CS 591, Lecture 1
Data Analytics: Theory and Applications
Boston University

Charalampos E. Tsourakakis

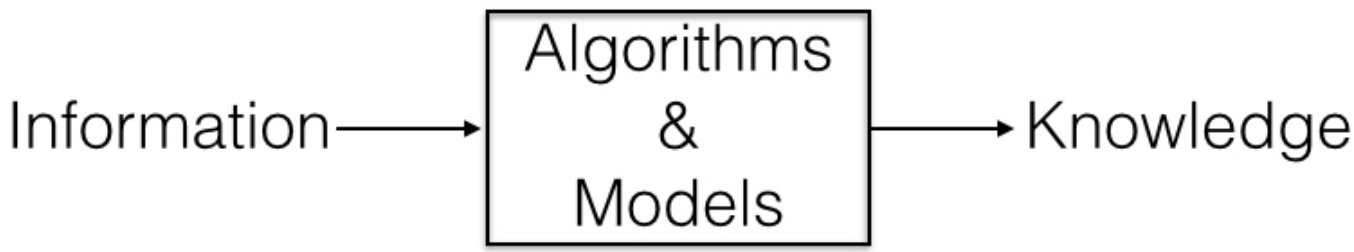
January 23rd, 2017


WELCOME TO CS 591!
imgflip.com

Welcome to CS 591!

"We are drowning in information and starving for knowledge"

John Naisbitt



Two types of techniques

- Traditional CS techniques where explicit instructions are coded

- Machine learning

  Study of algorithms that

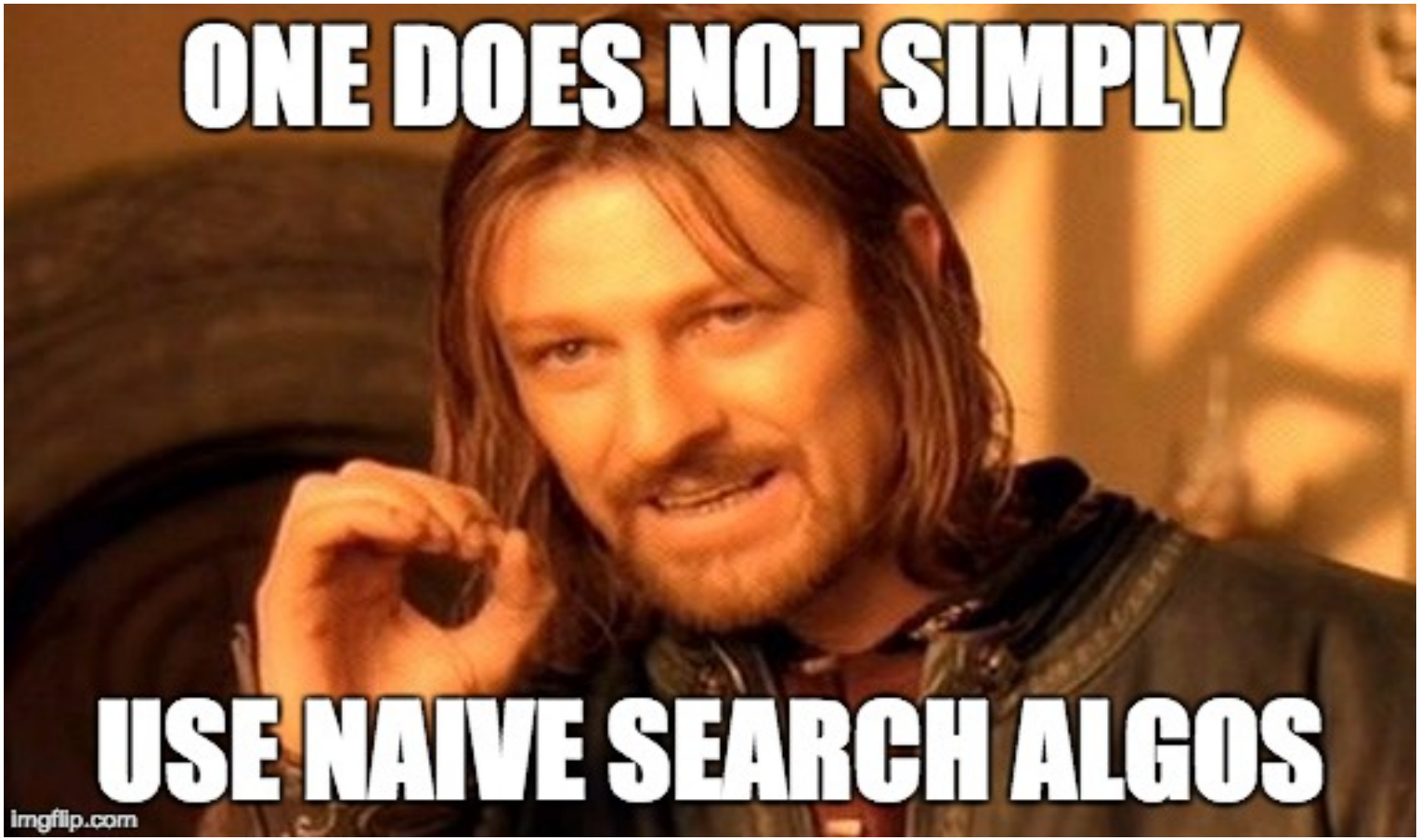
(a) improve their performance

(b) at some task

(c) with experience.

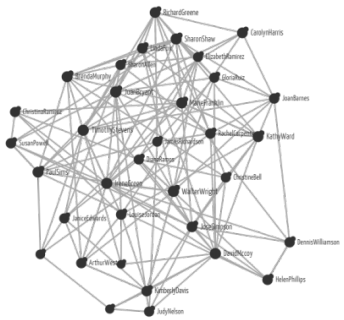
Searching massive volumes of data

Searching massive volumes of data

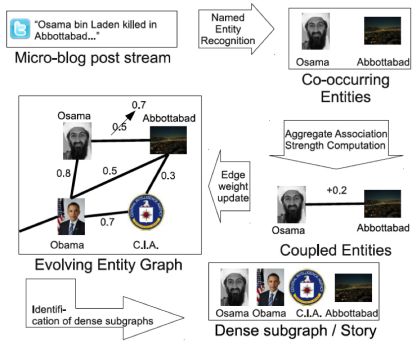


Graph mining

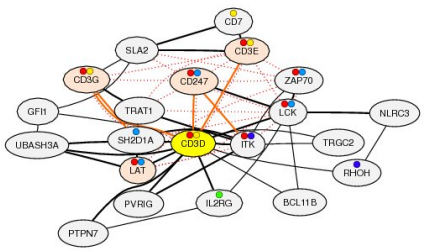
Who-calls-whom



Twitter



Genes



	Security	Social Media	CompBio
V	Humans	Entities	Genes
E	Phone call	Co-occurence	Correlation
Dense subgraph	Anomaly	Thematic coherence	Functionally related

Data streams

- Data streams
- Search engine queries

  How many distinct Google queries were performed during the last 21 days?
- Tweets
- GPS location data
- Sensor networks
- Bank transactions
- Facebook friendships
- ...

Face detection

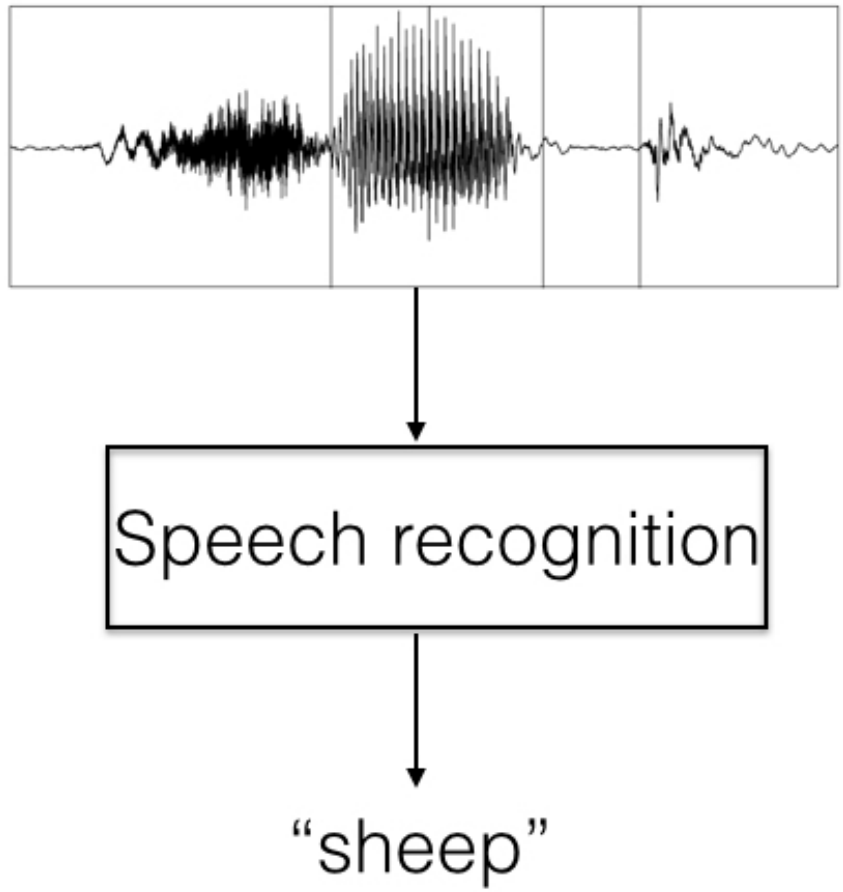
- Sliding window detector

  [Felzenszwalb et al., 2010]

  [Viola and Jones, 2004]

  [Dalal and Triggs, 2005]

Speech recognition

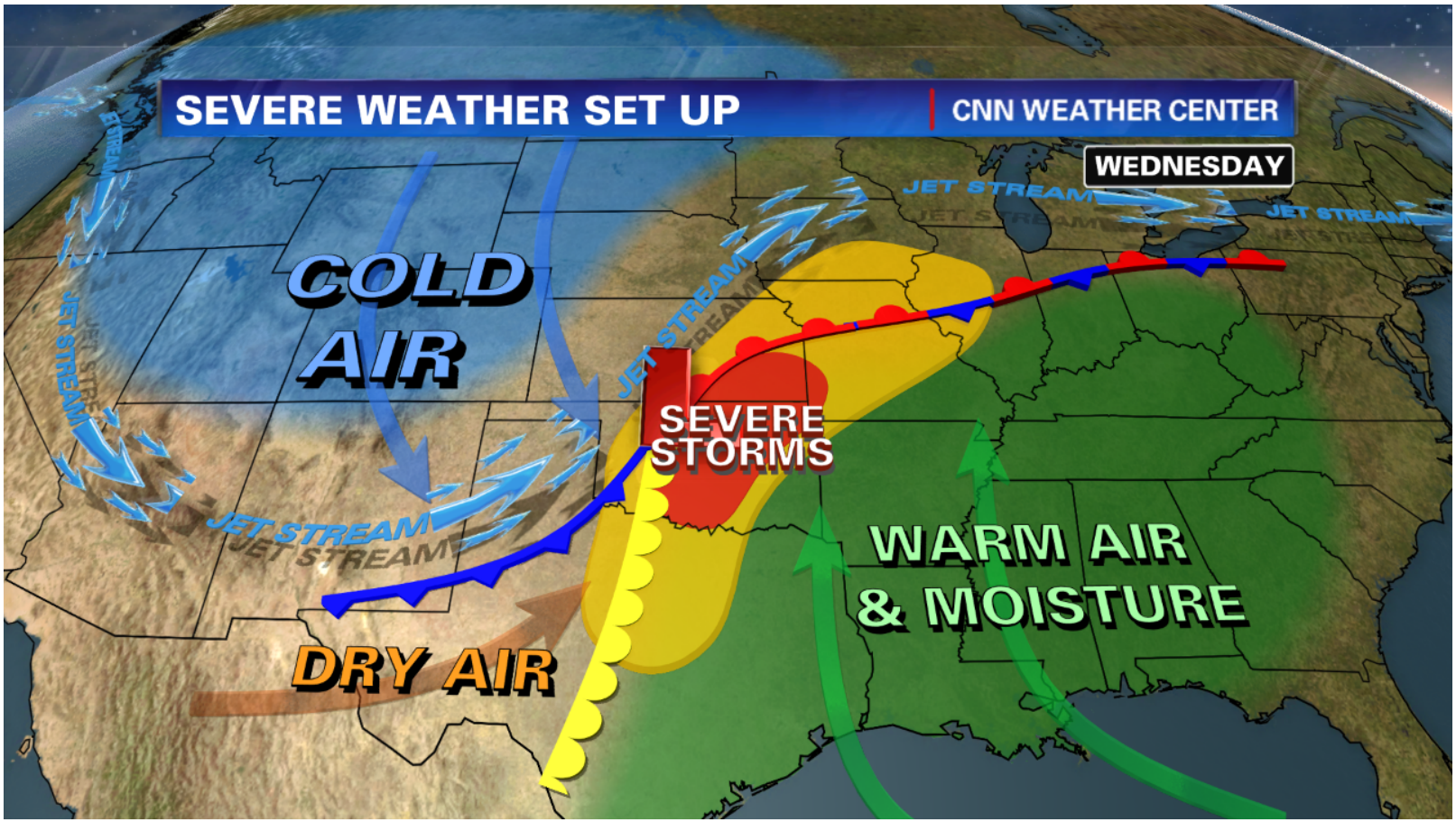


Source: [Mumford and Desolneux, 2010]

Video classification

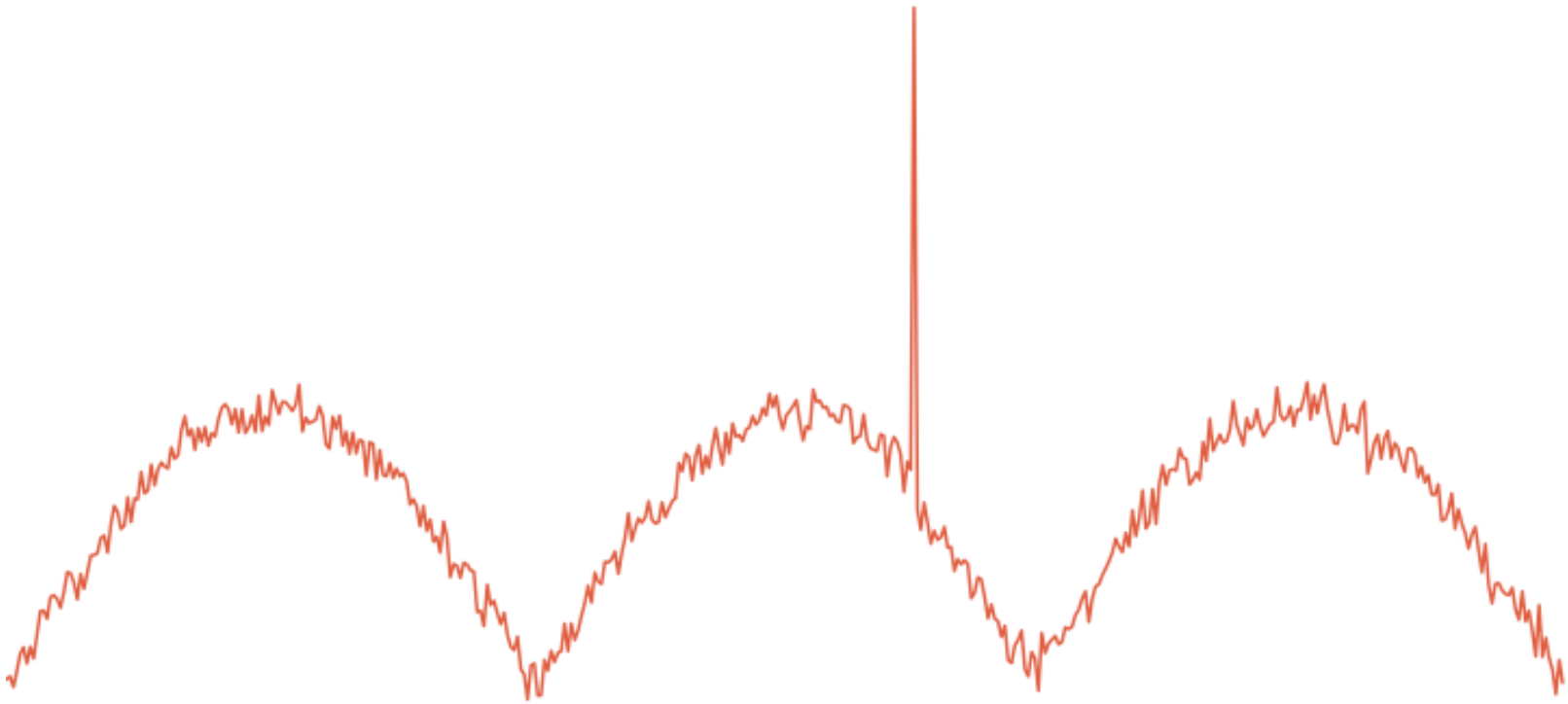
Is it a music or a fitness video?

Weather forecast



How do we predict the weather?

Early on anomaly detection



Can we predict time series anomalies early on?

Netflix Prize



	?	?	4
	?	4	2
	5	?	4
	?	5	?

Example of movie-rating data.

Scene completion

http://graphics.cs.cmu.edu/projects/scene-completion/

AlphaGo

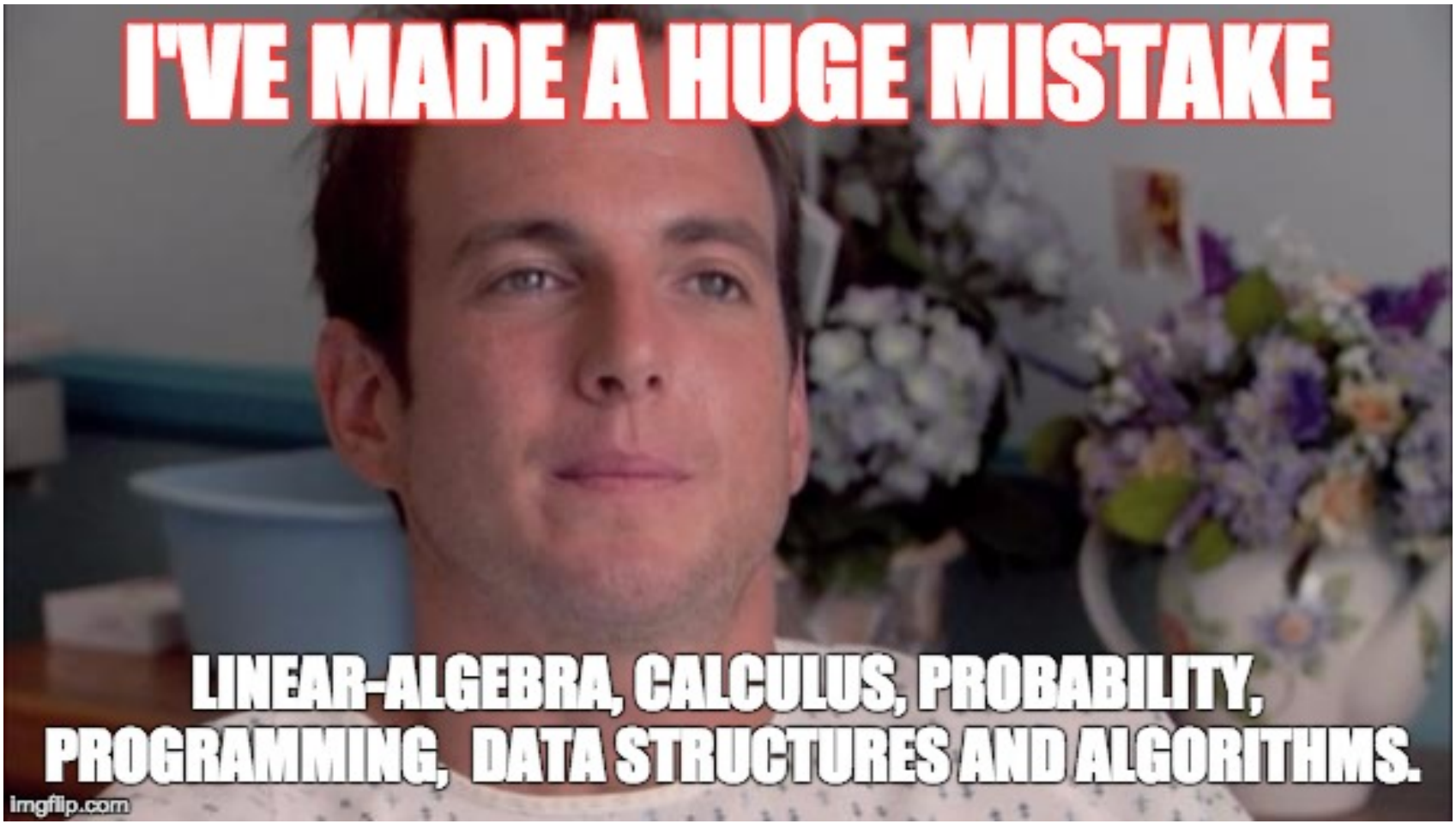
https://deepmind.com/research/alphago/

Self-driving cars

Logistics

Prerequisites

Don't be this person!



Grading

Registered students

- **Scribe** (10%, to be returned within **2** days)
- **Final exam** (20%)
- **Project** (70%) (**demanding!**)
- **Readings** is part of the class syllabus
- **Additional material** extends further what we discuss in the class

In order to **audit**:

- Two problems from the 1st part of the project
- Read a paper related to one of the class topics, and write a report

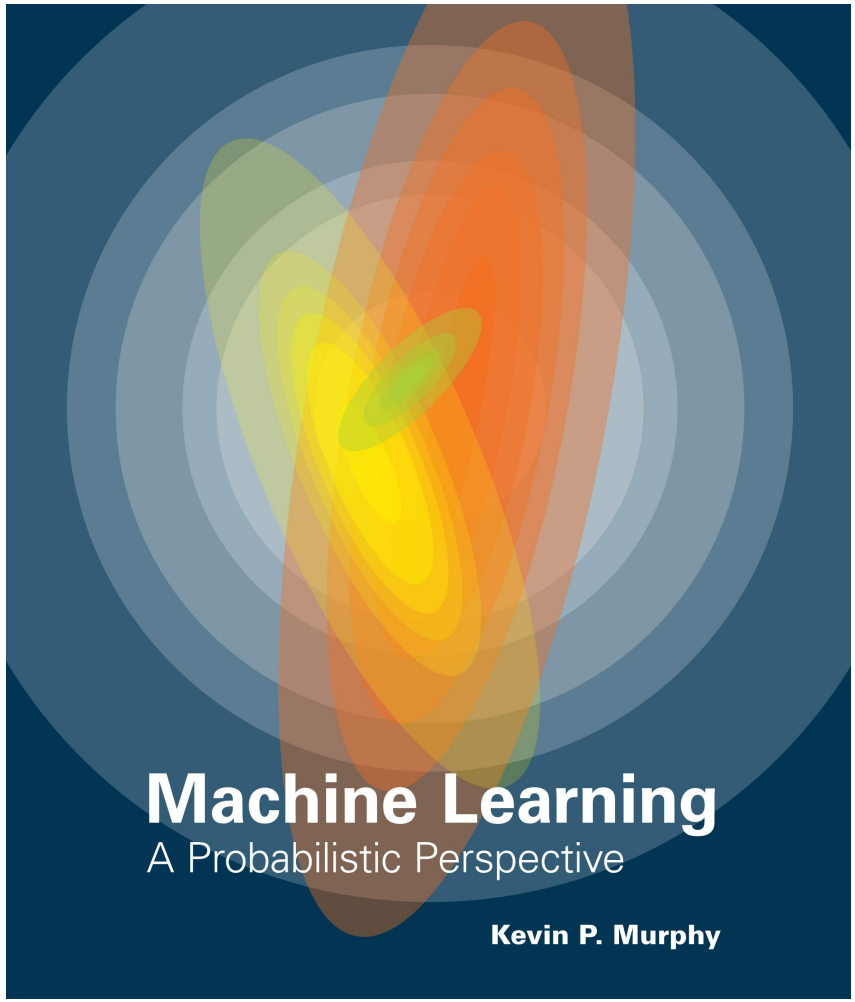
Project

- Will be out at the end of the 3rd week of classes on the class Web page
- Two parts: exercises (some involve programming), and an independent project on a topic of your choice.
- **Milestone:**
  - Probability and Hashing problems (problems 1,2)
  - Two page report on second part of the project (problem description, related work, project goals, preliminary experiments)
- For the second part you will be evaluated based on:
  - a written report (**LaTeX**),
  - your code,
  - a presentation (tentative date May 1st).
- Feel free to form groups up to 3 people for part 2.

Start early on!

Textbook

Highly recommended but **not** required:



More information

- Office hours

  Each Monday from 4.30 to 6.00.

  MCS 292

- Class web page

  `http://people.seas.harvard.edu/~babis/cs591_bu_sp17.html`

- Google group

  `https://groups.google.com/forum/#!forum/cs591_bu_spring17/`

Probability Recap with Applications on Image Restoration

Bayes' theorem

Bayes' theorem (aka Bayes' Law and Bayes' ule) is a direct application of conditional probabilities.

$$\mathbf{Pr}[H|D] = \frac{\mathbf{Pr}[D|H]\mathbf{Pr}[H]}{\mathbf{Pr}[D]}, \text{and } \mathbf{Pr}[D] > 0, \text{or } ...$$

$$\text{posterior} \propto \text{likelihood} \times \text{prior}.$$

Kahneman Tversky – Taxi accident


- A cab was involved in a hit and run accident at night
- 85% Green, 15% Blue
- A witness identified the cab as Blue
- The witness correctly identified each one of the two colors 80% of the time.

  What is the probability that the cab involved in the accident was Blue rather than Green?

Apply Bayes' theorem: whiteboard.

Bayesian Odds

$$\frac{\mathbf{Pr}\left[H_1|D\right]}{\mathbf{Pr}\left[H_2|D\right]} = \frac{\mathbf{Pr}\left[D|H_1\right]}{\mathbf{Pr}\left[D|H_2\right]} \frac{\mathbf{Pr}\left[H_1\right]}{\mathbf{Pr}\left[H_2\right]}$$

Now, let's define the txtredevents:

- Let G be the event of the delinquent being Green.
- Let B be the event of the delinquent being Blue.
- Finally, let W be the witness' report.

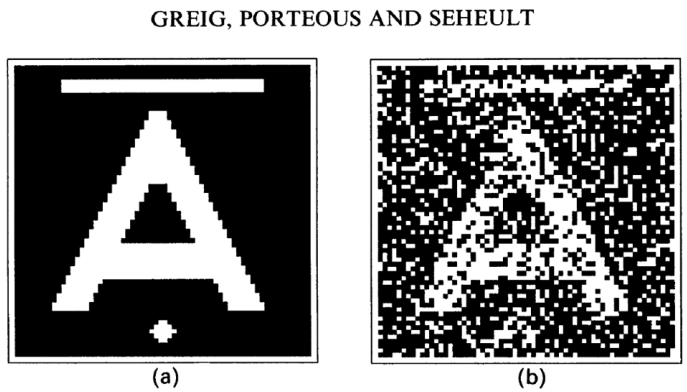
$$\frac{\mathbf{Pr}\left[B|W\right]}{\mathbf{Pr}\left[G|W\right]} = \frac{0.8}{0.2} \frac{0.15}{0.85} = \frac{12}{17}.$$

Since $\mathbf{Pr}\left[B|W\right] + \mathbf{Pr}\left[G|W\right] = 1$, we obtain that

$$\mathbf{Pr}\left[B|W\right] = 0.41 < \mathbf{Pr}\left[G|W\right]!$$

Exact MAP Estimation for Binary Images

Problem: Given (b) can we infer (a)? In other words, can we **restore** the image from its corrupted-by-noise version?



How to formulate the problem? Any ideas?

Exact MAP Estimation for Binary Images

Let's be Bayesian!

$x = (x_1, \ldots, x_n)$ the original image (shown in (a))

$y = (y_1, \ldots, y_n)$ the observed corrupted image (e.g., the one shown in (b))

Assumption: The records $y_1, \ldots, y_n$ are conditionally independent given x, and each has known conditional density $f(y_i|x_i)$ that depends only on x_i.

By Bayes' theorem:

$$p(x|y) \propto \underbrace{p(y|x)}_{\text{likelihood:how do we compute it?}} \times \underbrace{p(x)}_{\text{prior:what is a good prior?}}$$

Goal: output

$$\boxed{x^* = \arg\max p(x|y)}$$

Likelihood and Prior

Given our assumption, the likelihood function is

$$p(y|x) = \prod_{i=1}^{n} f(y_i|1)^{x_i} f(y_i|0)^{1-x_i}.$$

- What kind of patterns would we like the prior to enforce?

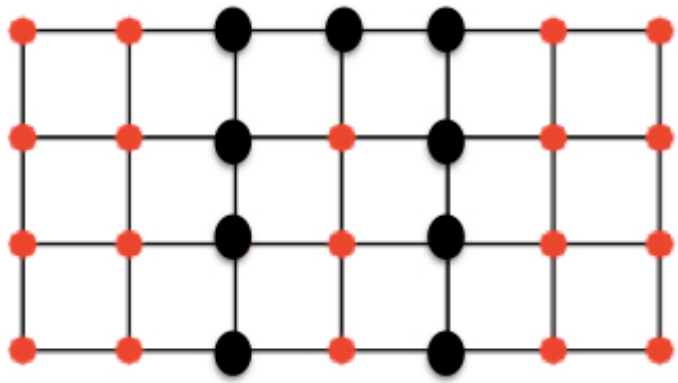
- Let's imagine how these characters would look on a binary image:

  a,b,c,f,y,x,z,1,&,$,@

  Homogeneous patches that occasionally change discontinuously

Prior $p(x)$

$$p(x) \propto \exp\left\{\frac{1}{2}\sum_{i \neq j} \beta_{ij}\left(x_i x_j + (1-x_i)(1-x_j)\right)\right\}$$



- For an edge (u, v) where $val(u) = val(v)$

$$x_u x_v + (1-x_v)(1-x_u) = x_u^2 + (1-x_u)^2 = 1.$$

- On the contrary for an edge where $val(u) \neq val(v)$

$$x_u(1-x_u) + (1-x_u)x_u = 0.$$

Exact MAP Estimation for Binary Images

Our MAP inference becomes equivalent to minimizing (details on whiteboard)

$$\sum_{i=1}^{n} x_i \max(0, -\lambda_i) + \sum_{i=1}^{n} \max(0, \lambda_i)(1 - x_i) + \frac{1}{2} \sum_{i \sim j} \beta_{ij} (x_i - x_j)^2,$$

where $\lambda_i = \frac{f(y_i|1)}{f(y_i|0)}$.

Let's rephrase this problem. Suppose $b_{ij} = b$ for all neighboring nodes for simplicity.

Exact MAP Estimation for Binary Images

- We have a $n \times m$ binary matrix
- We impose a grid structure
- We call two neighboring nodes bad if they have different values. We pay K units for each such pair.
- We are allowed to flip the value of any node, but we have to pay R units.
- The total cost is the sum of these two terms. How do we find the best assignment of values to nodes?

Any ideas? Is it NP-hard, poly-time solvable?

Exact MAP Estimation for Binary Images

Max flow problem!

- Source s, sink t
- Arc of capacity R from s to each node u with value 0.
- Arc of capacity R from each u node with value 1 to sink t.
- Directed arcs from each node u to its neighbors with capacity K.

Details on whiteboard.

Reading: [Kolmogorov and Zabin, 2004].

references I

Dalal, N. and Triggs, B. (2005).

Histograms of oriented gradients for human detection.

In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893. IEEE.

Felzenszwalb, P. F., Girshick, R. B., McAllester, D., and Ramanan, D. (2010).

Object detection with discriminatively trained part-based models.

IEEE transactions on pattern analysis and machine intelligence, 32(9):1627–1645.

Kolmogorov, V. and Zabin, R. (2004).

What energy functions can be minimized via graph cuts?

IEEE transactions on pattern analysis and machine intelligence, 26(2):147–159.

references II

Mumford, D. and Desolneux, A. (2010).

Pattern theory: the stochastic analysis of real-world signals.

CRC Press.

Viola, P. and Jones, M. J. (2004).

Robust real-time face detection.

International journal of computer vision, 57(2):137–154.