CS 1675: Intro to Machine Learning

Introduction

Prof. Adriana Kovashka
University of Pittsburgh
August 28, 2018

Course Info

- **Course website:** http://people.cs.pitt.edu/~kovashka/cs1675_fa18
- **Instructor**: Adriana Kovashka (kovashka@cs.pitt.edu)
  → Use "CS1675" at the beginning of your Subject
- **Office**: Sennott Square 5325
- **Class:** Tue/Thu, 11am-12:15pm
- **Office hours**: Tue 2-3:55pm, Thu 1-3:55pm

About the Instructor



Born 1985 in Sofia, Bulgaria



Got BA in 2008 at Pomona College, CA (Computer Science & Media Studies)



Got PhD in 2014 at University of Texas at Austin (Computer Vision)

About the TA

- Karin Cox
- **Office:** Sennott Square 6150
- **Office hours:** TBD
  - Do the Doodle by the end of Friday: https://doodle.com/poll/r8kbccezcrzuuh8a

Recitations

- **Time:** Friday, 9am and 1pm
- **Room:** Sennott Square 6110
- **Instructor:** TBD

Course Goals

- To learn the basic machine learning techniques, both from a theoretical and practical perspective
- To practice implementing and using these techniques for simple problems
- To understand the advantages/disadvantages of machine learning algorithms and how they relate to each other

Textbooks

- Christopher M. Bishop. *Pattern Recognition and Machine Learning.* Springer, 2006
- More resources available on course webpage

- Your notes from class are your best study material, slides are not complete with notes

Programming Language

- We’ll use Matlab
- It can be downloaded for free from MyPitt
- We’ll do a short tutorial; ask TA if you need further help

Course Structure

- Lectures
- Weekly assignments
- Two exams
- Participation component

Policies and Schedule

http://people.cs.pitt.edu/~kovashka/cs1675_fa18

Should I take this class?

- It will be a lot of work!
  - I expect you'll spend **6-8 hours** on HW each week
  - But you will learn a lot
- Some parts will be hard and require that you pay close attention!
  - But I will have periodic ungraded pop quizzes to see how you're doing
  - I will also pick on students randomly to answer questions
  - Use instructor's and TA's office hours!!!

Questions?

Plan for Today

- Blitz introductions
- What is machine learning?
  - Example problems and tasks
  - ML in a nutshell
  - Challenges
  - Measuring performance
- Review
  - Linear algebra
  - Calculus
- Matlab tutorial

Blitz introductions (10 sec)

- What is your name?
- What one thing outside of school are you passionate about?
- What do you hope to get out of this class?

- **Every time you speak, please remind me your name**

What is machine learning?

- Finding patterns and relationships in data
- Using these patterns to make useful *predictions* or to *summarize* the data automatically
- E.g.
  - predict how much a user will like a movie, even though that user never rated that movie
  - identify common types of movies without knowing about genres

Example machine learning tasks

- Netflix challenge
  - Given lots of data about how users rated movies (training data)
  - But we don't know how user *i* will rate movie *j* and want to predict that (test data)
  - Why is that hard? How can we do it?

Example machine learning tasks

- Spam or not?

Sebring, Tracy
To: Batra, Dhruv
ECE 4424 proposal

CUSP has approved ECE 4424 with the following cha
copy of the proposal with these items addressed? (se
Thanks!!!
Tracy

vs

nadia bamba
January 19, 2015 5:57 AM
To: undisclosed recipients: ;
Hide Details
Reply-To: nadia bamba
From Miss Nadia BamBa,

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.

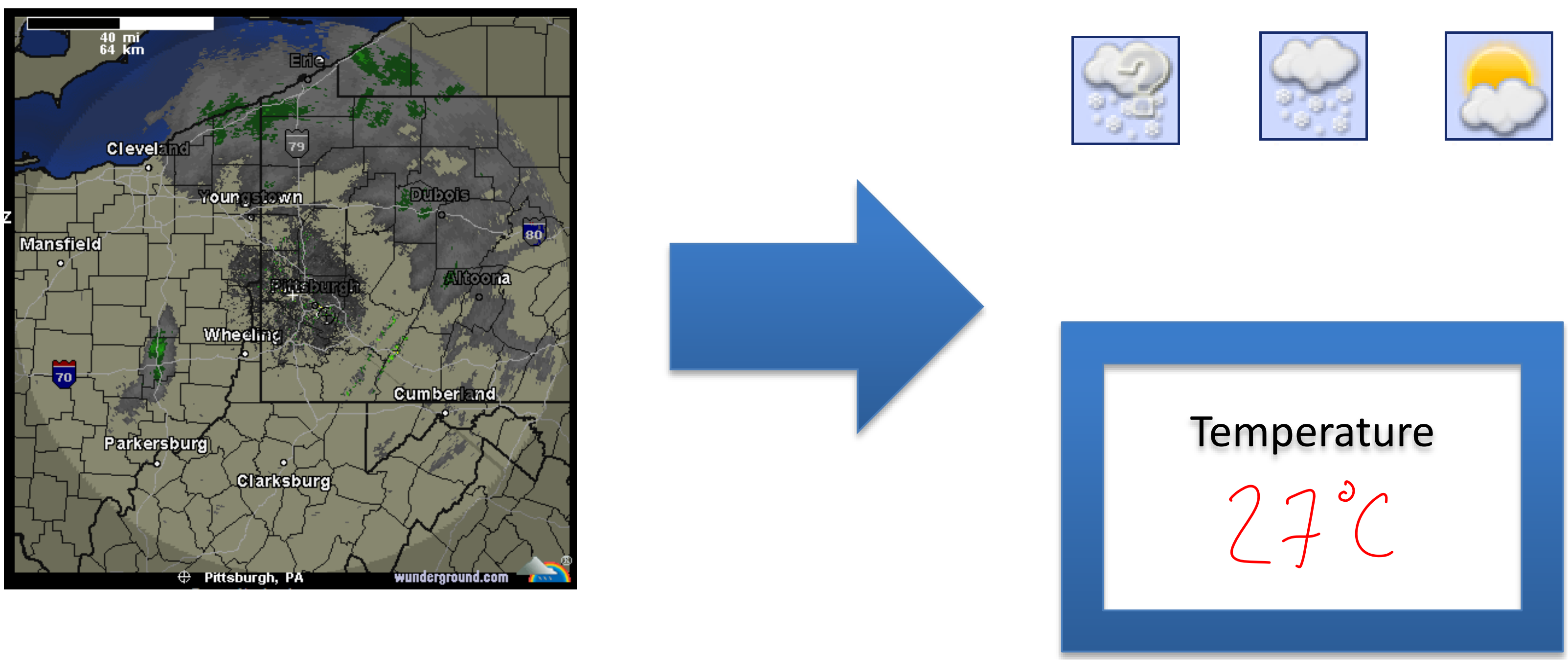
Moreover i am willing to offer you 15% of the total sum as compensation for effort input after the successful transfer of this fund to your designated account overseas,

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

Slide credit: Dhruv Batra

Example machine learning tasks

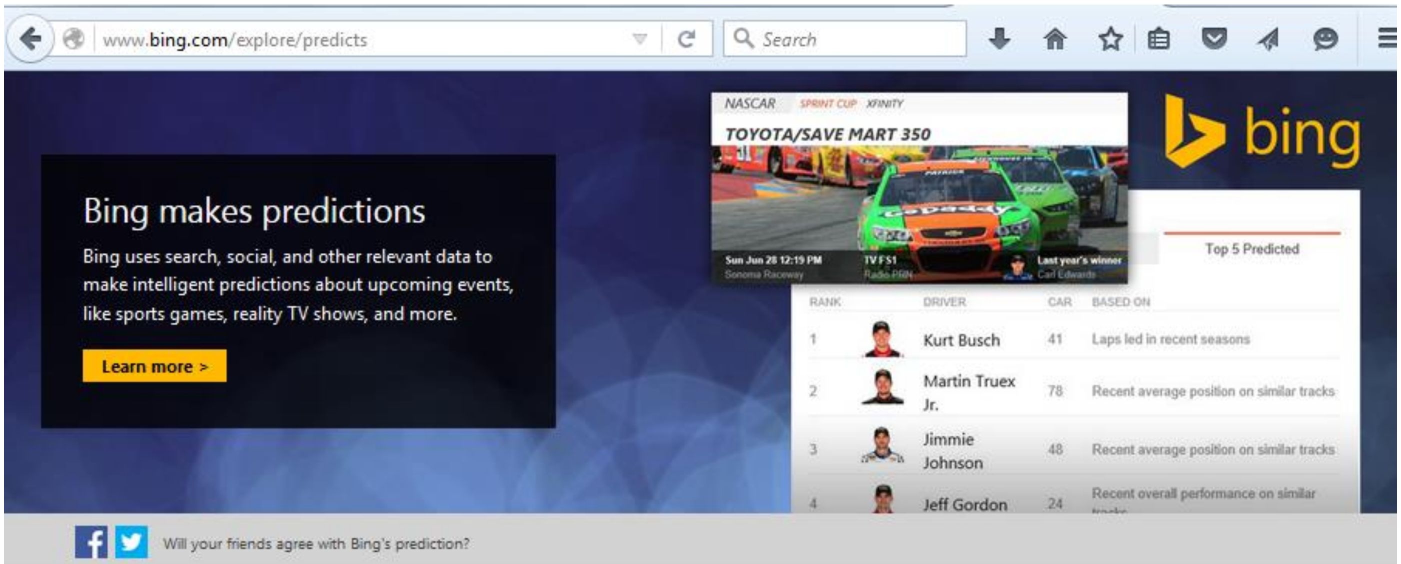
- Weather prediction

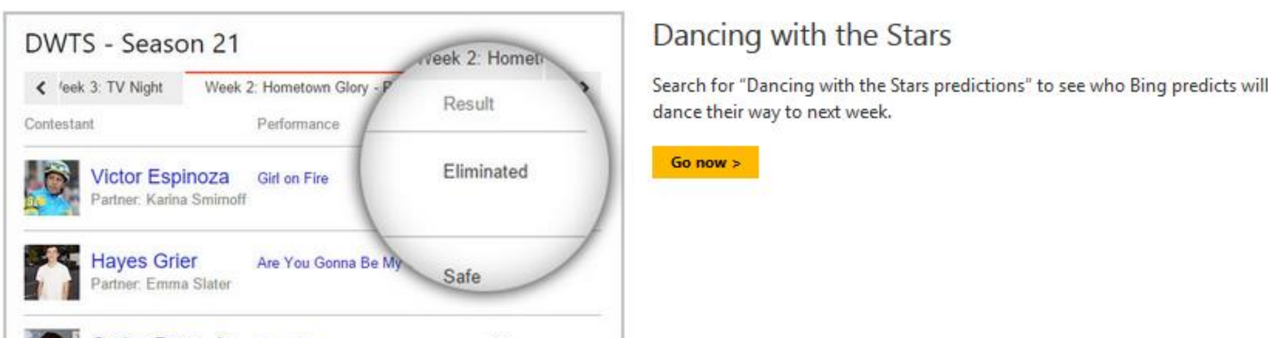


Slide credit: Carlos Guestrin

Example machine learning tasks

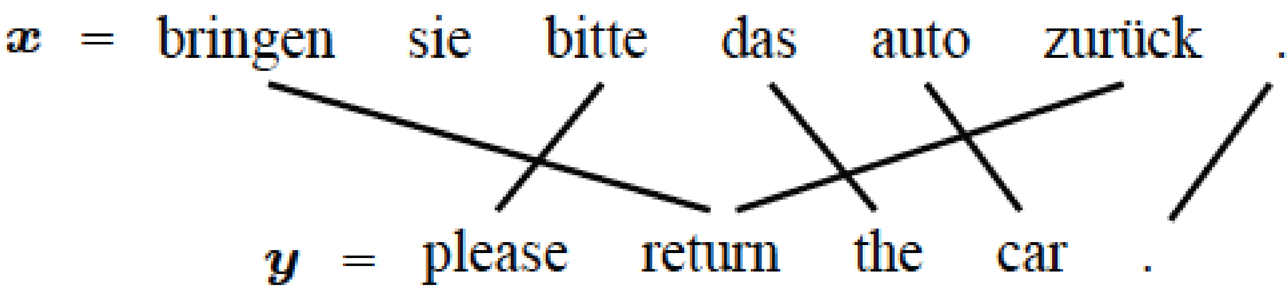
- Who will win <contest of your choice>?





Example machine learning tasks

- Machine translation



Slide credit: Dhruv Batra, figure credit: Kevin Gimpel

Example machine learning tasks

- Speech recognition

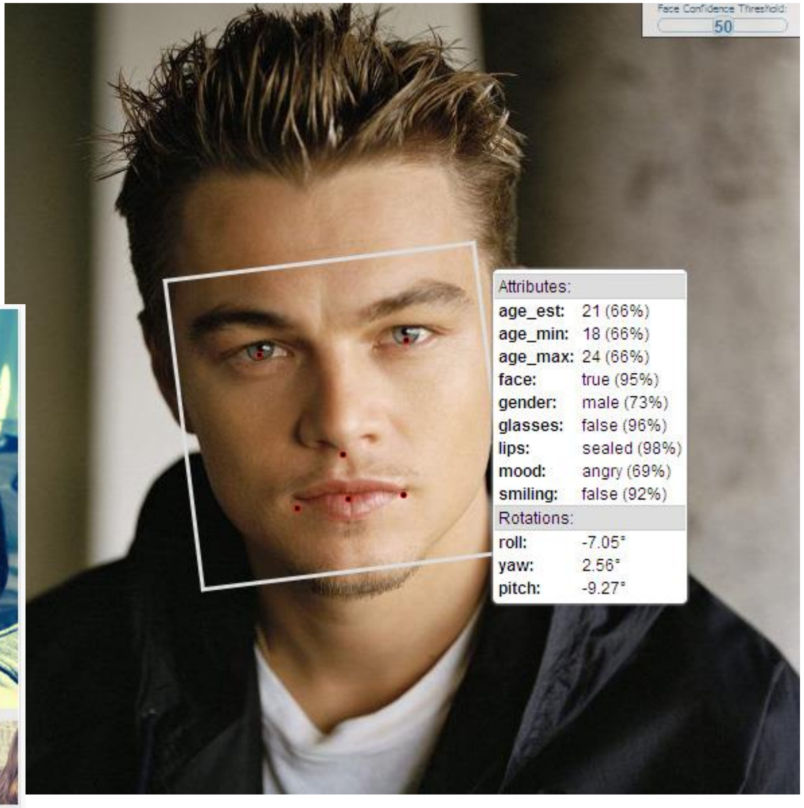
Example machine learning tasks

- Pose estimation

Slide credit: Noah Snavely

Example machine learning tasks

- Face recognition



Slide credit: Noah Snavely

Example machine learning tasks

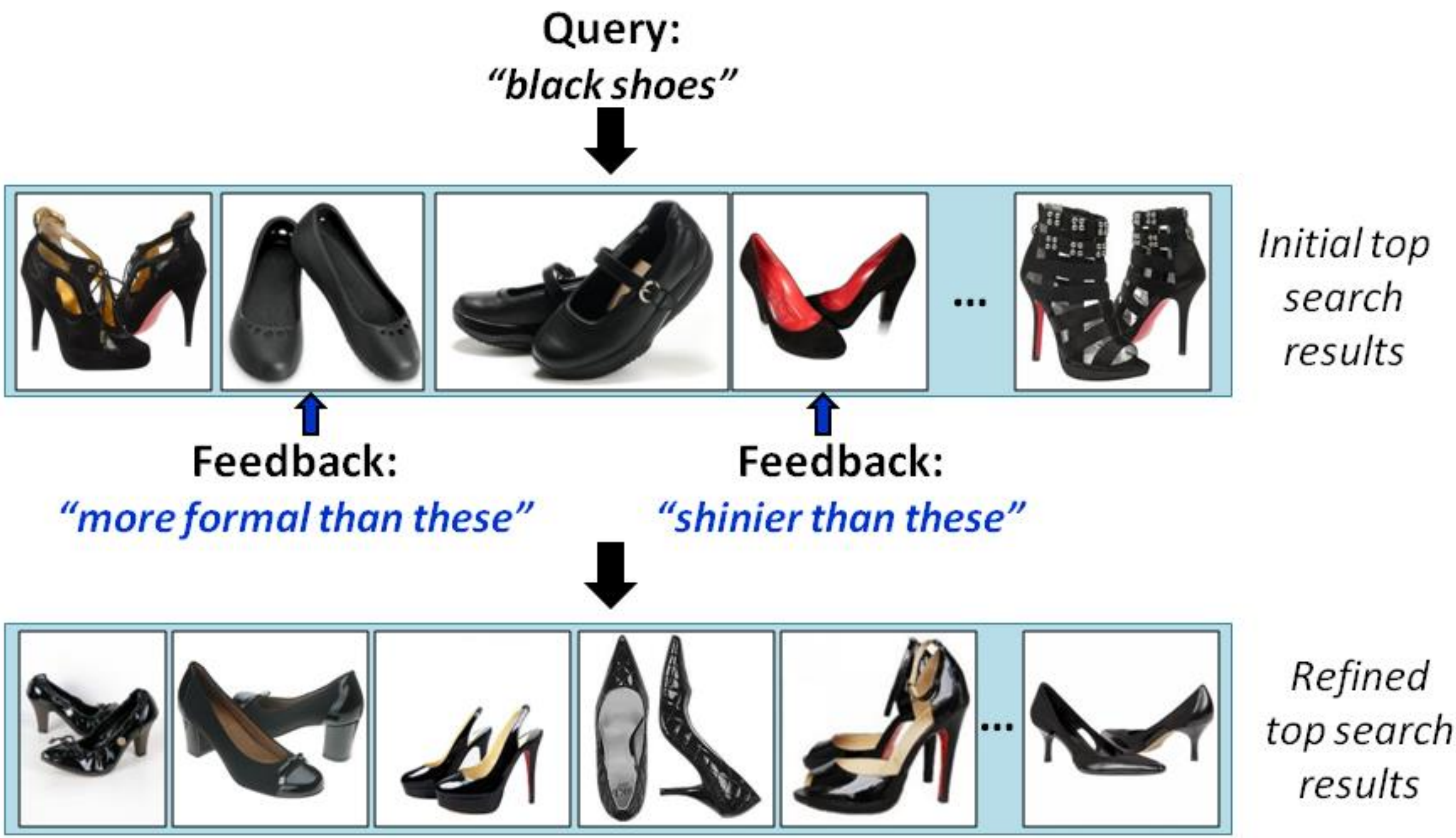
- Image categorization

Pizza
Wine
Stove

Slide credit: Dhruv Batra

Example machine learning tasks

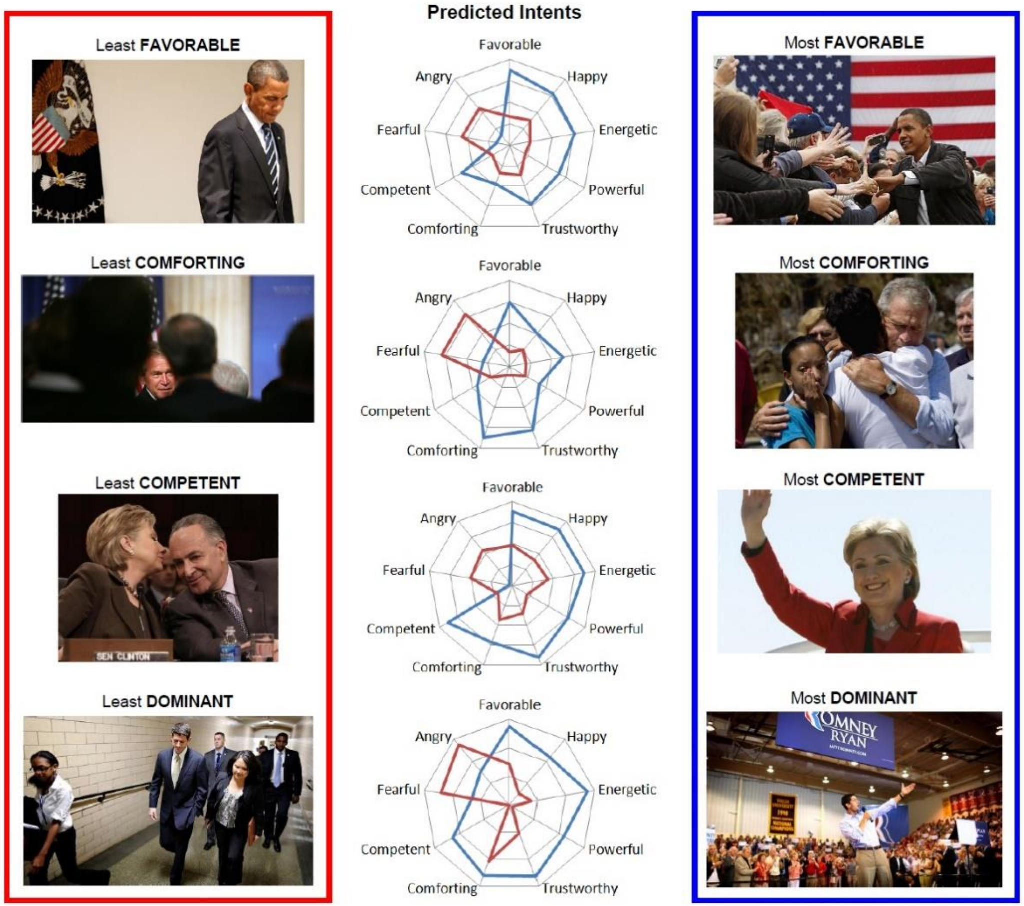
- Image retrieval



Kovashka et al., "WhittleSearch: Image Search with Relative Attribute Feedback", CVPR 2012

Example machine learning tasks

- Inferring visual persuasion



Joo et al., “Visual Persuasion: Inferring Communicative Intents of Images”, CVPR 2014

Example machine learning tasks

- Answering questions about images

What color are her eyes?
What is the mustache made of?

How many slices of pizza are there?
Is this a vegetarian pizza?

Is this person expecting company?
What is just under the tree?

Does it appear to be rainy?
Does this person have 20/20 vision?

Antol et al., “VQA: Visual Question Answering”, ICCV 2015

Example machine learning tasks

- What else?
- What are some problems from everyday life that can be helped by machine learning?

ML in a Nutshell

- Tens of thousands of machine learning algorithms
- Decades of ML research oversimplified:
  - Learn a mapping from input to output f: X → Y
  - X: emails, Y: {spam, notspam}

Slide credit: Pedro Domingos

ML in a Nutshell

$$y = f(\mathbf{x})$$

output prediction function features

- **Training:** given a *training set* of labeled examples $\{(\mathbf{x}_1,y_1), \ldots, (\mathbf{x}_N,y_N)\}$, estimate the prediction function f by minimizing the prediction error on the training set
- **Testing:** apply f to a never before seen *test example* $\mathbf{x}$ and output the predicted value $y = f(\mathbf{x})$

Slide credit: L. Lazebnik

ML in a Nutshell

- Apply a prediction function to a feature representation of the image to get the desired output:

f() = “apple”

f() = “tomato”

f() = “cow”

Slide credit: L. Lazebnik

ML in a Nutshell

Training

Training Images

Features → Training → Learned model

Training Labels → Training

Testing

Test Image

Features → Learned model → Prediction

Slide credit: D. Hoiem and L. Lazebnik

Training vs Testing

- What do we want?
  - High accuracy on training data?
  - No, high accuracy on *unseen/new/test data!*
  - Why is this tricky?
- Training data
  - Features (x) and labels (y) used to learn mapping f
- Test data
  - Features used to make a prediction
  - Labels only used to see how well we've learned f!!!
- Validation data
  - Held-out set of the *training data*
  - Can use both features and labels to tune *parameters* of the model we're learning

Why do we hope this would work?

- Statistical estimation view:
  - x and y are *random variables*
  - $D = (x_1,y_1), (x_2,y_2), \ldots, (x_N,y_N) \sim P(X,Y)$
  - Both training & testing data sampled IID from P(X,Y)
    - IID: Independent and Identically Distributed
  - Learn on training set, have some hope of *generalizing* to test set

Adapted from Dhruv Batra

ML in a Nutshell

- Every machine learning algorithm has:
  - Data representation (x, y)
  - Problem representation
  - Evaluation / objective function
  - Optimization

Adapted from Pedro Domingos

Data representation

- Let’s brainstorm what our “X” should be for various “Y” prediction tasks…

Problem representation

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

Slide credit: Pedro Domingos

Evaluation / objective function

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

Slide credit: Pedro Domingos

Optimization

- Discrete / combinatorial optimization
  - E.g. graph algorithms

- Continuous optimization
  - E.g. linear programming

$$
\begin{array}{ll}
\text{maximize} & \mathbf{c}^{\mathrm{T}}\mathbf{x} \\
\text{subject to} & A\mathbf{x} \leq \mathbf{b} \\
\text{and} & \mathbf{x} \geq \mathbf{0}
\end{array}
$$

Adapted from Dhruv Batra, image from Wikipedia

Defining the Learning Task

Improve on task, T, with respect to
performance metric, P, based on experience, E.

T: Categorize email messages as spam or legitimate
P: Percentage of email messages correctly classified
E: Database of emails, some with human-given labels

T: Recognizing hand-written words
P: Percentage of words correctly classified
E: Database of human-labeled images of handwritten words

T: Playing checkers
P: Percentage of games won against an arbitrary opponent
E: Playing practice games against itself

T: Driving on four-lane highways using vision sensors
P: Average distance traveled before a human-judged error
E: A sequence of images and steering commands recorded while observing a human driver.

Slide credit: Ray Mooney

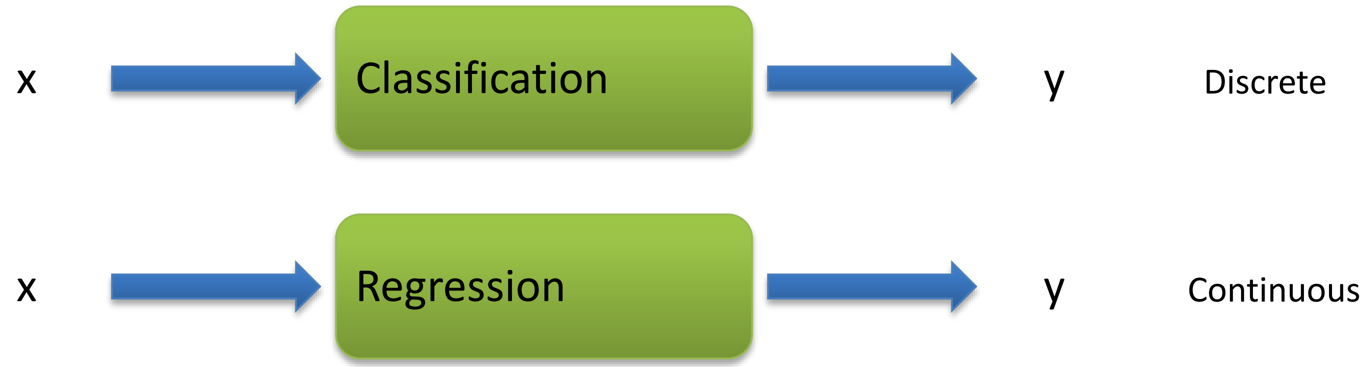
Types of Learning

- 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

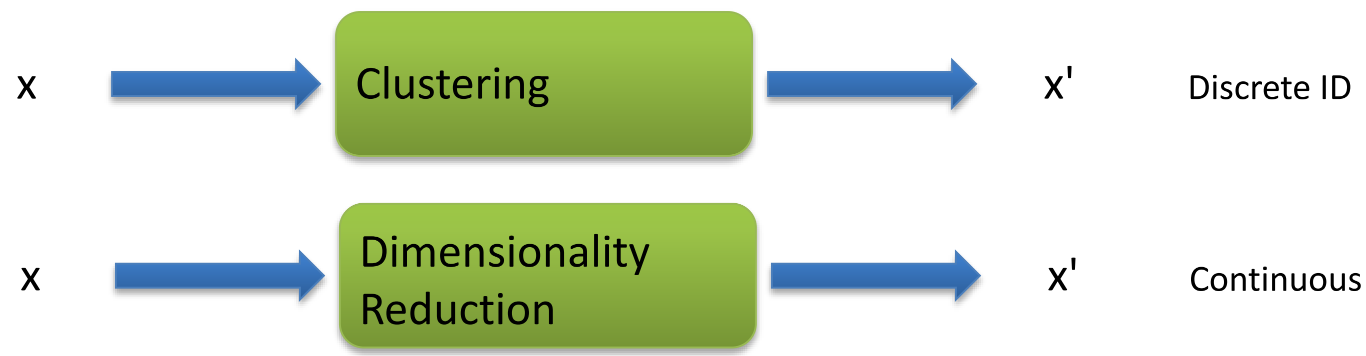
Slide credit: Dhruv Batra

Types of Prediction Tasks

Supervised Learning

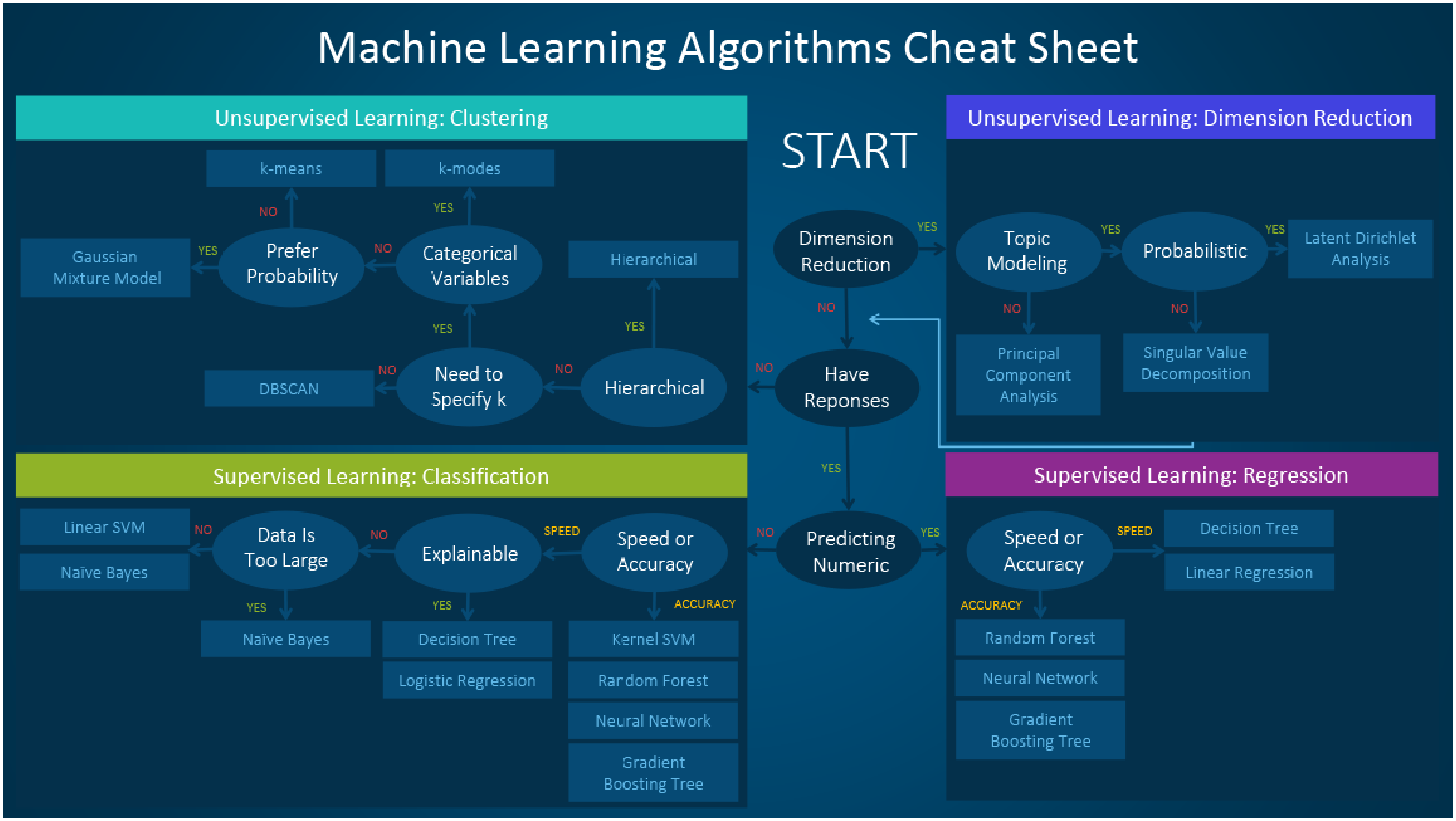


Unsupervised Learning



Adapted from Dhruv Batra

Navigating ML World



https://blogs.sas.com/content/subconsciousmusings/2017/04/12/machine-learning-algorithm-use/

Example of Solving a ML Problem

- Spam or not?

Sebring, Tracy
To: Batra, Dhruv
ECE 4424 proposal

CUSP has approved ECE 4424 with the following cha
copy of the proposal with these items addressed? (se
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.

Slide credit: Dhruv Batra

Intuition

- Spam Emails
  - a lot of words like
    - “money”
    - “free”
    - “bank account”
- Regular Emails
  - word usage pattern is more spread out

Slide credit: Dhruv Batra, Fei Sha

Simple strategy: Let's count!

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

From Miss Nadia BamBa,

Greeting, Permit me to inform you of my desire of going i
Nadia BamBa the only Daughter of late Mr and Mrs Jame
cocoa merchant in Abidjan, the economic capital of Ivory
his business associates on one of their outing to discus c
on the 21st October 2002, my father took me very speci

Before the death of my father in a private hospital here i
bedside and told me that he had a sum of $6, 8000.000(S
Dollars) left in a suspense account in a Bank here in Abic
Daughter for the next of kin in deposit of the fund.

Sebring, Tracy
To: Batra, Dhruv
ECE 4424 proposal

CUSP has approved ECE 4424 with the following changes: Can
copy of the proposal with these items addressed? (see below)
Thanks!!!
Tracy

This is X

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

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

This is Y

= 1 or 0?

Adapted from Dhruv Batra, Fei Sha

Weigh counts and sum to get prediction

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

From Miss Nadia BamBa,

Greeting, Permit me to inform you o
Nadia BamBa the only Daughter of la
cocoa merchant in Abidjan, the econ
his business associates on one of th
on the 21st October 2002, my father

Before the death of my father in a p
bedside and told me that he had a su
Dollars) left in a suspense account i
Daughter for the next of kin in depos

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

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

Why these words?

Adapted from Dhruv Batra, Fei Sha

Why not just hand-code these weights?

- We’re letting the data do the work rather than develop hand-code classification rules
  - The *machine* is *learning* to program itself

- But there are challenges…

Klingon vs Mlingon Classification

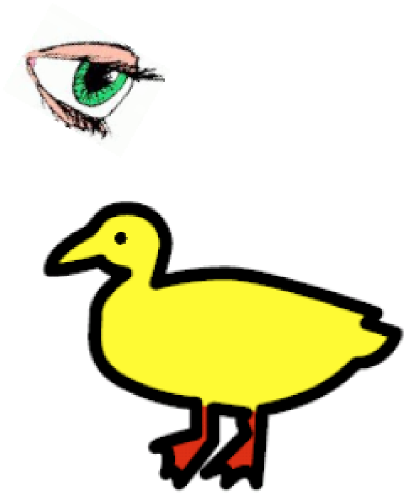
- Training Data
  - Klingon: klix, kour, koop
  - Mlingon: moo, maa, mou

- Testing Data: kap

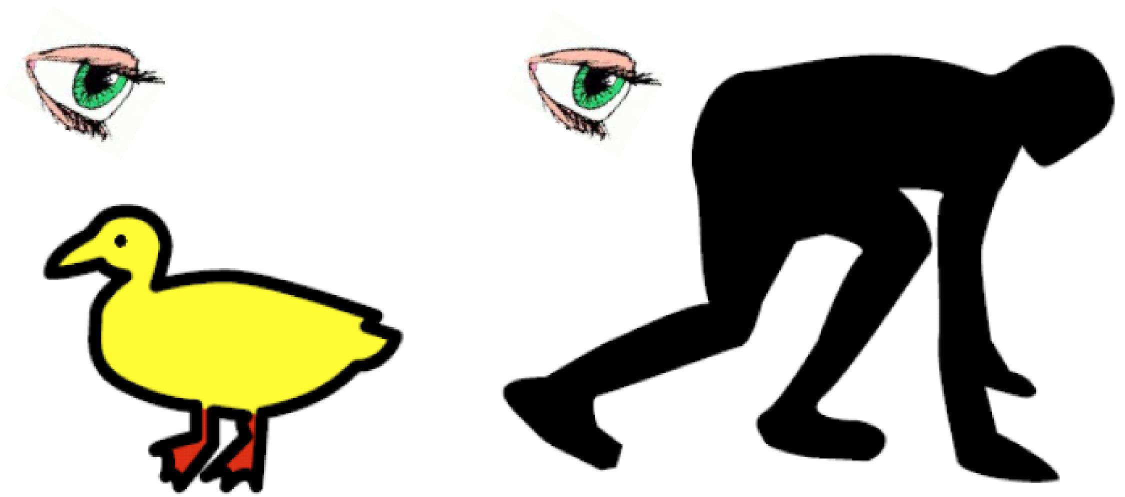
- Which language? Why?

Slide credit: Dhruv Batra

“I saw her duck”

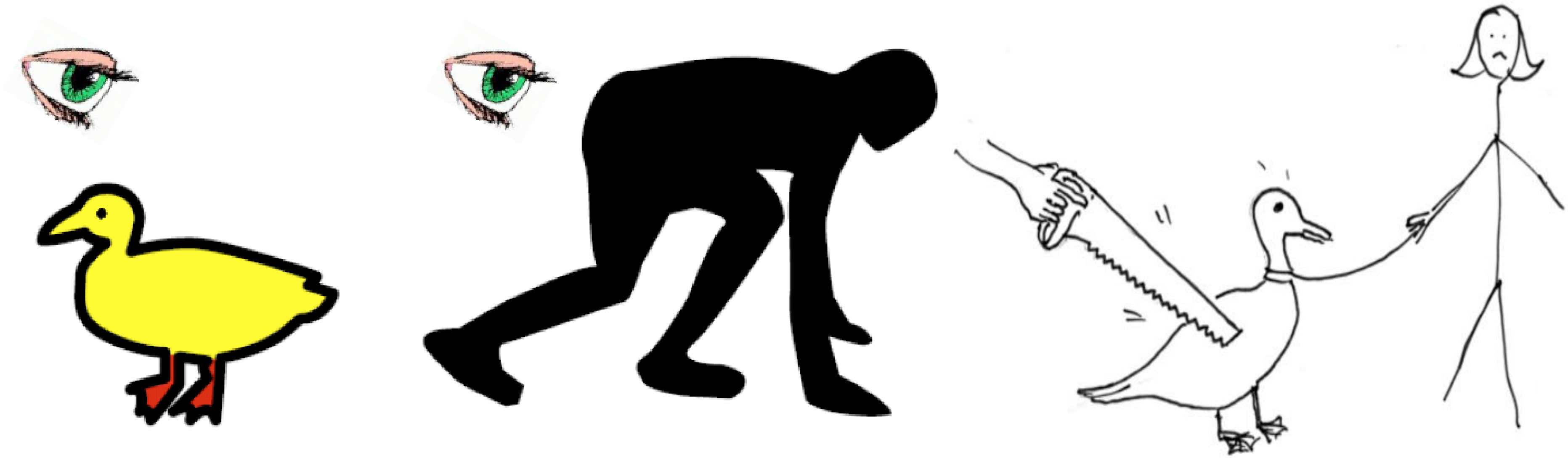


“I saw her duck”



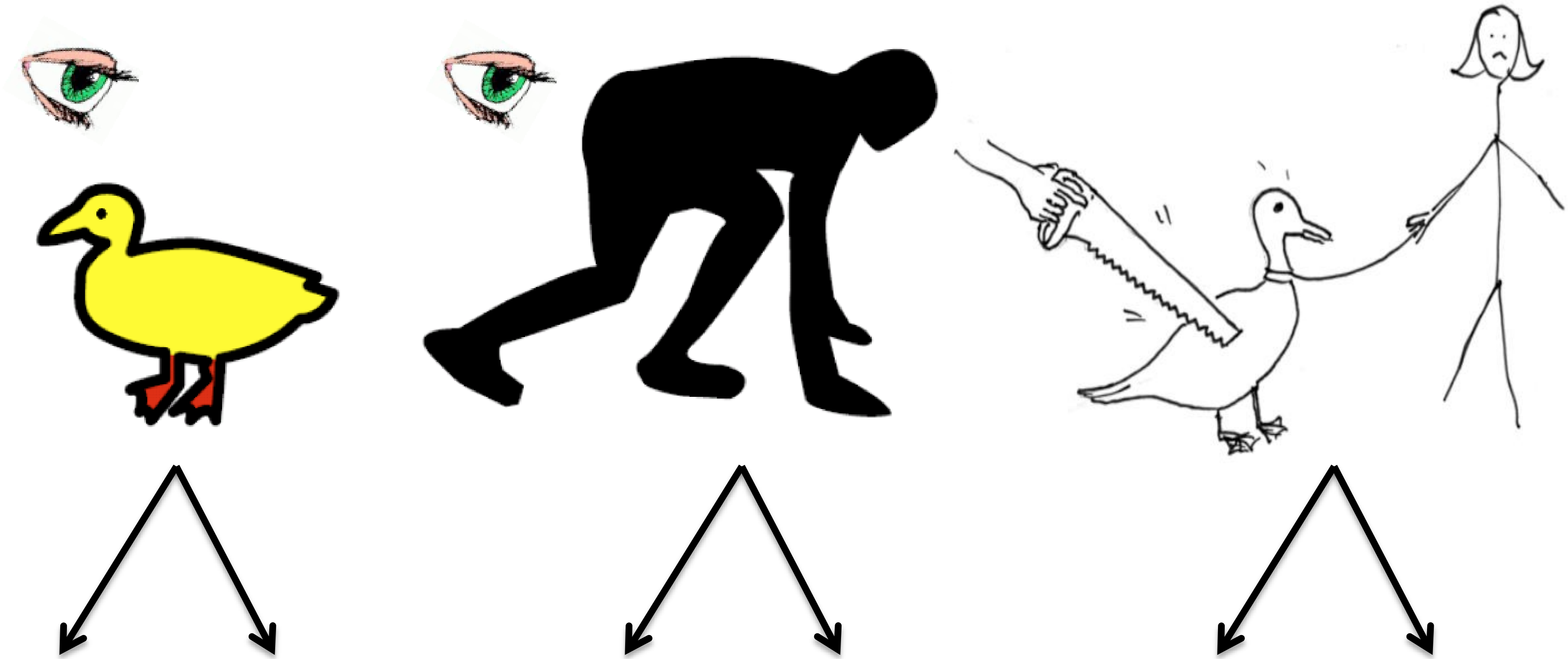
Slide credit: Dhruv Batra, figure credit: Liang Huang

“I saw her duck”



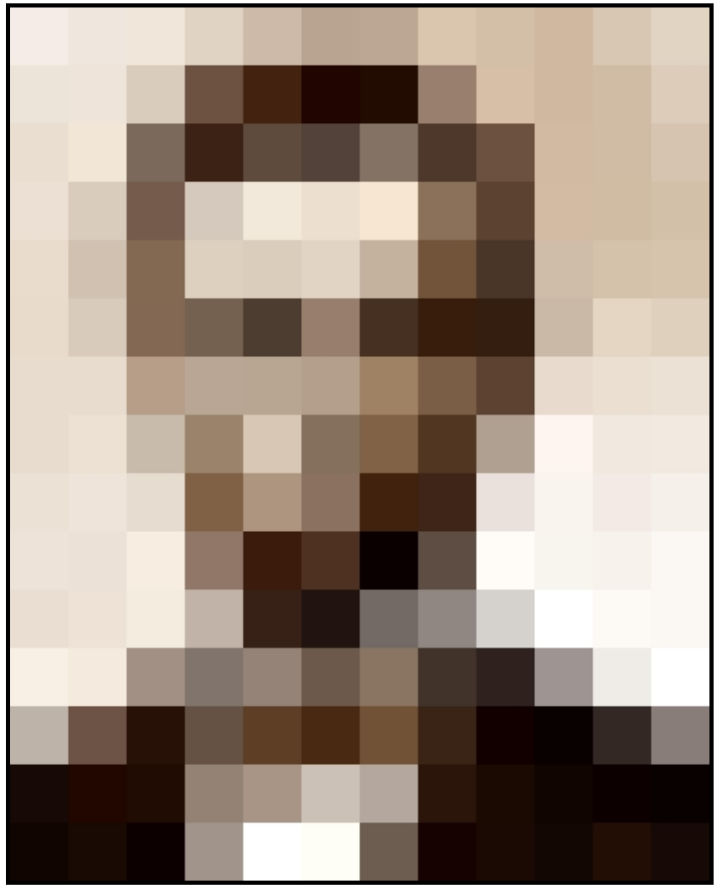
Slide credit: Dhruv Batra, figure credit: Liang Huang

“I saw her duck with a telescope…”



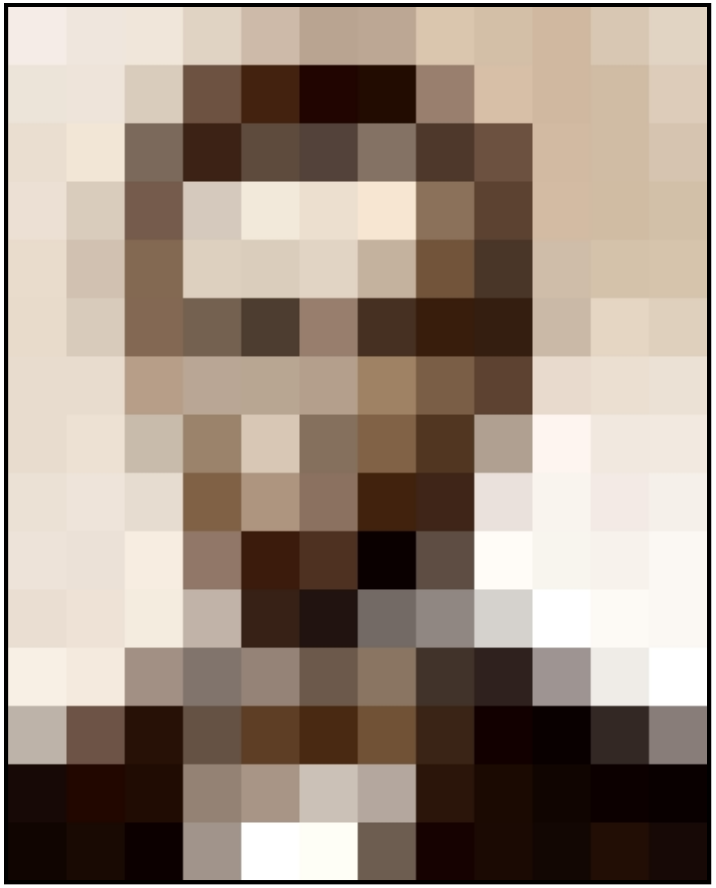
Slide credit: Dhruv Batra, figure credit: Liang Huang

What humans see



Slide credit: Larry Zitnick

What computers see



Slide credit: Larry Zitnick

Challenges

- Some challenges: ambiguity and context
- Machines take data representations too literally
- Humans are much better than machines at generalization, which is needed since test data will rarely look exactly like the training data

Challenges

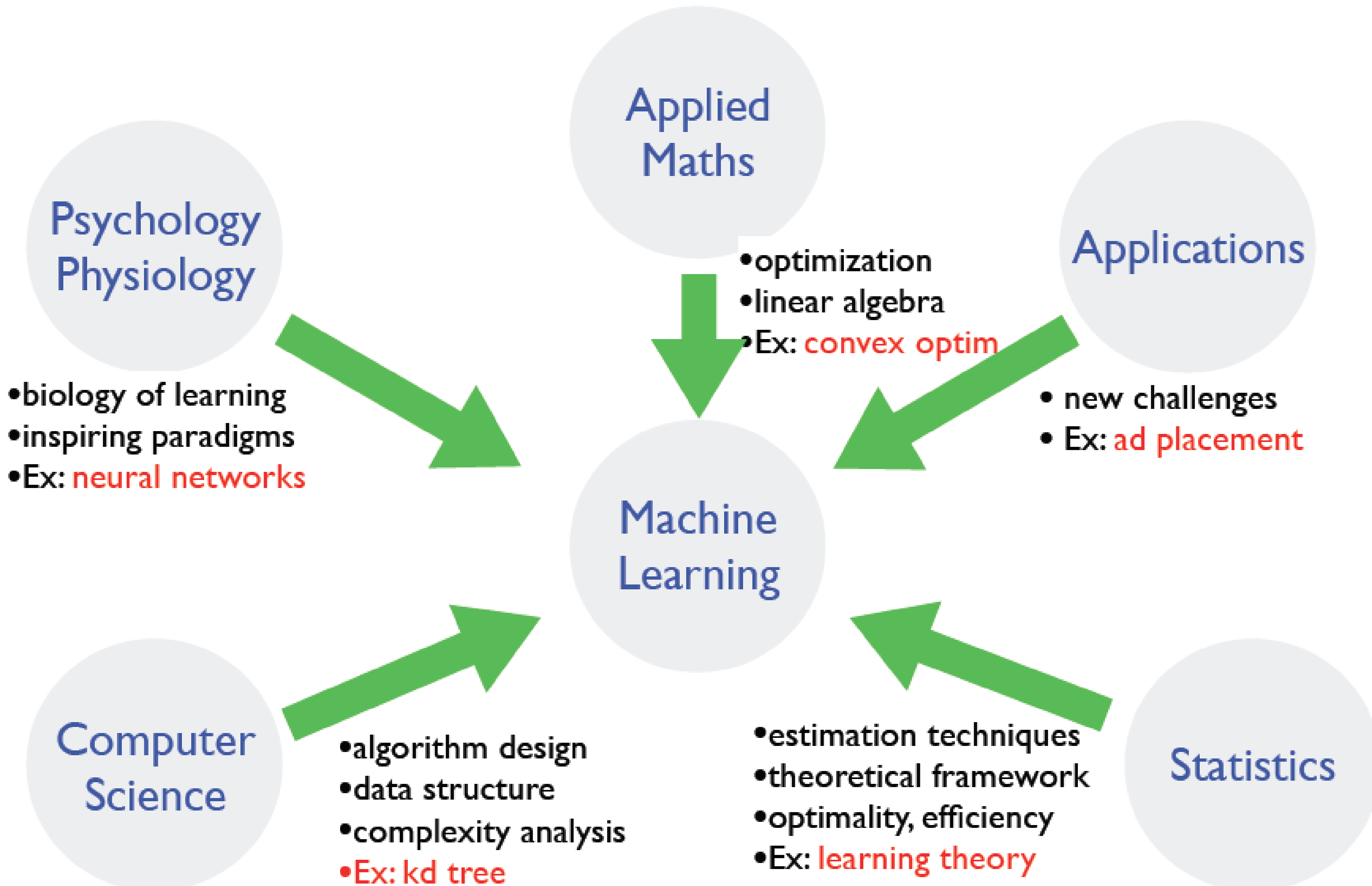
- Why might it be hard to:
  - Predict if a viewer will like a movie?
  - Recognize cars in images?
  - Translate between languages?

The Time is Ripe to Study ML

- Many basic effective and efficient algorithms available.
- Large amounts of on-line data available.
- Large amounts of computational resources available.

Slide credit: Ray Mooney

Where does ML fit in?



Slide credit: Dhruv Batra, Fei Sha

Plan for Today

- Blitz introductions
- What is machine learning?
  - Example problems and tasks
  - ML in a nutshell
  - Challenges
  - Measuring performance
- Review
  - Linear algebra
  - Calculus
- Matlab tutorial

Measuring Performance

- If *y* is discrete:
  - Accuracy: # correctly classified / # all test examples
  - Precision/recall
    - True Positive, False Positive, True Negative, False Negative
    - **P**recision = TP / (TP + FP) = # predicted true pos / # predicted pos
    - **R**ecall = TP / (TP + FN) = # predicted true pos / # true pos
  - F-measure

    = 2PR / (P + R)

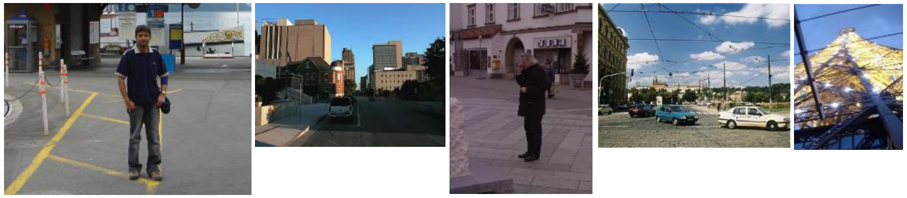
- Want evaluation metric to be in some range, e.g. [0 1]
  - 0 = worst possible classifier, 1 = best possible classifier

Precision / Recall / F-measure

True positives
(images that contain people)

True negatives
(images that *do not* contain people)

Predicted positives
(images *predicted to* contain people)



Predicted negatives
(images *predicted not to* contain people)

- Precision = TP / (TP + FP)
- Recall = TP / (TP + FN)
- F-measure :

Accuracy:

Measuring Performance

- If *y* is continuous:
  - Sum-of-Squared-Differences (SSD) error between predicted and true *y*:

$$\mathbf{E} = \sum_{\mathbf{i}=1}^{\mathbf{n}} (\mathbf{f}(\mathbf{x_i}) - \mathbf{y_i})^2$$

Linear algebra review

See http://cs229.stanford.edu/section/cs229-linalg.pdf for more

Vectors and Matrices

- Vectors and matrices are just collections of ordered numbers that represent something: movements in space, scaling factors, word counts, movie ratings, pixel brightnesses, etc.
- We'll define some common uses and standard operations on them.

Vector

- A column vector $\mathbf{v} \in \mathbb{R}^{n \times 1}$ where

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

- A row vector $\mathbf{v}^T \in \mathbb{R}^{1 \times n}$ where

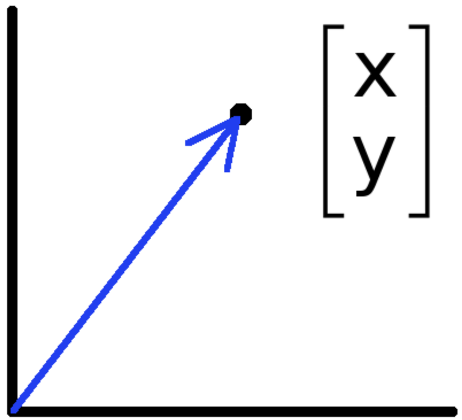
$$\mathbf{v}^T = \begin{bmatrix} v_1 & v_2 & \dots & v_n \end{bmatrix}$$

T denotes the transpose operation

Vector

- You'll want to keep track of the orientation of your vectors when programming in MATLAB.
- You can transpose a vector *V* in MATLAB by writing **V'**.

Vectors have two main uses



- Vectors can represent an offset in 2D or 3D space
- Points are just vectors from the origin

- Data can also be treated as a vector
- Such vectors don't have a geometric interpretation, but calculations like "distance" still have value

Matrix

- A matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ is an array of numbers with size $m \downarrow$ by $n \rightarrow$, i.e. m rows and n columns.

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ \vdots & & & & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{bmatrix}$$

- If $m = n$, we say that $\mathbf{A}$ is square.

Matrix Operations

- Addition

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} + \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} a+1 & b+2 \\ c+3 & d+4 \end{bmatrix}$$

  – Can only add a matrix with matching dimensions, or a scalar.

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} + 7 = \begin{bmatrix} a+7 & b+7 \\ c+7 & d+7 \end{bmatrix}$$

- Scaling

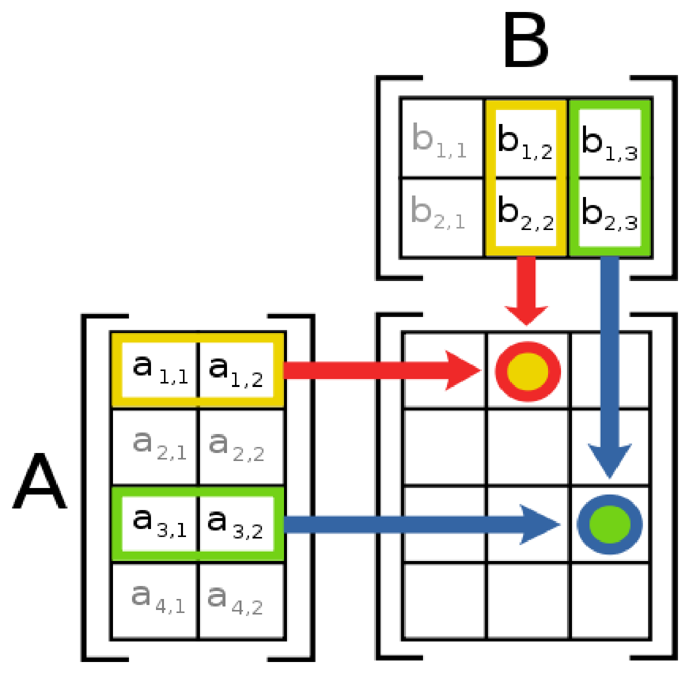
$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \times 3 = \begin{bmatrix} 3a & 3b \\ 3c & 3d \end{bmatrix}$$

Matrix Multiplication

- Let X be an *a*x*b* matrix, Y be an *b*x*c* matrix
- Then Z = X*Y is an *a*x*c* matrix
- Second dimension of first matrix, and first dimension of first matrix have to be the same, for matrix multiplication to be possible
- Practice: Let X be an 10x5 matrix. Let’s factorize it into 3 matrices…

Matrix Multiplication

- The product AB is:



- Each entry in the result is (that row of A) dot product with (that column of B)

Matrix Multiplication

- Example:

A x B

$$\begin{bmatrix} 1 & \boxed{3} \\ 5 & \boxed{7} \end{bmatrix}$$

$$\begin{bmatrix} \boxed{0} & \boxed{2} \\ 4 & 6 \end{bmatrix} \begin{bmatrix} \square & \boxed{14} \\ \square & \square \end{bmatrix}$$

$$0 \cdot 3 + 2 \cdot 7 = \boxed{14}$$

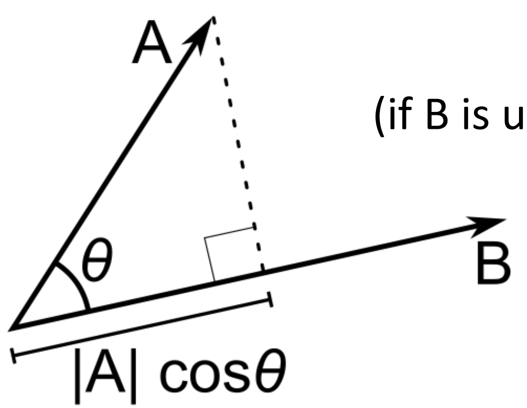
  - Each entry of the matrix product is made by taking the dot product of the corresponding row in the left matrix, with the corresponding column in the right one.

Inner Product

- Multiply corresponding entries of two vectors and add up the result

$$\mathbf{x}^T\mathbf{y} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^{n} x_i y_i \qquad \text{(scalar)}$$

- x·y is also |x||y|Cos(*angle between x and y*)
- If B is a unit vector, then A·B gives the length of A which lies in the direction of B (projection)



(if B is unit-length hence norm is 1)

Different types of product

- $\boldsymbol{x}, \boldsymbol{y}$ = column vectors (nx1)
- $\boldsymbol{X}, \boldsymbol{Y}$ = matrices (mxn)
- x, y = scalars (1x1)

- $\boldsymbol{x}^T \boldsymbol{y} = \boldsymbol{x} \cdot \boldsymbol{y}$ = inner product (1xn x nx1 = scalar)
- $\boldsymbol{x} \otimes \boldsymbol{y} = \boldsymbol{x}\, \boldsymbol{y}^T$ = outer product (nx1 x 1xn = matrix)

- $\boldsymbol{X} * \boldsymbol{Y}$ = matrix product
- $\boldsymbol{X}\ .\!*\ \boldsymbol{Y}$ = element-wise product

Inverse

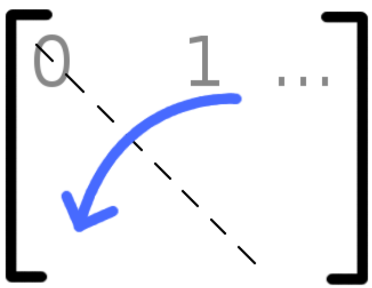
- Given a matrix **A**, its inverse $\mathbf{A}^{-1}$ is a matrix such that $\mathbf{AA}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$

- E.g. $\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}^{-1} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{3} \end{bmatrix}$

- Inverse does not always exist. If $\mathbf{A}^{-1}$ exists, **A** is *invertible* or *non-singular*. Otherwise, it's *singular*.

Matrix Operations

- Transpose – flip matrix, so row 1 becomes column 1



$$\begin{bmatrix} 0 & 1 \\ 2 & 3 \\ 4 & 5 \end{bmatrix}^T = \begin{bmatrix} 0 & 2 & 4 \\ 1 & 3 & 5 \end{bmatrix}$$

- A useful identity:

$$(ABC)^T = C^T B^T A^T$$

Norms

- L1 norm

$$\|\boldsymbol{x}\|_1 := \sum_{i=1}^{n} |x_i|$$

- L2 norm

$$\|\boldsymbol{x}\| := \sqrt{x_1^2 + \cdots + x_n^2}$$

- L^p norm (for real numbers $p \geq 1$)

$$\|\mathbf{x}\|_p := \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$$

Matrix Rank

- Column/row rank

$$\begin{aligned} \text{col-rank}(\mathbf{A}) = &\ \text{the maximum number of linearly independent column vectors of } \mathbf{A} \\ \text{row-rank}(\mathbf{A}) = &\ \text{the maximum number of linearly independent row vectors of } \mathbf{A} \end{aligned}$$

- Column rank always equals row rank
- Matrix rank $\quad \text{rank}(\mathbf{A}) \triangleq \text{col-rank}(\mathbf{A}) = \text{row-rank}(\mathbf{A})$
- If a matrix is not full rank, inverse doesn’t exist
  - Inverse also doesn’t exist for non-square matrices

Matrix Operation Properties

- Matrix addition is commutative and associative
  - A + B = B + A
  - A + (B + C) = (A + B) + C
- Matrix multiplication is associative and distributive but *not* commutative
  - A(B*C) = (A*B)C
  - A(B + C) = A*B + A*C
  - A*B != B*A

Special Matrices

- Identity matrix **I**
  - Square matrix, 1's along diagonal, 0's elsewhere
  - **I** · [another matrix] = [that matrix]

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

- Diagonal matrix
  - Square matrix with numbers along diagonal, 0's elsewhere
  - A diagonal · [another matrix] scales the rows of that matrix

$$\begin{bmatrix} 3 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & 2.5 \end{bmatrix}$$

Special Matrices

- Symmetric matrix

$$\mathbf{A}^T = \mathbf{A}$$

$$\begin{bmatrix} 1 & 2 & 5 \\ 2 & 1 & 7 \\ 5 & 7 & 1 \end{bmatrix}$$

Matrix Operations

- MATLAB example:

$$AX = B$$

$$A = \begin{bmatrix} 2 & 2 \\ 3 & 4 \end{bmatrix}, B = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

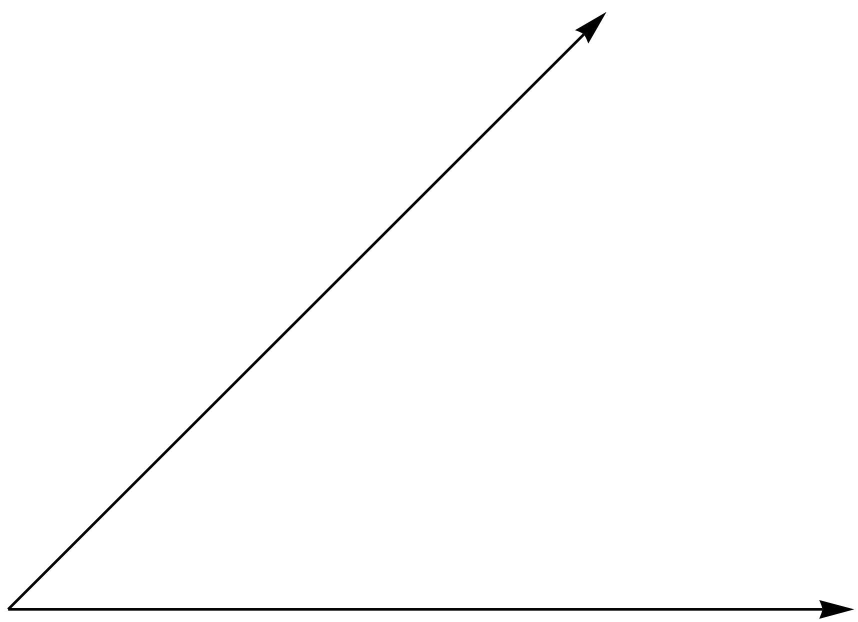
```
>> x = A\B
x =
     1.0000
    -0.5000
```

Linear independence

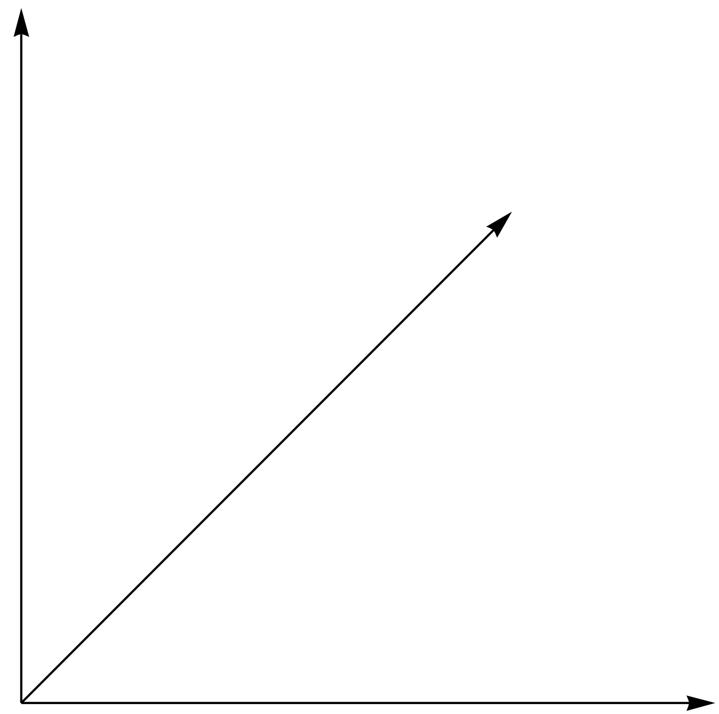
- Suppose we have a set of vectors $\mathbf{v_1}, \ldots, \mathbf{v_n}$
- If we can express $\mathbf{v}_1$ as a linear combination of the other vectors $\mathbf{v}_2 \ldots \mathbf{v}_n$, then $\mathbf{v}_1$ is linearly *dependent* on the other vectors.
  - The direction $\mathbf{v}_1$ can be expressed as a combination of the directions $\mathbf{v}_2 \ldots \mathbf{v}_n$. (E.g. $\mathbf{v}_1 = .7\ \mathbf{v}_2 - .7\ \mathbf{v}_4$)
- If no vector is linearly dependent on the rest of the set, the set is linearly *independent*.
  - Common case: a set of vectors $\mathbf{v_1}, \ldots, \mathbf{v_n}$ is always linearly independent if each vector is perpendicular to every other vector (and non-zero)

Linear independence

Linearly independent set



Not linearly independent



Singular Value Decomposition (SVD)

- There are several computer algorithms that can “factor” a matrix, representing it as the product of some other matrices
- The most useful of these is the Singular Value Decomposition
- Represents any matrix **A** as a product of three matrices: $\mathbf{U\Sigma V^T}$
- MATLAB command: [U,S,V] = svd(A);

Singular Value Decomposition (SVD)

$$\mathbf{U\Sigma V^T = A}$$

- Where **U** and **V** are rotation matrices, and **Σ** is a scaling matrix. For example:

$$\begin{array}{cccc} U & \Sigma & V^T & A \\ \begin{bmatrix} -.40 & .916 \\ .916 & .40 \end{bmatrix} \times & \begin{bmatrix} 5.39 & 0 \\ 0 & 3.154 \end{bmatrix} \times & \begin{bmatrix} -.05 & .999 \\ .999 & .05 \end{bmatrix} = & \begin{bmatrix} 3 & -2 \\ 1 & 5 \end{bmatrix} \end{array}$$

Singular Value Decomposition (SVD)

- In general, if **A** is *m* x *n*, then **U** will be *m* x *m*, **Σ** will be *m* x *n*, and $\mathbf{V}^\mathbf{T}$ will be *n* x *n*.

$$\overset{U}{\begin{bmatrix} -.39 & -.92 \\ -.92 & .39 \end{bmatrix}} \times \overset{\Sigma}{\begin{bmatrix} 9.51 & 0 & 0 \\ 0 & .77 & 0 \end{bmatrix}} \times \overset{V^T}{\begin{bmatrix} -.42 & -.57 & -.70 \\ .81 & .11 & -.58 \\ .41 & -.82 & .41 \end{bmatrix}} = \overset{A}{\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}}$$

Singular Value Decomposition (SVD)

- **U** and **V** are always rotation matrices.
  - Geometric rotation may not be an applicable concept, depending on the matrix. So we call them "unitary" matrices – each column is a unit vector.
- **Σ** is a diagonal matrix
  - The number of nonzero entries = rank of **A**
  - The algorithm always sorts the entries high to low

$$\overset{U}{\begin{bmatrix} -.39 & -.92 \\ -.92 & .39 \end{bmatrix}} \times \overset{\Sigma}{\begin{bmatrix} 9.51 & 0 & 0 \\ 0 & .77 & 0 \end{bmatrix}} \times \overset{V^T}{\begin{bmatrix} -.42 & -.57 & -.70 \\ .81 & .11 & -.58 \\ .41 & -.82 & .41 \end{bmatrix}} = \overset{A}{\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}}$$

Singular Value Decomposition (SVD)

$$\mathbf{M} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{T}$$

Illustration from Wikipedia

Calculus review

Differentiation

The derivative provides us information about the rate of change of a function.

The derivative of a function is also a function.

Example:

> The derivative of the rate function is the acceleration function.

Derivative = rate of change

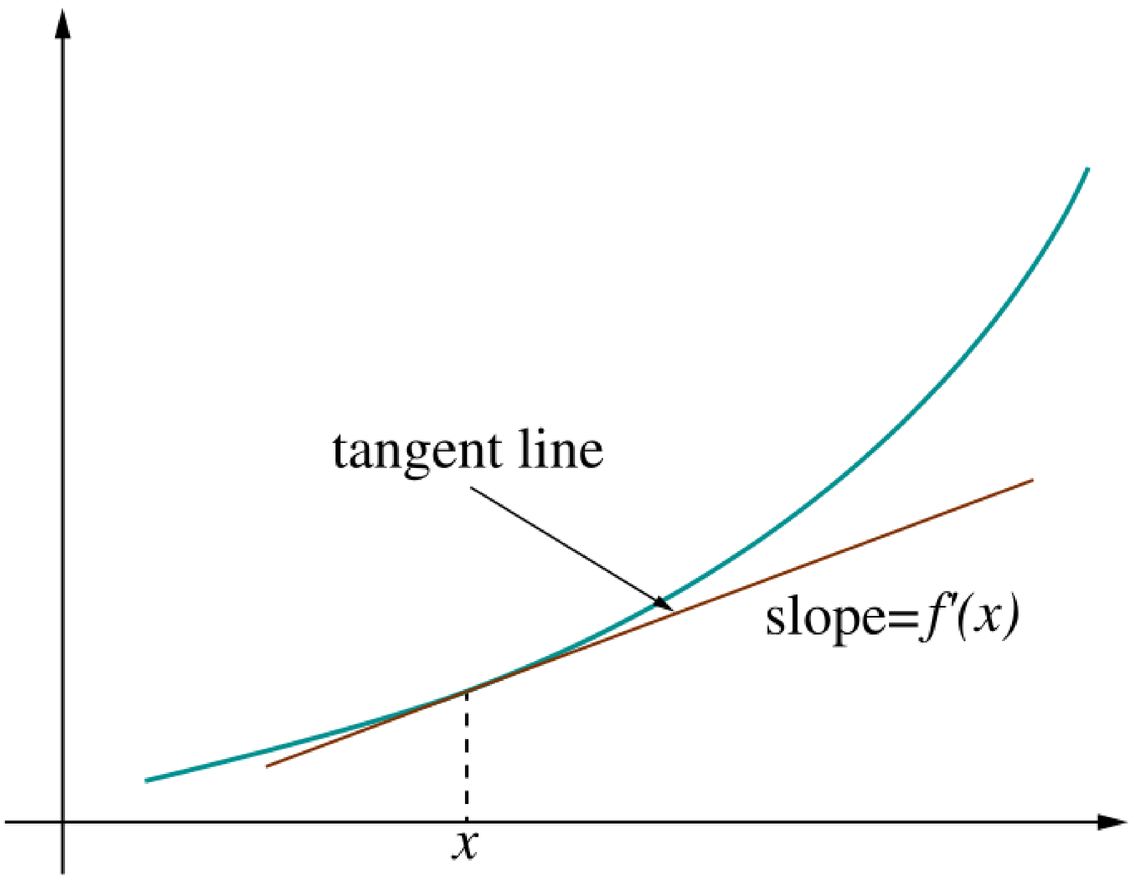


Image: Wikipedia

Derivative = rate of change

- Linear function y = mx + b
- Slope $m = \frac{\text{change in } y}{\text{change in } x} = \frac{\Delta y}{\Delta x},$

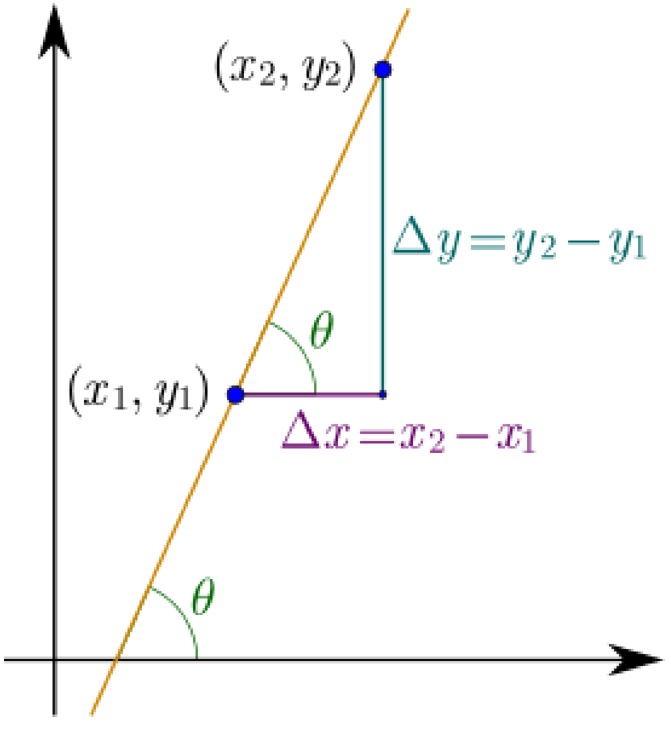


Image: Wikipedia

Ways to Write the Derivative

Given the function f(x), we can write its derivative in the following ways:

- $f'(x)$

- $\frac{d}{dx}f(x)$

The derivative of x is commonly written dx.

Differentiation Formulas

The following are common differentiation formulas:

- The derivative of a constant is 0.

$$\frac{d}{du} c = 0$$

- The derivative of a sum is the sum of the derivatives.

$$\frac{d}{du}(f(u) + g(u)) = f'(u) + g'(u)$$

Examples

- The derivative of a constant is 0.

$$\frac{d}{du}7 =$$

- The derivative of a sum is the sum of the derivatives.

$$\frac{d}{dt}(t+4) =$$

More Formulas

- The derivative of *u* to a constant power:

$$\frac{d}{du}u^n = n * u^{n-1}du$$

- The derivative of *e*:

$$\frac{d}{du}e^u = e^u du$$

- The derivative of *log*:

$$\frac{d}{du}\log(u) = \frac{1}{u}du$$

More Examples

- The derivative of *u* to a constant power:

$$\frac{d}{dx}3x^3 =$$

- The derivative of *e*:

$$\frac{d}{dy}e^{4y} =$$

- The derivative of *log*:

$$\frac{d}{dx}3\log(x) =$$

Product and Quotient

The product rule and quotient rules are commonly used in differentiation.

- Product rule:

$$\frac{d}{du}(f(u)*g(u)) = f(u)g'(u) + g(u)f'(u)$$

- Quotient rule:

$$\frac{d}{du}\left(\frac{f(u)}{g(u)}\right) = \frac{g(u)f'(u) - f(u)g'(u)}{(g(u))^2}$$

Chain Rule

The chain rule allows you to combine any of the differentiation rules we have already covered.

- First, do the derivative of the outside and then do the derivative of the inside.

$$\frac{d}{du} f(g(u)) = f'(g(u)) * g'(u) * du$$

Try These

$$f(z) = z + 11$$

$$s(y) = 4ye^{2y}$$

$$g(y) = 4y^3 + 2y$$

$$p(x) = \frac{\log(x^2)}{x}$$

$$h(x) = e^{3x}$$

$$q(z) = (e^z - z)^3$$

Solutions

$$f'(z) = 1$$

$$s'(y) = 8ye^{2y} + 4e^{2y}$$

$$g'(y) = 12y^2 + 2$$

$$p'(x) = \frac{2 - \log(x^2)}{x^2}$$

$$h'(x) = 3e^{3x}$$

$$q'(z) = 3(e^z - z)^2(e^z - 1)$$

Matlab

Matlab tutorial

http://www.cs.pitt.edu/~kovashka/cs1675_fa18/tutorial.m
http://www.cs.pitt.edu/~kovashka/cs1675_fa18/myfunction.m
http://www.cs.pitt.edu/~kovashka/cs1675_fa18/myotherfunction.m

Please cover whatever we don’t finish at home.

Other tutorials and exercises

- https://people.cs.pitt.edu/~milos/courses/cs2750/Tutorial/
- http://www.math.udel.edu/~braun/M349/Matlab_probs2.pdf
- http://www.facstaff.bucknell.edu/maneval/help211/basicexercises.html
  - Do Problems 1-8, 12
  - Most also have solutions
  - Ask the TA if you have any problems