ECE 6504: Advanced Topics in Machine Learning

Probabilistic Graphical Models and Large-Scale Learning

Dhruv Batra

Virginia Tech

What is this class about?

Some of the most exciting developments in Machine Learning, AI, Statistics & related fields in the last 3 decades

First Caveat

- This is an ADVANCED Machine Learning class
  - This should not be your first introduction to ML
  - You will need a formal class; not just self-reading/coursera
  - If you took ECE 4984/5984, you're in the right place
  - If you took ECE 5524 or equivalent, see list of topics taught in ECE 4984/5984.

Topics Covered in Intro to ML&P

- Basics of Statistical Learning
  - Loss function, MLE, MAP, Bayesian estimation, bias-variance tradeoff, overfitting, regularization, cross-validation
- Supervised Learning
  - Naïve Bayes, Logistic Regression, Nearest Neighbour, Neural Networks, Support Vector Machines, Kernels
  - Ensemble Methods: Bagging, Boosting
- Unsupervised Learning
  - Clustering: k-means, Gaussian mixture models, EM
  - Dimensionality reduction: PCA, SVD, LDA
- Perception
  - Applications to Vision, Natural Language Processing

What is this class about?

- Making global predictions from local observations
- Learning such models from large quantities of data

Exciting Developments

- Probabilistic Graphical Models
  - Directed: Bayesian Networks (Bayes Nets)
  - Undirected: Markov/Conditional Random Fields
  - Structured Prediction
- Large-Scale Learning
  - Online learning
  - Distributed learning
- Deep Learning
  - Convolutional Nets
  - Distributed backprop
  - Dropout

Not covered in this class

What is Machine Learning?

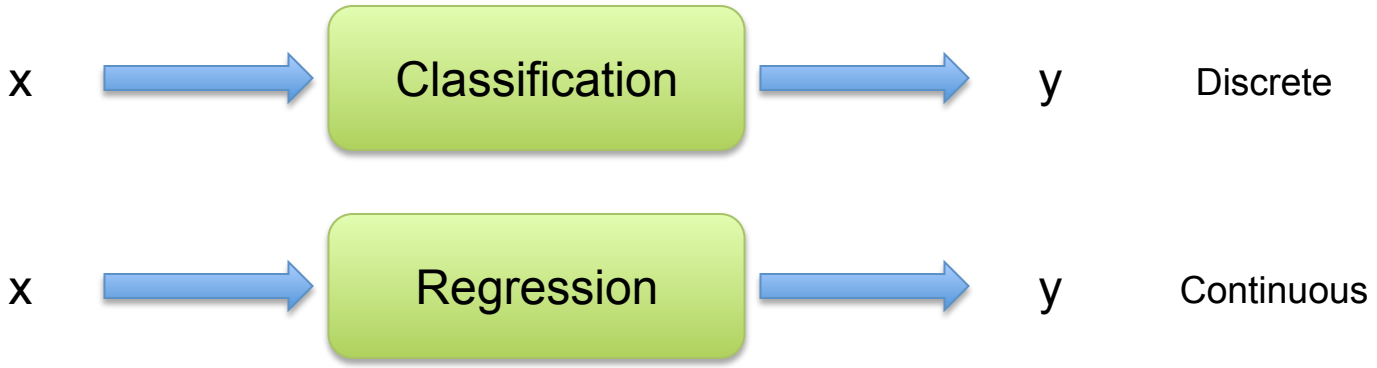
- What is learning?

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

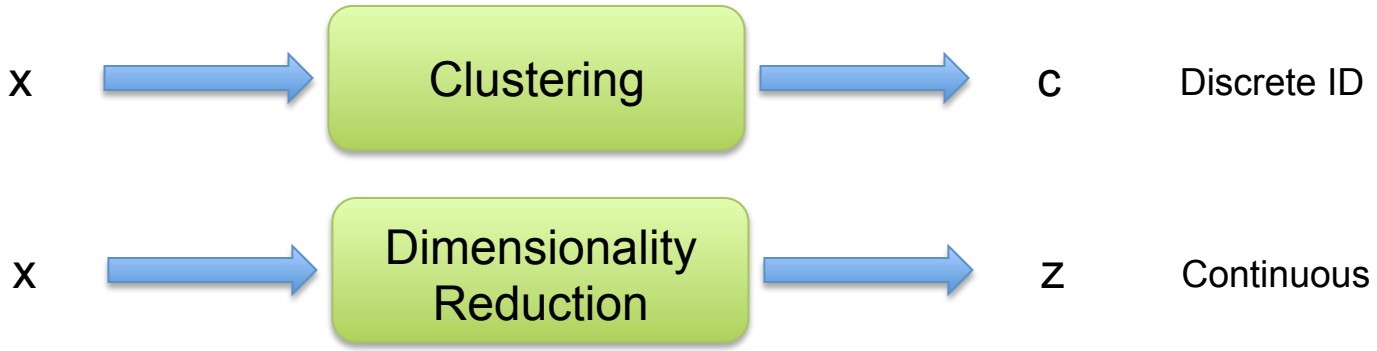
- [Tom Mitchell] algorithms that
  - improve their performance (P)
  - at some task (T)
  - with experience (E)

Tasks

Supervised Learning



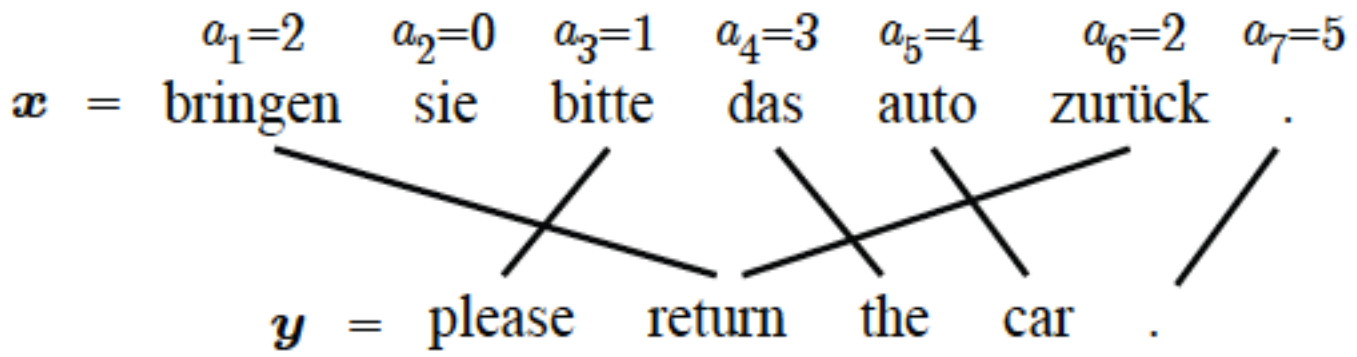
Unsupervised Learning



Classification

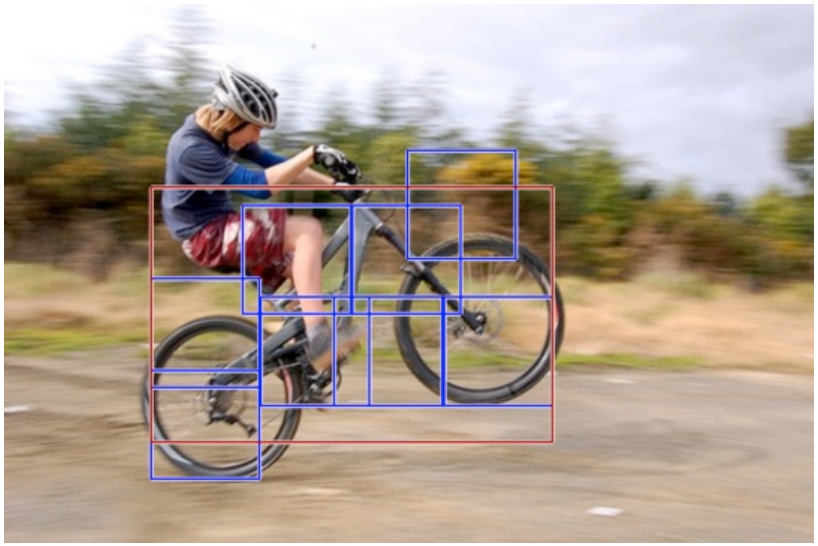
Speech Recognition

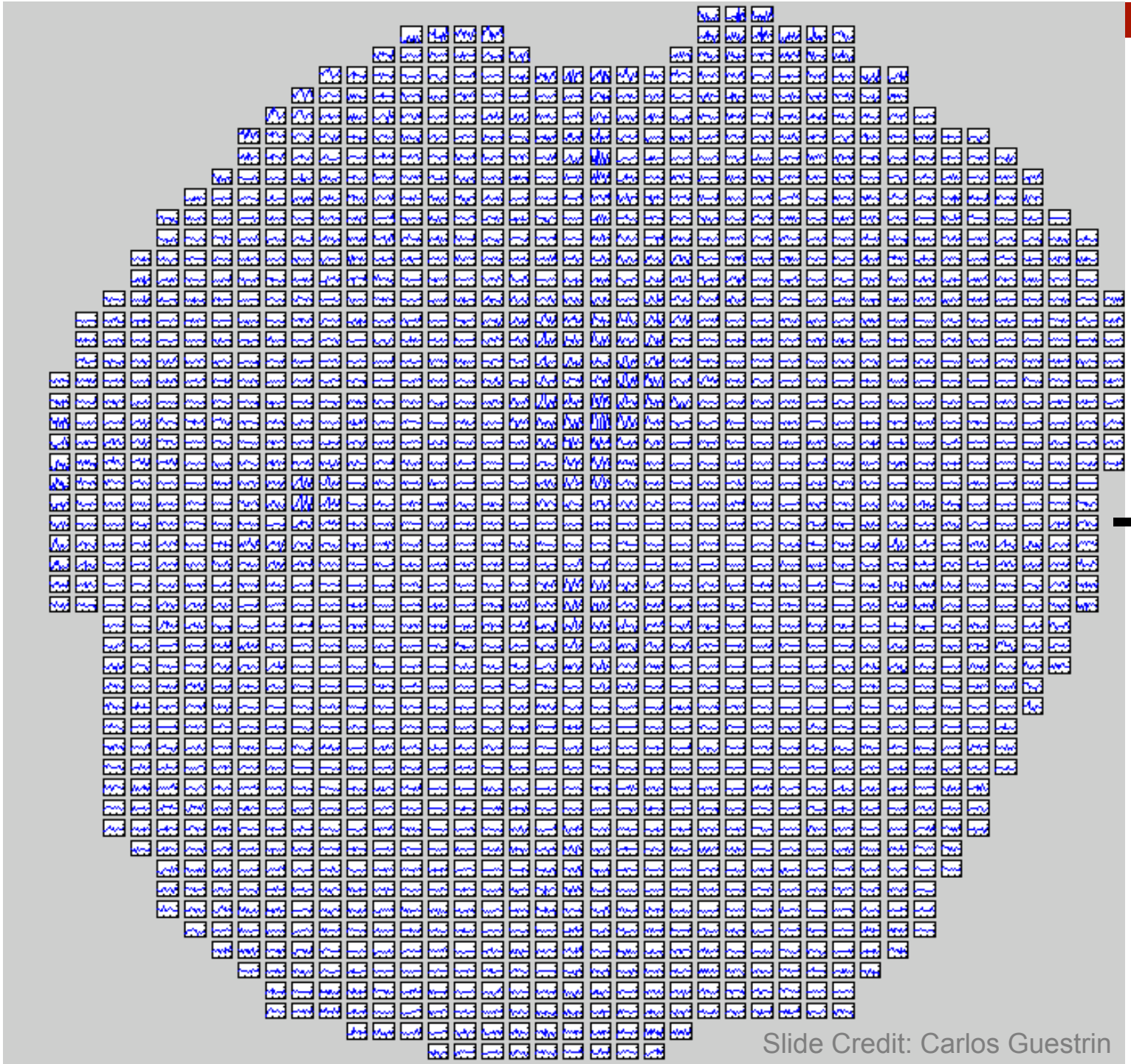
Machine Translation



Object/Face detection

- Many new digital cameras now detect faces
  - Canon, Sony, Fuji, …





Reading a noun (vs verb)

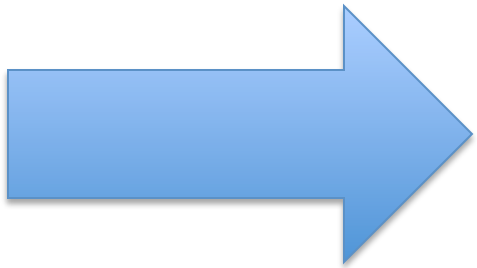
[Rustandi et al., 2005]

Slide Credit: Carlos Guestrin

Regression

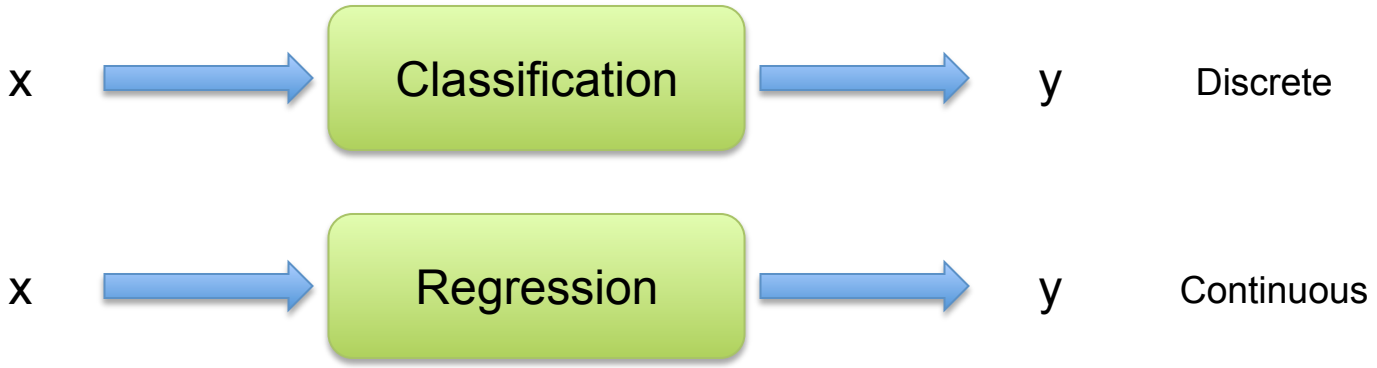
Stock market

Weather Prediction

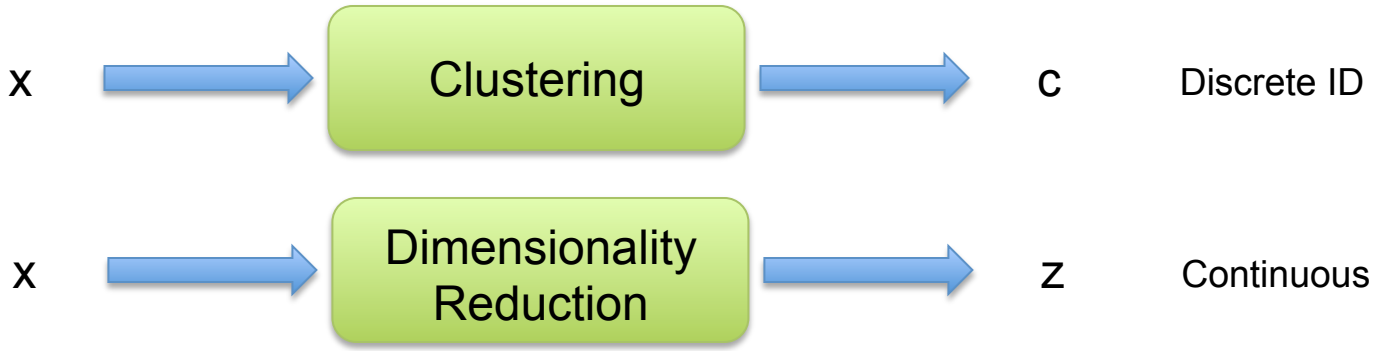


Tasks

Supervised Learning



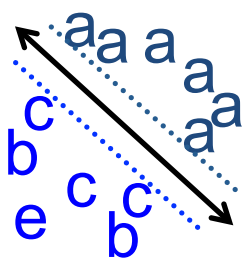
Unsupervised Learning

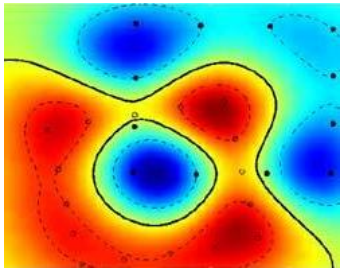


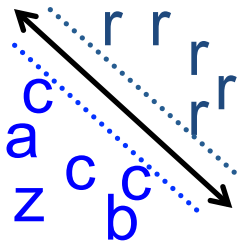
Need for Joint Prediction

Handwriting recognition

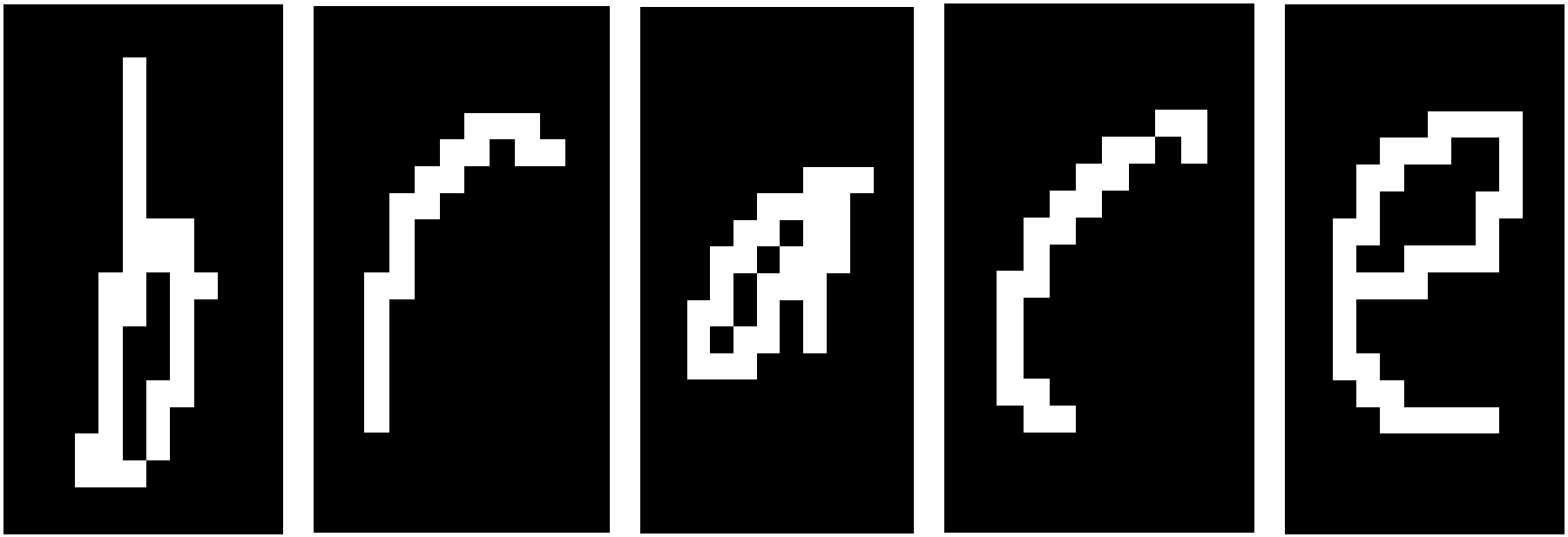
Character recognition, e.g., kernel SVMs







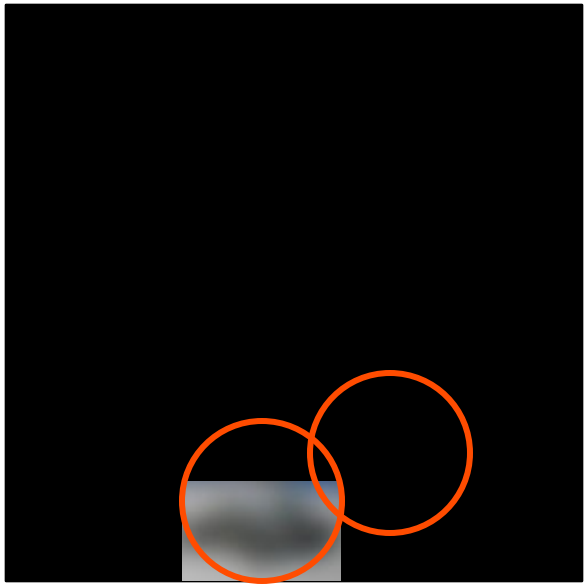
Handwriting recognition 2

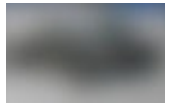


Local Ambiguity

[Smyth et al., 1994]

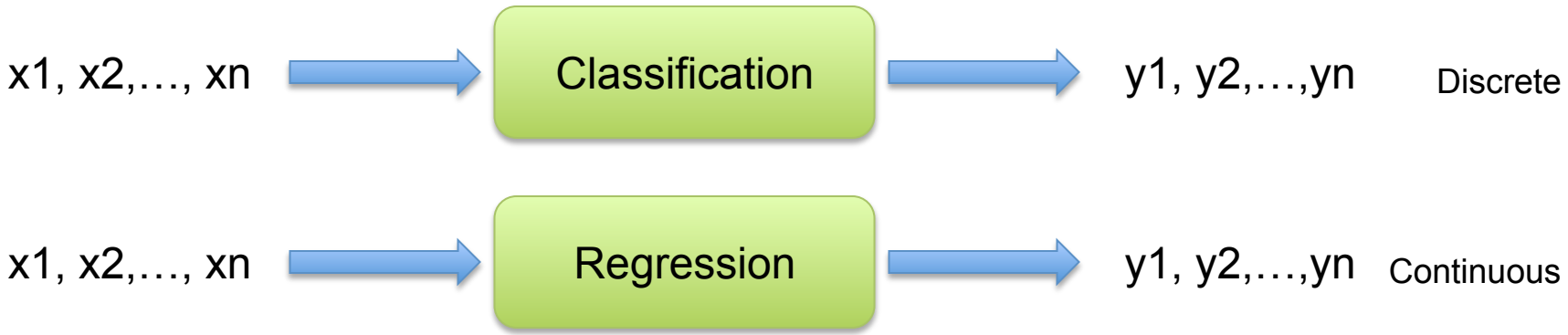
Local Ambiguity





slide credit: Fei-Fei Li, Rob Fergus & Antonio Torralba

Joint Prediction



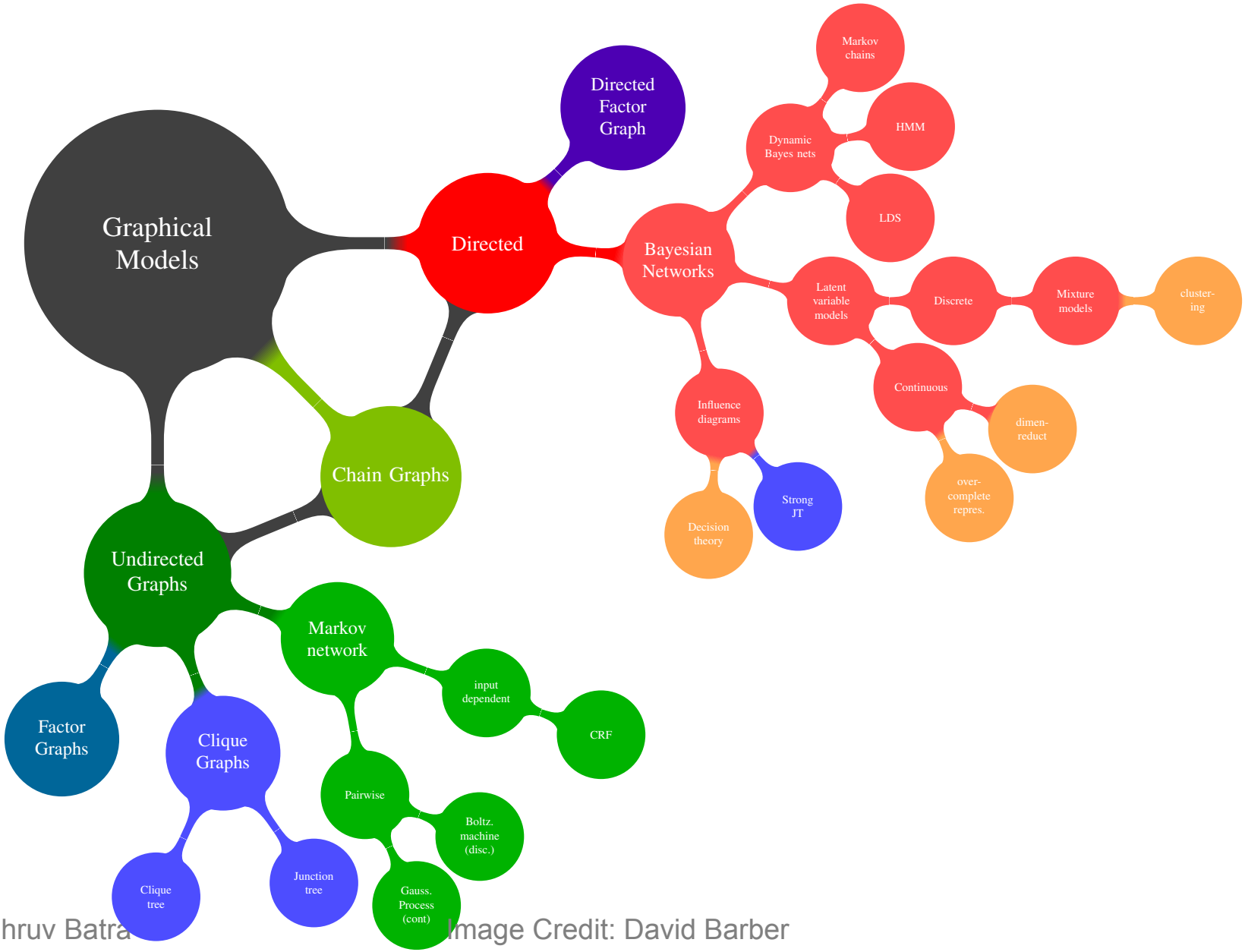
How many parameters?

- $P(X_1, X_2, \ldots, X_n)$
- Each X_i takes k states
- What if all X_i are independent?

Probabilistic Graphical Models

- One of the most exciting advancements in statistical AI in the last 10-20 years
- Marriage
  - Graph Theory + Probability
- Compact representation for exponentially-large probability distributions
  - Exploit conditional independencies
- Generalize
  - naïve Bayes
  - logistic regression
  - Many more …

Types of PGMs



Main Issues in PGMs

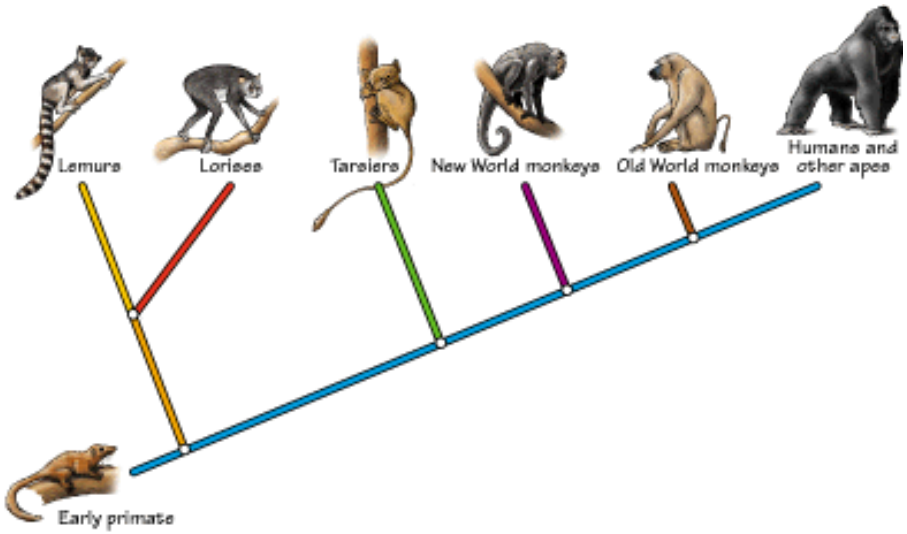
- Representation
  - How do we store $P(X_1, X_2, \ldots, X_n)$
  - What does my model mean/imply/assume? (Semantics)
- Inference
  - How do I answer questions/queries with my model? such as
  - Marginal Estimation: $P(X_5 \mid X_1, X_4)$
  - Most Probable Explanation: argmax $P(X_1, X_2, \ldots, X_n)$
- Learning
  - How do we learn parameters and structure of $P(X_1, X_2, \ldots, X_n)$ from data?
  - What model is the right for my data?

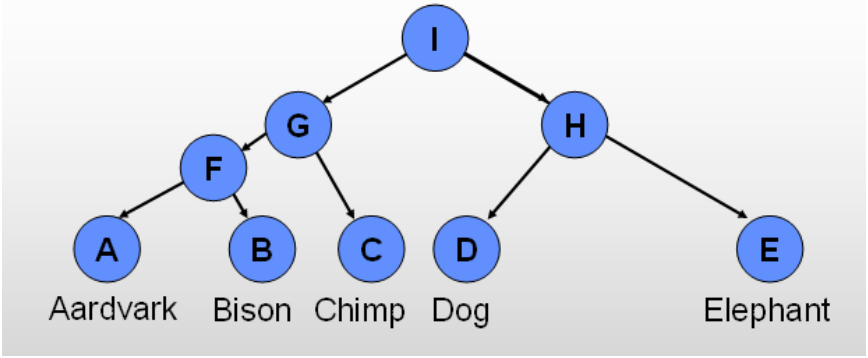
Key Ingredient

- Exploit independence assumptions
  - Encoded in the graph structure

- Structured Prediction vs Unstructured Prediction

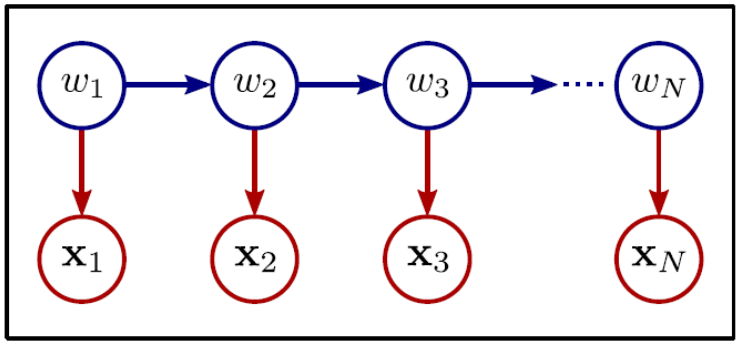
Application: Evolutionary Biology





[Friedman et al.]

Application: Computer Vision



Chain model

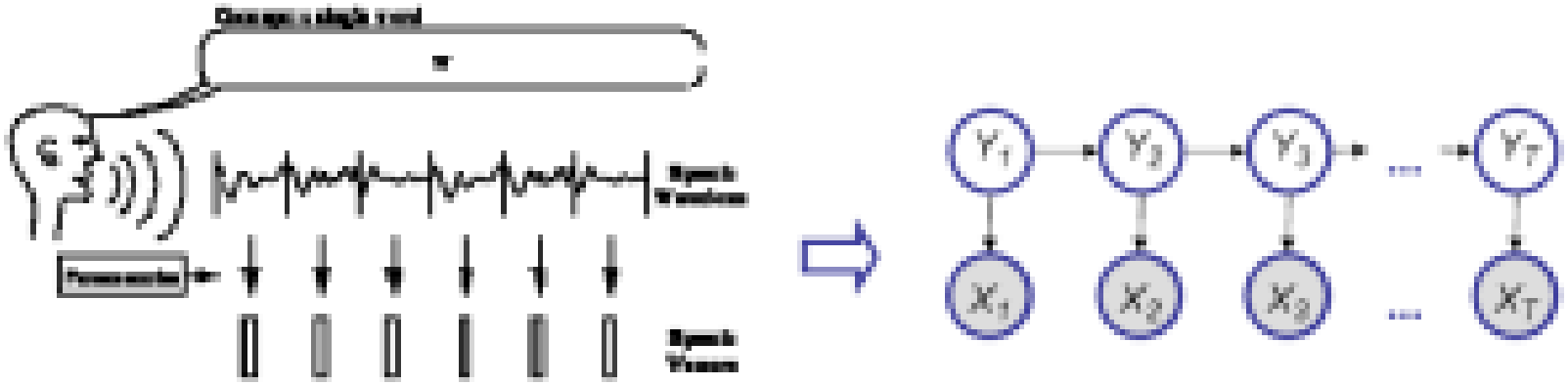
(hidden Markov model)

Interpreting sign

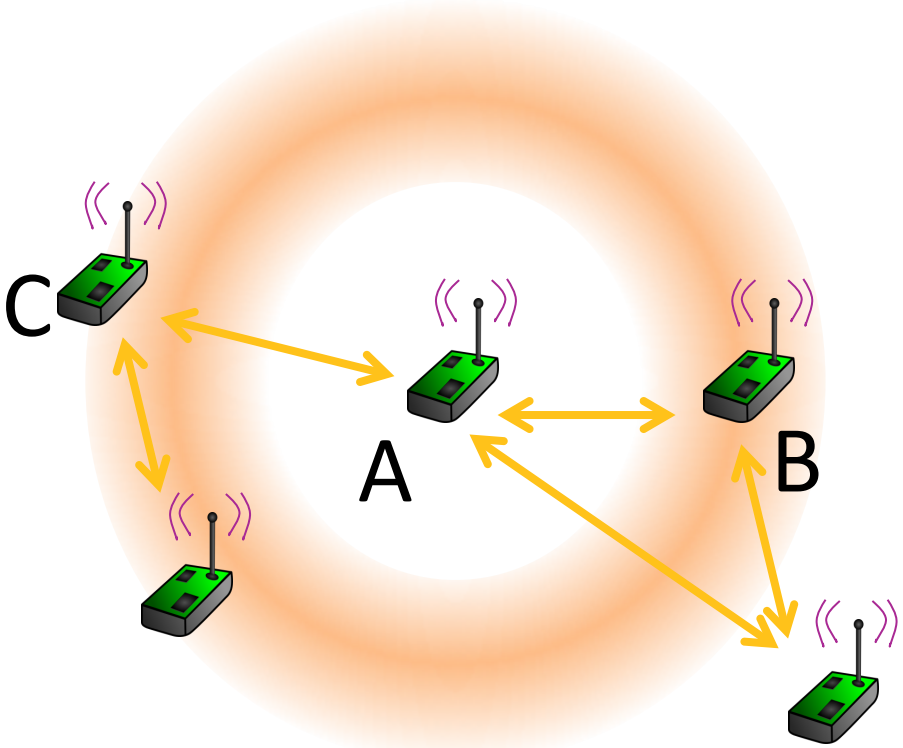
language sequences

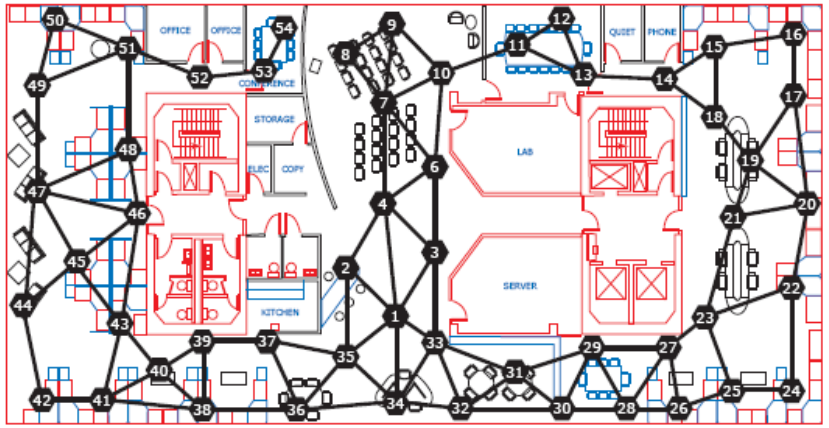
Image Credit: Simon JD Prince

Application: Speech



Application: Sensor Network





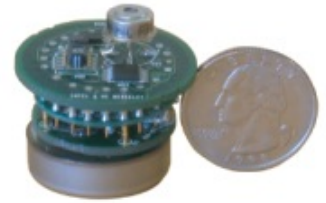


Image Credit: Carlos Guestrin & Erik Sudderth

Application: Medical Diagnosis

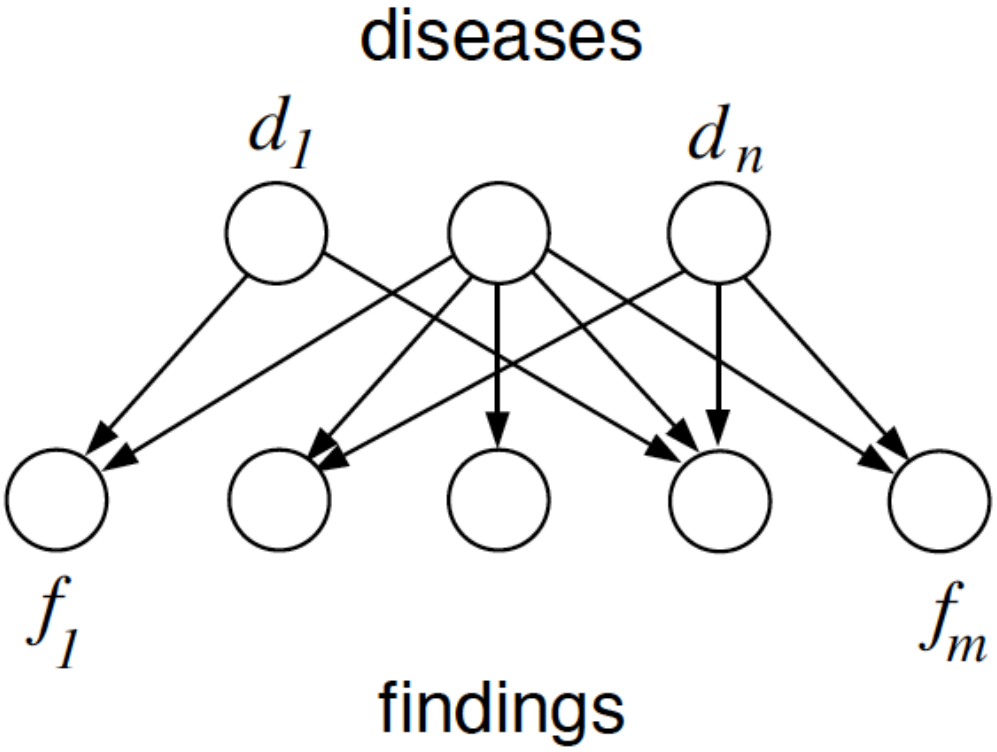
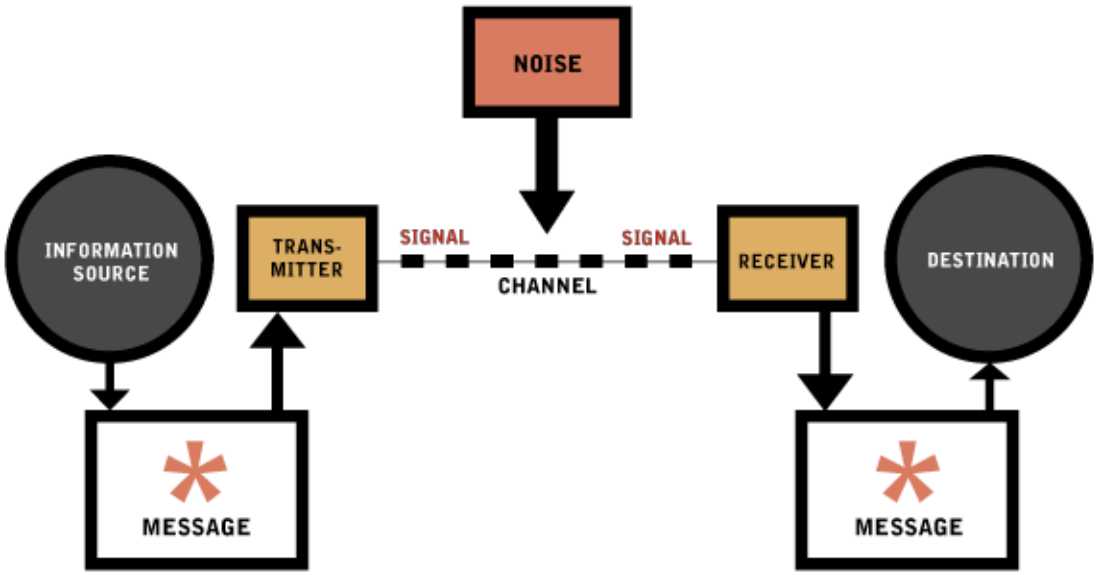
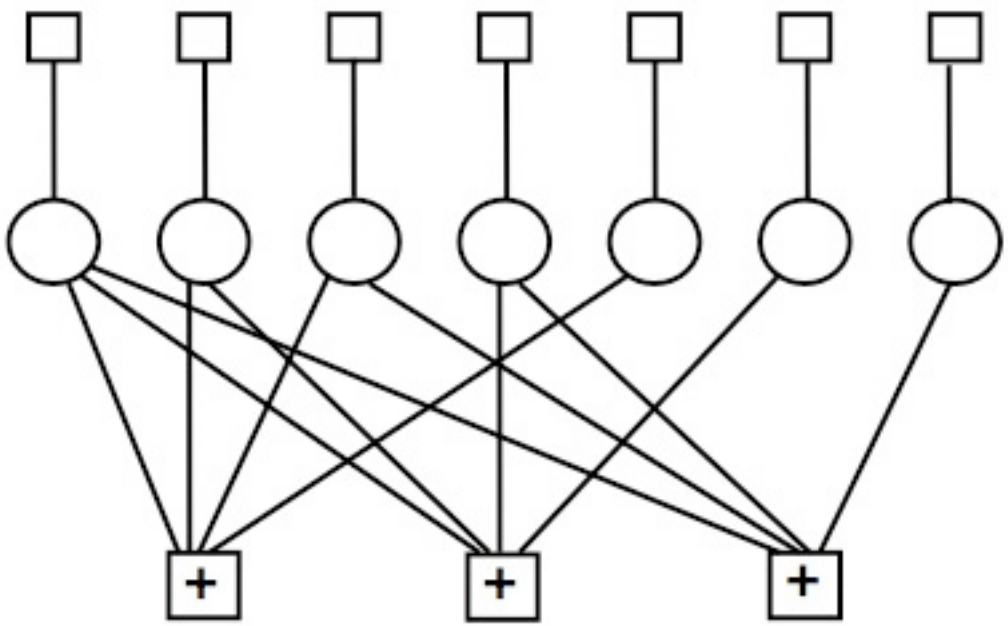


Image Credit: Erik Sudderth

Application: Coding



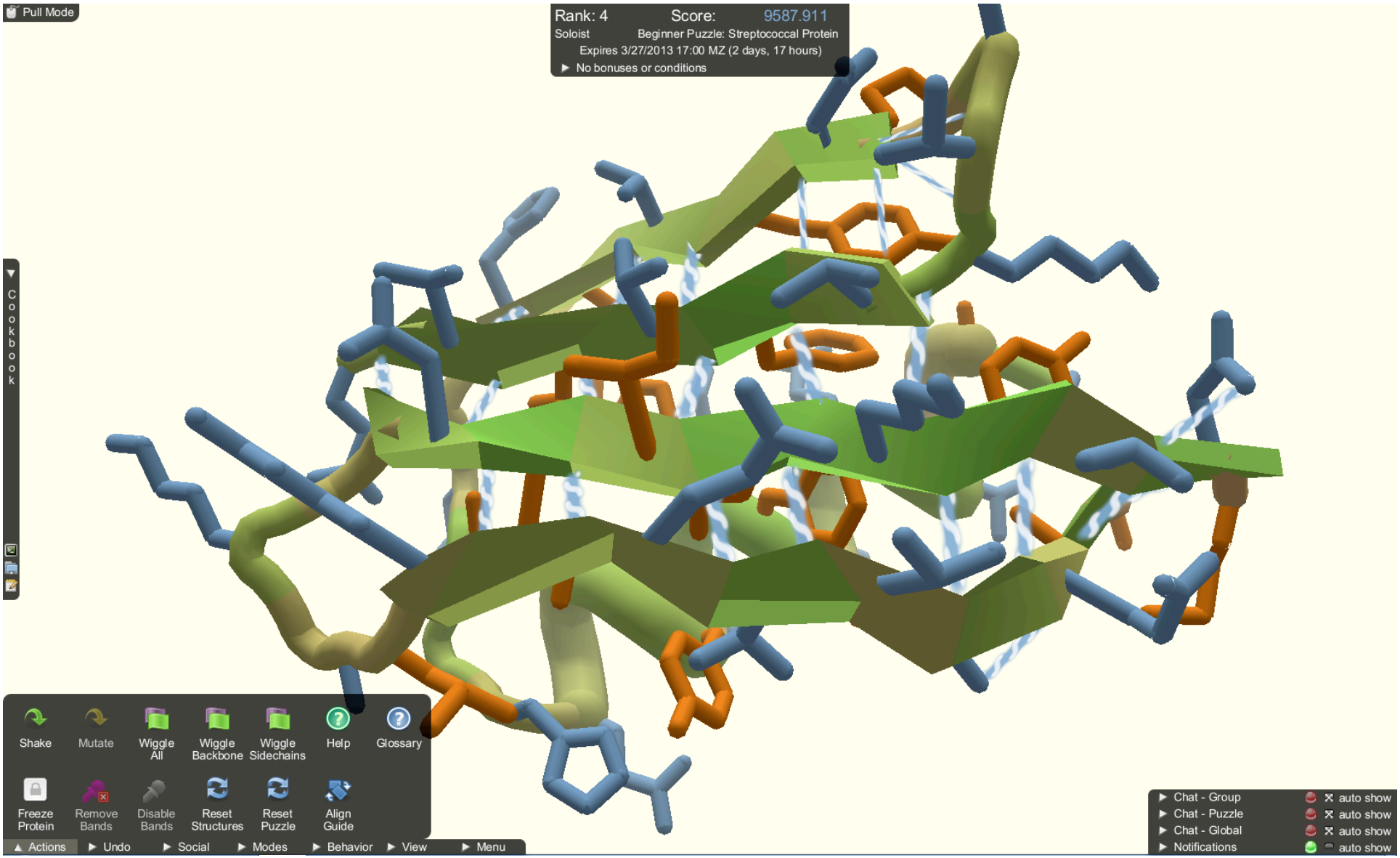


Observed Bits

True Bits

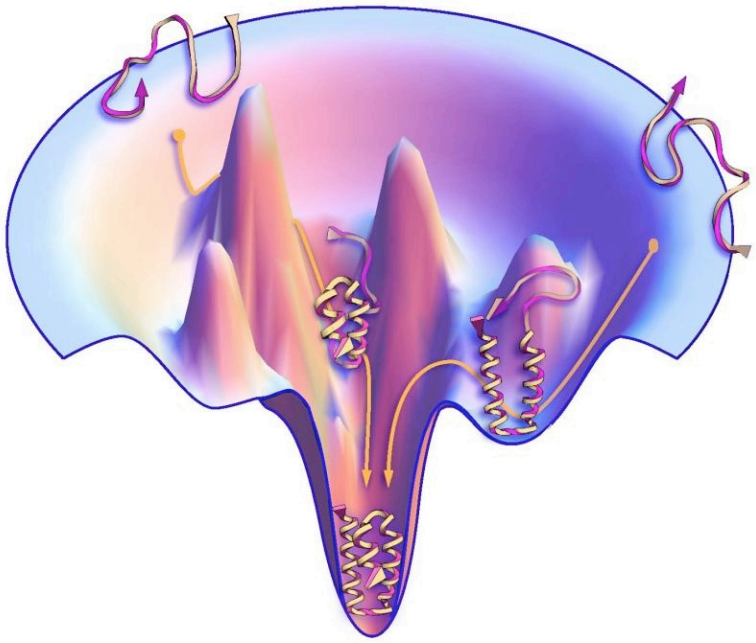
Parity Constraints

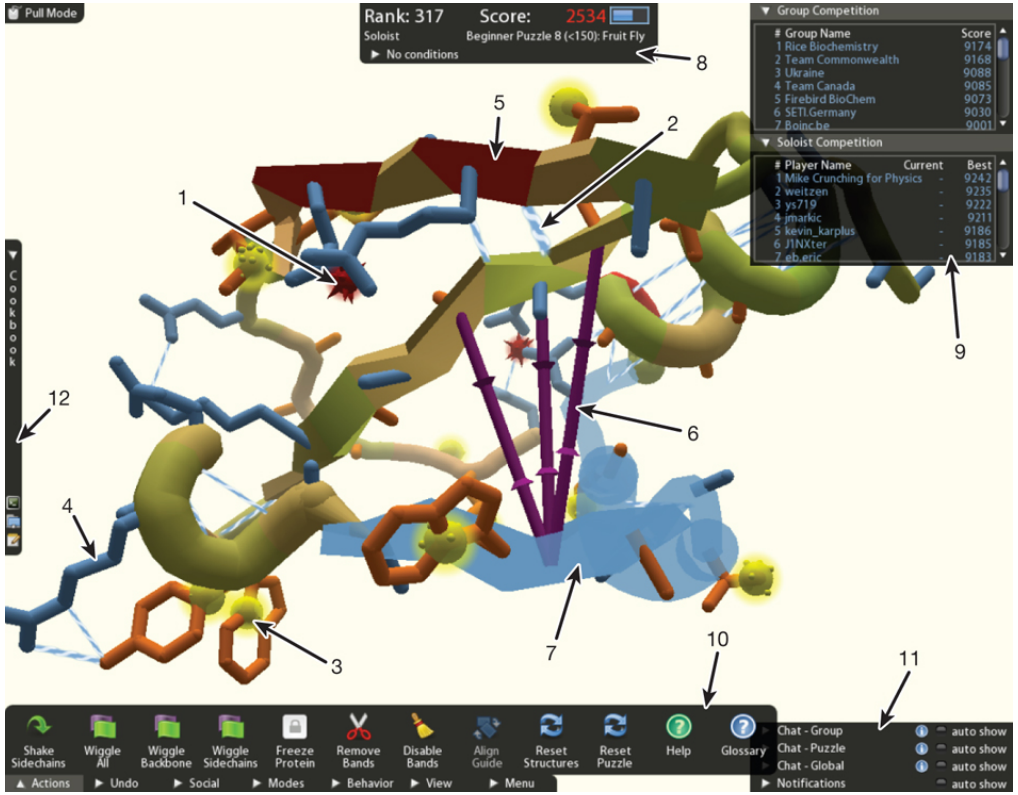
Application: Protein Folding



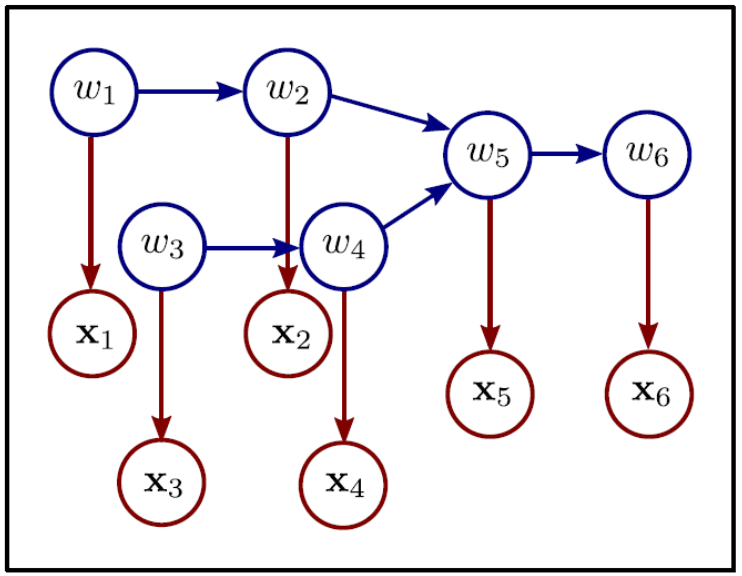
Application: Protein Folding

- Foldit
  - http://youtu.be/bTINNFQxs_A?t=175
  - http://www.youtube.com/watch?v=IGYJyur4FUA

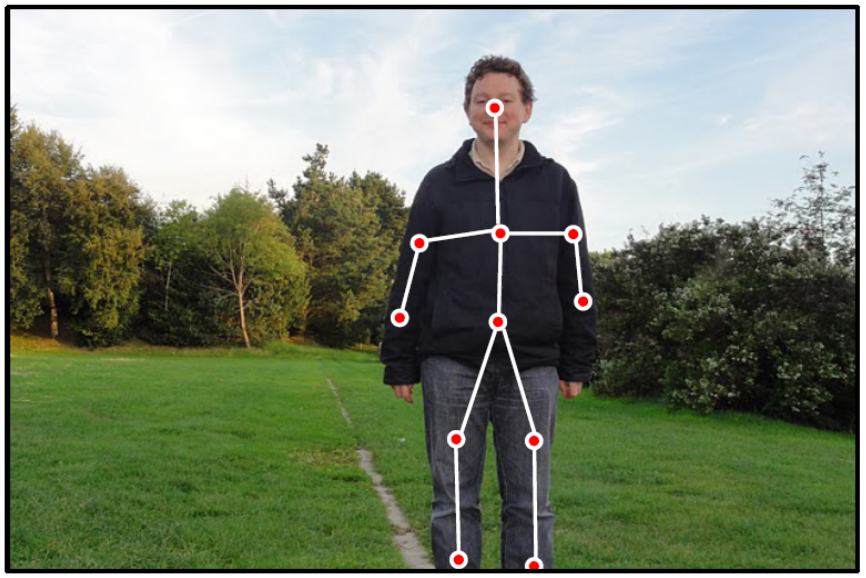




Application: Computer Vision

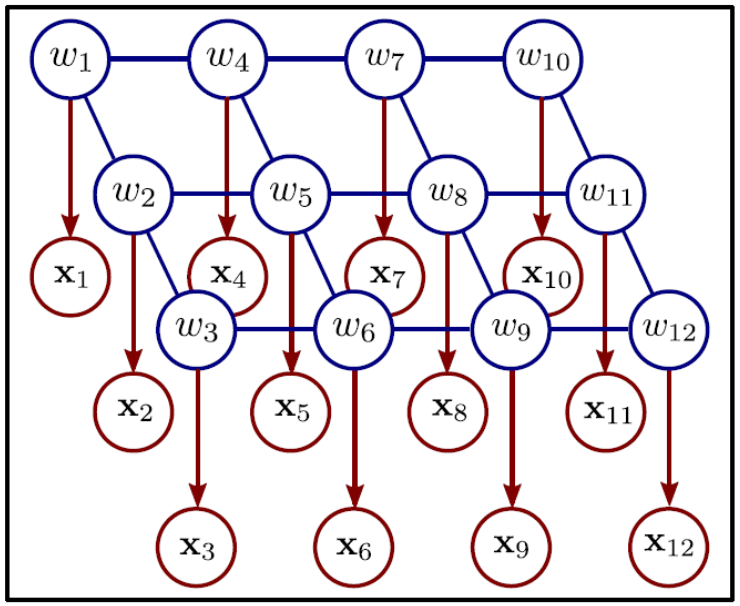


Tree model



Parsing the human body

Application: Computer Vision



Grid model

Markov random field

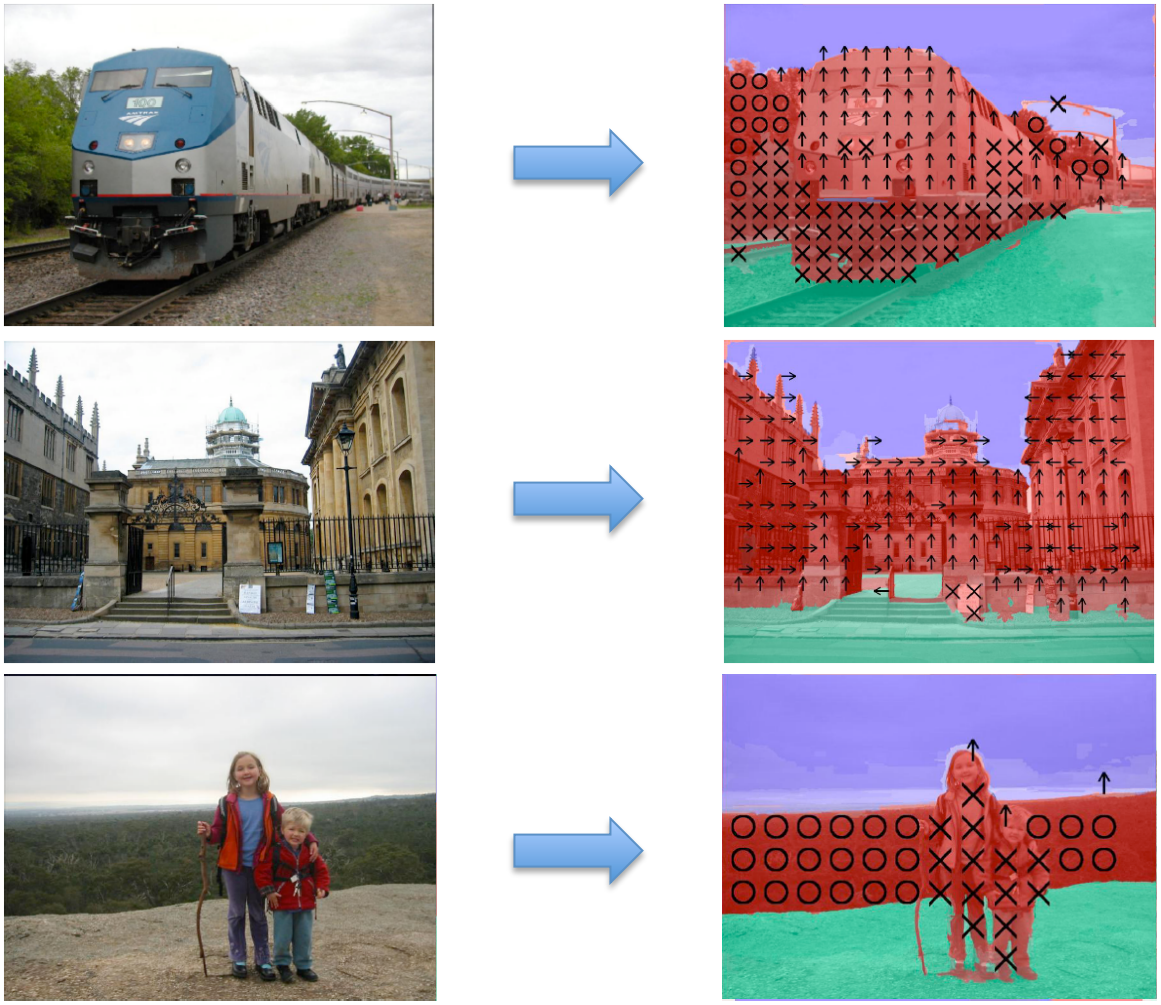
(blue nodes)

Semantic

segmentation

Application: Computer Vision

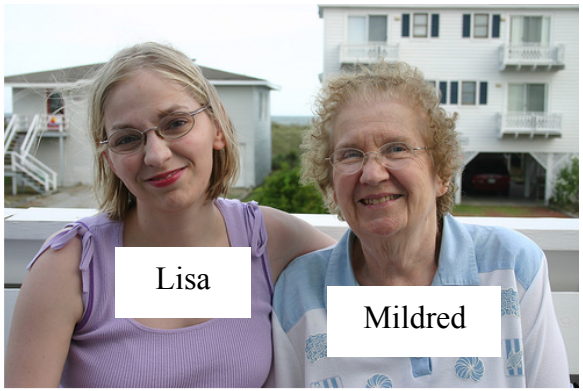
- Geometric Labelling
  - [Hoiem et al. IJCV '07], [Hoiem et al. CVPR '08], [Saxena PAMI '08], [Ramalingam et al. CVPR '08].

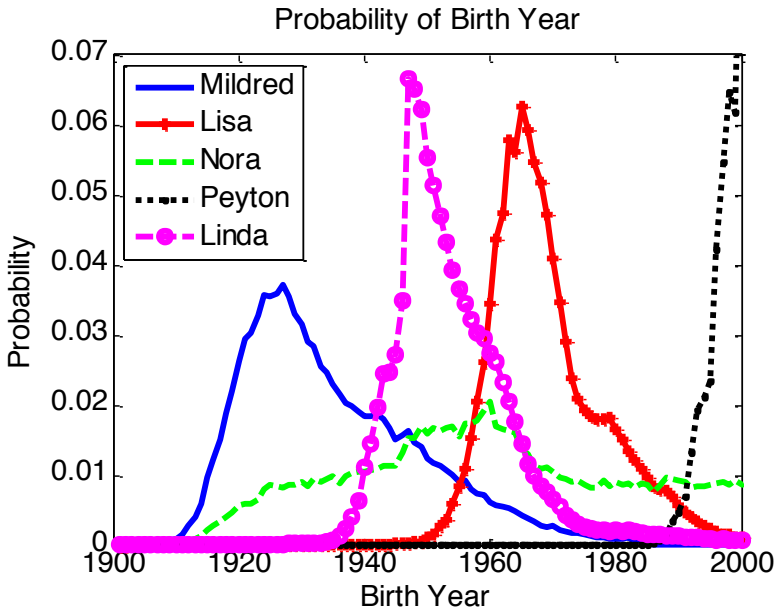


Application: Computer Vision

- Name-Face Association
  - [Berg et al. CVPR '04, Phd-Thesis '07], [Gallagher et al. CVPR '08].

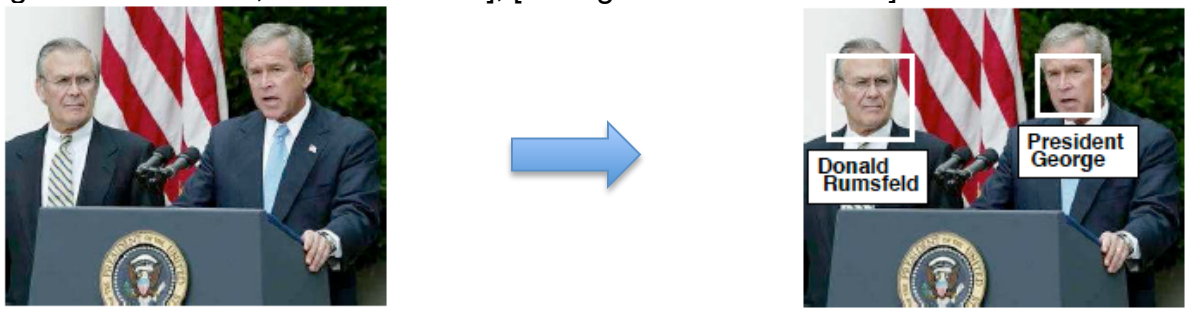
Mildred and Lisa



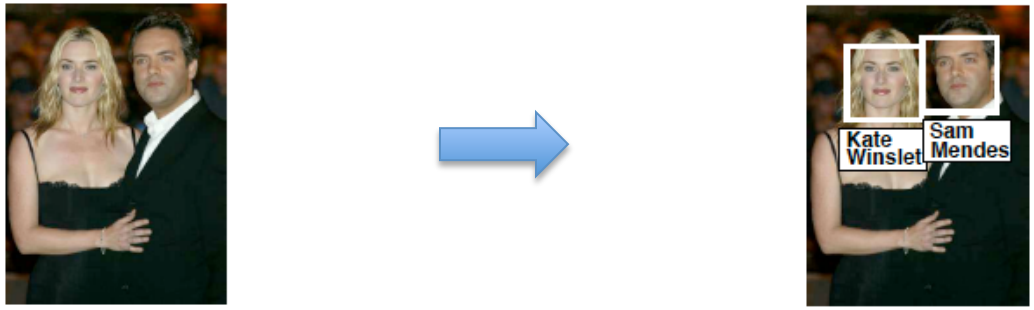


Application: Computer Vision

- Name-Face Association
  - [Berg et al. CVPR '04, Phd-Thesis '07], [Gallagher et al. CVPR '08].



President George W. Bush makes a statement in the Rose Garden while Secretary of **Defense Donald Rumsfeld** looks on, July 23, 2003. Rumsfeld said the United States would release graphic photographs of the dead sons of **Saddam Hussein** to prove they were killed by American troops. Photo by Larry Downing/Reuters



British director **Sam Mendes** and his partner actress **Kate Winslet** arrive at the London premiere of 'The Road to Perdition', September 18, 2002. The films stars **Tom Hanks** as a Chicago hit man who has a separate family life and co-stars **Paul Newman** and Jude Law. REUTERS/Dan Chung

And many

many

many

many

many

more…

Course Information

- Instructor: Dhruv Batra
  - dbatra@vt
  - Office Hours: Fri 1-2pm
  - Location: 468 Whittemore

Syllabus

- Directed Graphical Models (Bayes Nets)
  - Representation: Directed Acyclic Graphs (DAGs), Conditional Probability Tables (CPTs), d-Separation, v-structures, Markov Blanket, I-Maps
  - Parameter Learning: MLE, MAP, EM
  - Structure Learning: Chow-Liu, Decomposable scores, hill climbing
  - Inference: Marginals, MAP/MPE, Variable Elimination
- Undirected Graphical Models (MRFs/CRFs)
  - Representation: Junction trees, Factor graphs, treewidth, Local Makov Assumptions, Moralization, Triangulation
  - Inference: Belief Propagation, Message Passing, Linear Programming Relaxations, Dual-Decomposition, Variational Inference, Mean Field
  - Parameter Learning: MLE, gradient descent
  - Structured Prediction: Structured SVMs, Cutting-Plane training
- Large-Scale Learning
  - Online learning: perceptrons, stochastic (sub-)gradients
  - Distributed Learning: Dual Decomposition, Alternating Direction Method of Multipliers (ADMM)

Syllabus

- You will learn about the methods you heard about, and then some.
- You will understand algorithms, theory, applications, and implementations
- **It's going to be FUN and HARD WORK ☺**

Prerequisites

- Intro Machine Learning
  - Classifiers, regressors, loss functions, MLE, MAP
- Linear Algebra
  - Matrix multiplication, eigenvalues, positive semi-definiteness…
- Graph Concepts
  - Nodes, edges, trees, cycles, depth-first search
- Algorithms
  - Dynamic programming, basic data structures, complexity…
- Programming
  - Matlab for HWs. Your language of choice for project
- Ability to deal with “abstract mathematical concepts”
- This will be an in-depth class.

Textbook

- No required book.
  - We will assign readings from online/free books, papers, etc
- Reference Books:
  - [On Library Reserve]
    Probabilistic Graphical Models: Principles and Techniques
    Daphne Koller and Nir Friedman
  - [Free PDF from author]
    Bayesian reasoning and machine learning
    David Barber
    http://web4.cs.ucl.ac.uk/staff/D.Barber/pmwiki/pmwiki.php?n=Brml.HomePage
  - [Free PDF from authors]
    Graphical models, exponential families, and variational inference.
    Martin J. Wainwright and Michael I. Jordan.

Grading

- 5 homeworks (50%)
  - First one goes out Jan 30
    - Start early, Start early, Start early, Start early, Start early, Start early, Start early, Start early, Start early, Start early
- Final project (25%)
  - Projects done individually, or groups of two students
- Final (20%)
  - Take home
  - 3-5 days
- Class Participation / Paper Reading (5%)
  - Contribute to class discussions on Scholar
  - Ask questions, answer questions
  - Reading assigned papers

Re-grading Policy

- Homework assignments and midterm
  - **Within 3 days** of receiving grades: see me

- This is an advanced grad class.
  - The goal is understanding the material and making progress towards our research.

Homeworks

- Homeworks are hard, start early!
  - Due in 2 weeks via Scholar (Assignments tool)
  - Theory + Implementation (similar format as 4984/5984)
  - HW1 out 1/30

- "Free" Late Days
  - 5 late days for the semester
    - Use for HW, project proposal/report
    - Cannot use for midterm or final
  - After late days are used up:
    - Half credit within 48 hours
    - Zero credit after 48 hours

- All homeworks must be submitted even for zero credit

Project

- Goal
  - Chance to try Graphical Models
  - Encouraged to apply to your research (computer vision, communication, UAVs, computational biology…)
  - Must be done this semester. No double counting.
  - 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

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

Spring 2013 Projects

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

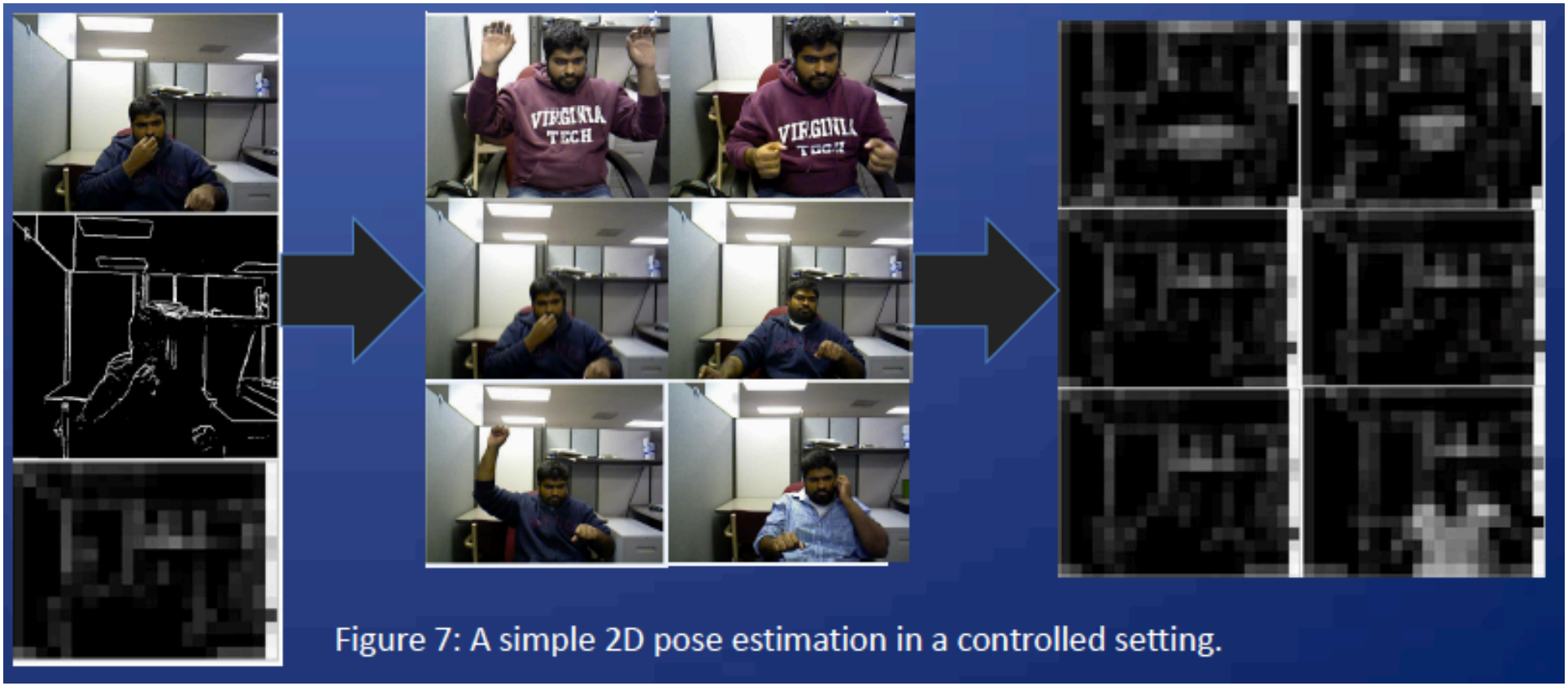
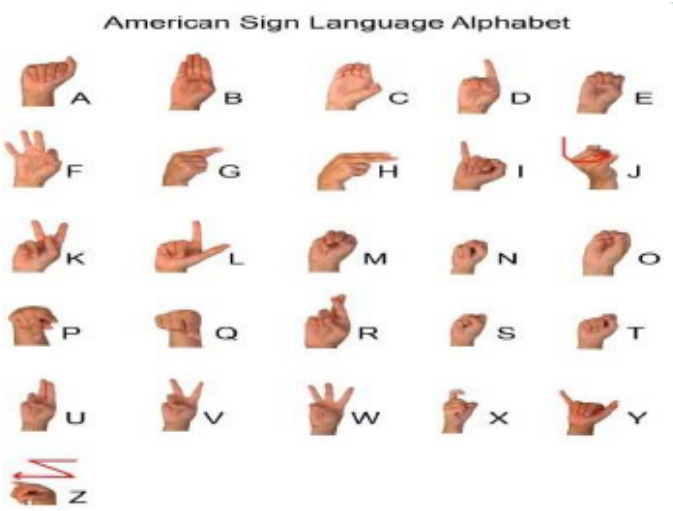
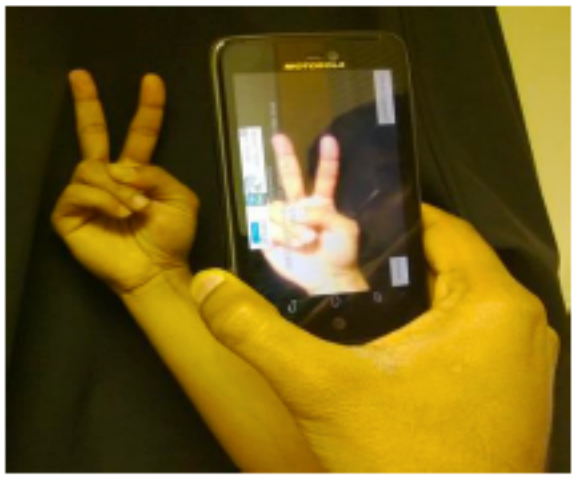


Figure 7: A simple 2D pose estimation in a controlled setting.

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).
  - 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 for the programming part

- Zero tolerance on plagiarism
  - Neither ethical nor in your best interest
  - Always credit your sources
  - Don’t cheat. We will find out.

Audit / Sit in

- Audit
  - ECE Audit Request form
    - http://www.ece.vt.edu/graduate/forms/index.html
  - Deadline: Jan 27

- Sitting in
  - Talk to instructor.

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 hints/answers. We will monitor.

- Class websites:
  - https://scholar.vt.edu/portal/site/s14ece6504
  - https://filebox.ece.vt.edu/~s14ece6504/

- Office Hours

Other Relevant Classes

- Data Analytics (CS 5526)
  - Instructor: X Deng
  - Offered: Spring
- Optimization (ISE 5406)
  - Instructor: BM Fraticelli
  - Offered: Spring
- Convex Optimization (ECE 5734)
  - Instructor: MH Farhood
  - Offered: Spring
- Advanced Computer Vision (ECE 6504)
  - Instructor: Devi Parikh
  - Offered: Spring

Guest Lectures

- Rosalyn Moran, VT CRI
  - Graphical Models for Neuroscience
  - Variational Inference

Misc Notes

- Mix of power-point + writing on board
  - Slides + notes available on scholar
- Difficulty level of this class
  - On par with Spring 2013 4984/5984
  - Significantly more than Fall 2013 4984
  - More than Fall 2013 5984
- Exciting topic; Advanced Class
  - Focus on depth, not breadth
  - We will go as slow as necessary and bearable ☺

Plan for Today

- Nothing!

Todo

- Readings
  - Probability Refresher: Barber Chap 1
  - Graph Theory Refresher: Barber Chap 2