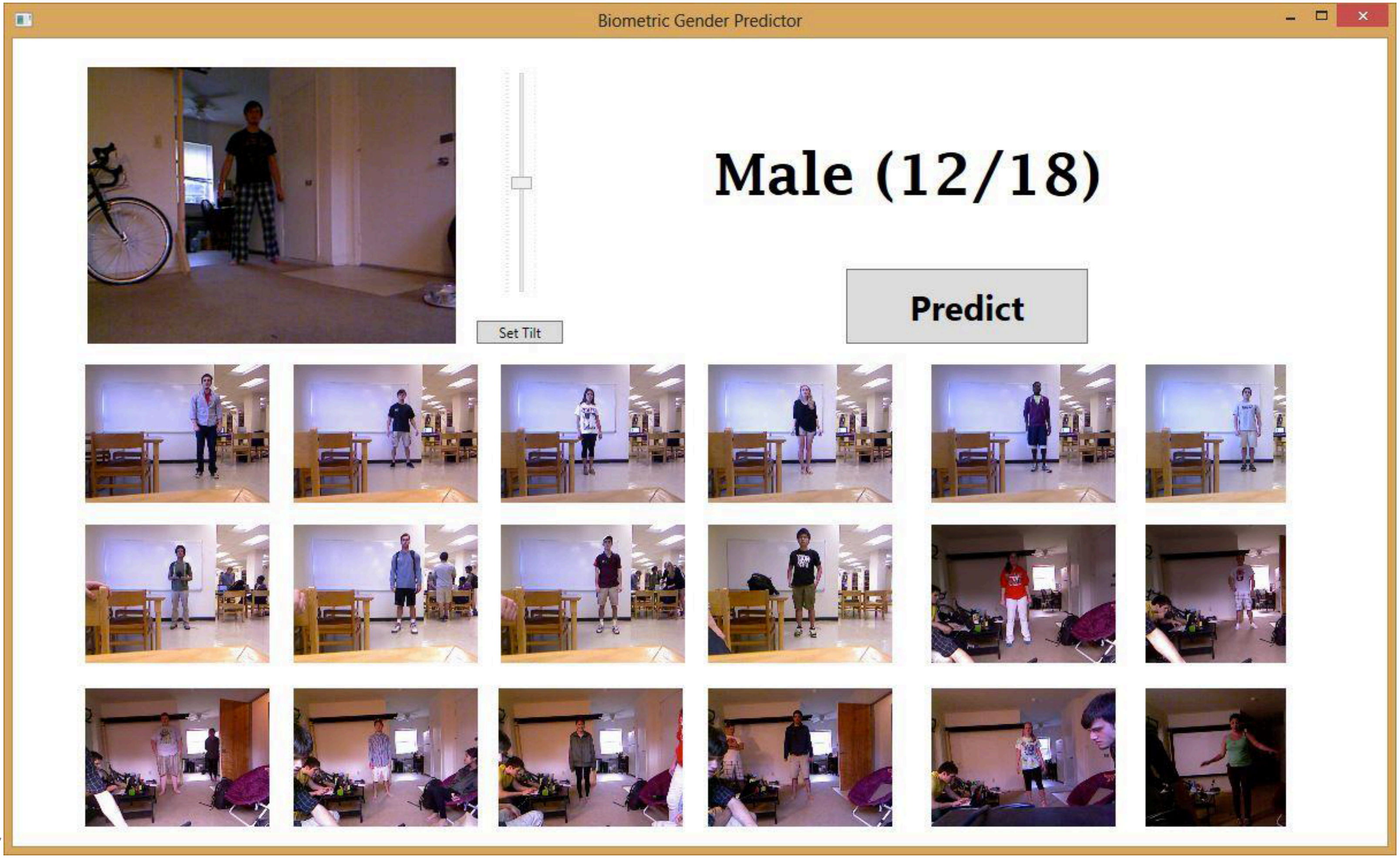
Spring 2013 Projects

- Gesture Activated Interactive Assistant
  - Gordon Christie & Ujwal Krothpalli, Grad Students
  - http://youtu.be/VFPAHY7th9A?t=42s

Figure 7: A simple 2D pose estimation in a controlled setting.

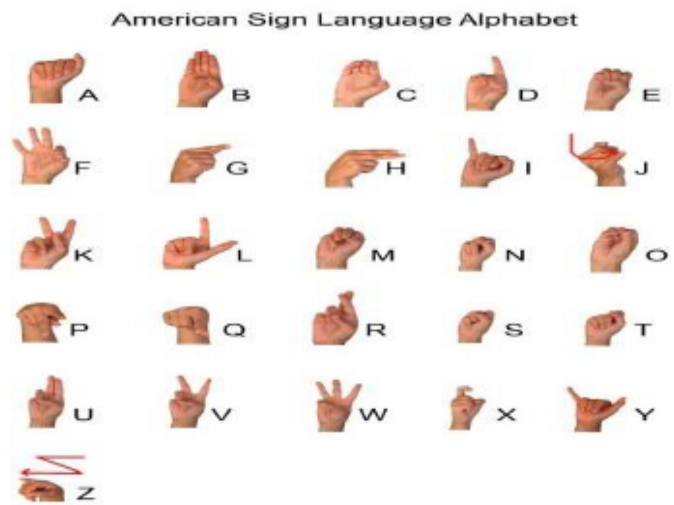
Spring 2013 Projects

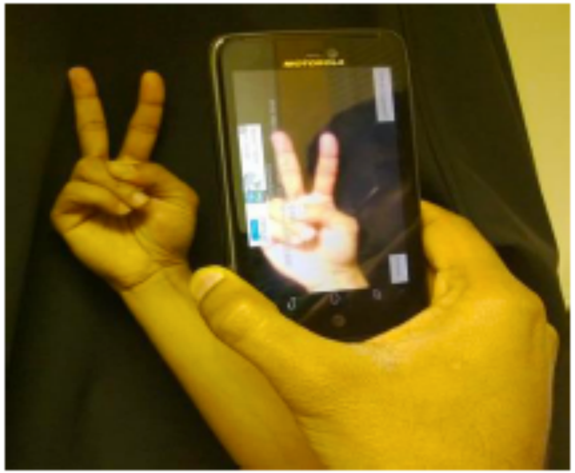
- Gender Classification from body proportions
  - Igor Janjic & Daniel Friedman, Juniors

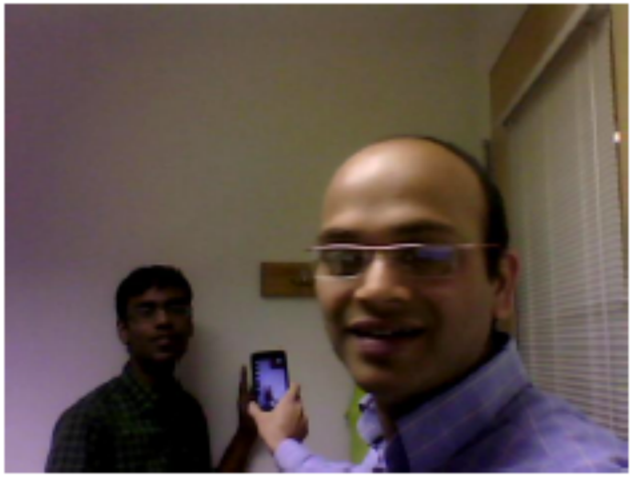


Spring 2013 Projects

- American Sign Language Detection
  - Vireshwar Kumar & Dhiraj Amuru, Grad Students







Collaboration Policy

- Collaboration
  - Only on HW and project (not allowed in exams & HW0).
  - You may discuss the questions
  - Each student writes their own answers
  - Write on your homework anyone with whom you collaborate
  - Each student must write their own code
- Zero tolerance on plagiarism
  - Neither ethical nor in your best interest
  - Always credit your sources
  - Don't cheat. We will find out. Consequences are serious.

Waitlist / Audit / Sit in

- Waitlist
  - Do HW0. Come to first few classes.
  - Let's see how many people drop.
  - Remember: Offered again next year.
- Audit
  - Can't audit Special Studies.
  - Once we get a permanent number:
    Do enough work (your choice) to get 50% grade.
- Sitting in
  - Talk to me.

Communication Channels

- Primary means of communication -- Scholar Forum
  - No direct emails to Instructor unless private information
  - Instructor can mark/provide answers to everyone
  - Class participation credit for answering questions!
  - No posting answers. We will monitor.

- Class websites:
  - https://scholar.vt.edu/portal/site/f16ece5424
  - https://filebox.ece.vt.edu/~f16ece5424/

- Office Hours

How to do well in class?

- Come to class!
  - Sit in front; ask questions!

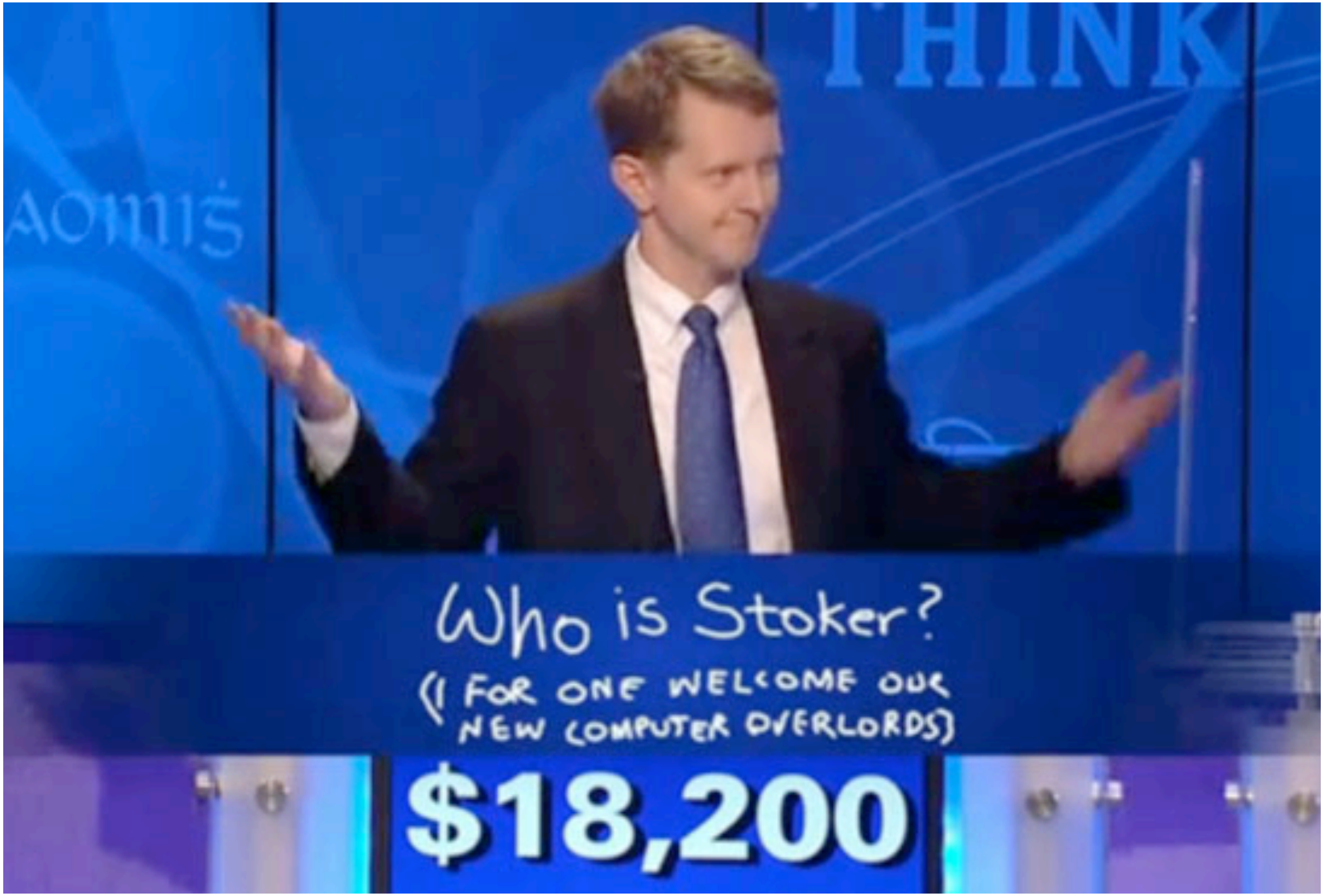
- Start early on the assignments
  - Seriously, start early.
    - You will need the time.
      - These assignments are hard.
        - » Start early.
          - » Start early.
            - » Start early.
- One point
  - No laptops or screens in class

Todo

- HW0
  - Due Thursday by 11:55pm
  - No ‘late days’

- Readings
  - Probability Refresher: Barber Chap 1
  - Overview of ML: Barber Section 13.1

Welcome


THINK
Who is Stoker?
(I FOR ONE WELCOME OUR NEW COMPUTER OVERLORDS)
$18,200