Lecture 1: January 12, 2015
cs 573: Probabilistic Reasoning
Professor Nevatia
Spring 2015

Introduction

- Course name: Probabilistic Reasoning
  - csci 573
- Instructor: Prof. Ram Nevatia
  - My background, interests
- Attendance sheet, in class students only
- Today's objective
  - Describe course content
  - Conduct of the class
  - Required work, grading
  - Pre-requisites

General Information

- Course web page https://courses.uscden.net
  - Must be registered for class to have access
  - Usually, a preliminary version of slides will be posted prior to class, complete version after the class.
  - Lecture videos, course notes….
- Office hours:
  - Instructor: MW 1:30-3:00 P.M., PHE 204, 213-740-6427, nevatia AT usc.edu
  - This week only: Today, 1- 2 P.M., none on Wednesday
  - TA: Tom Collins, collinst AT usc.edu , office hours and place TBD
- Book:
  - Required: *Probabilistic Graphical Models* by Koller and Friedman, MIT Press 2010. Note: errata for early printings may be found at http://pgm.stanford.edu/errata/
- Koller video lectures available at: https://class.coursera.org/pgm/lecture/preview

About Enrollment

- On-campus section is fully subscribed
- Capacity limited by physical space and ability to provide individual attention to students
  - Not possible to add another section due to non-availability of a qualified instructor
- Heavy demand is a surprise to us
  - Previous years have seen much lower demand and enrollments
- Common for students to drop in first 1-2 weeks of class
  - Will fill in if and as students drop

Prerequisites

- Undergraduate level course in probability theory
- Good skills with basic mathematics such as calculus and linear algebra
- Programming skills: ability to convert algorithms into programs
- May have some overlap with cs561 (basic AI course) and cs567 (Machine Learning).
  - Neither is a pre-requisite

Requirements and Grading

- Assignments: 7-8 assignments, mostly written (mathematical) but 1-2 programming assignments may be included. *Large* projects are not planned.
- Grading:
  - Assignments 30%,
  - Exams 1 and 2: 30% each
  - Class attendance and participation 10% (does not apply to DEN students)
- Exams:
  - Exam 1, during 7th or 8th week of classes; exact date will be announced at least 1 week in advance
  - Exam 2, April 29, 2015; last day of scheduled classes
- Programming
  - No programming projects are planned but some programming may be necessary to solve numerical problems
  - Students may choose their own language but some packages may be specific to some languages

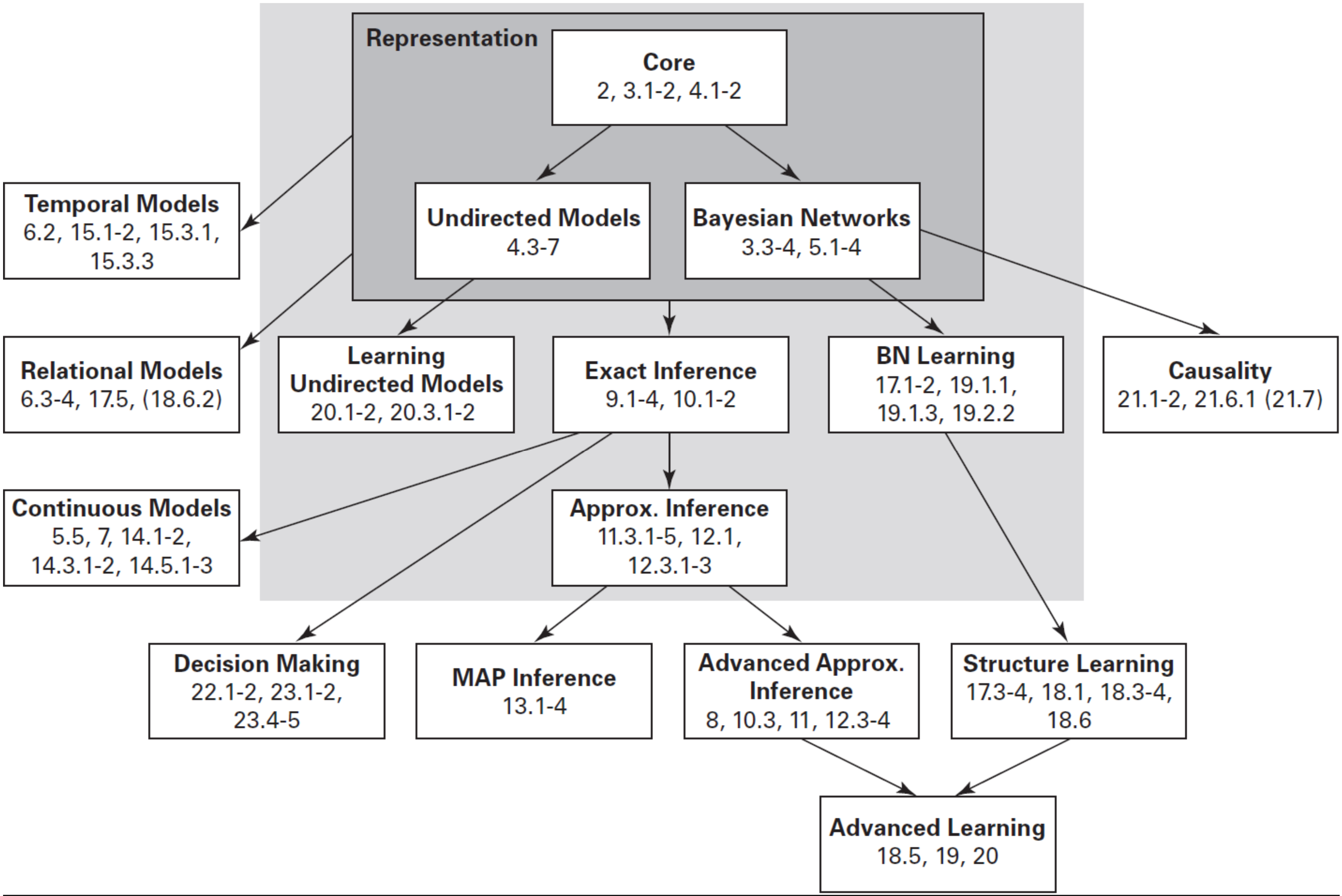
Course Objectives

- In-depth coverage of issues related to probabilistic reasoning. These include:
  - Probabilistic Representations (~ 4 Weeks)
    - Bayesian networks, undirected graphs, dynamic networks
  - Probabilistic inference (~ 4 Weeks)
    - Exact and approximate, including for temporal graphs
  - Learning of parameters and structure of probabilistic graphical models (~ 3 Weeks)
  - Decision making under uncertainty (~ 2 Weeks)
- First two topics will get the most attention. There are other courses that cover learning and planning in more detail.
- Focus on concepts and algorithms, not applications or commercial systems

Syllabus in terms of Book Sections

- Note: This is only a plan, actual coverage may vary some. Not all material in each section may be covered and some external material may be included.
- Representations: Chapters 2-5 all except sections marked with "*" in the book, Chapter 6: 6.1, 6.2 only
- Inference: Chapter 9 (except 9.5 and 9.6), 10, 11.1 to 11.3, 12.1 to 12.3, 13.1 to 13.3, 15.1, 15.2; additional notes
- Parameter Estimation: 17.1, 17.2 (exclude 17.2.4, 17.2.5), 19.2.2 (exclude 19.2.2.5, 19.2.2.6), Additional Notes
- Decision Theory: Ch. 22
- Selected parts of Ch. 7, 14, 11.4, 23

Book Plan



Representation
Core
2, 3.1-2, 4.1-2
Temporal Models
6.2, 15.1-2, 15.3.1, 15.3.3
Undirected Models
4.3-7
Bayesian Networks
3.3-4, 5.1-4
Relational Models
6.3-4, 17.5, (18.6.2)
Learning Undirected Models
20.1-2, 20.3.1-2
Exact Inference
9.1-4, 10.1-2
BN Learning
17.1-2, 19.1.1, 19.1.3, 19.2.2
Causality
21.1-2, 21.6.1 (21.7)
Continuous Models
5.5, 7, 14.1-2, 14.3.1-2, 14.5.1-3
Approx. Inference
11.3.1-5, 12.1, 12.3.1-3
Decision Making
22.1-2, 23.1-2, 23.4-5
MAP Inference
13.1-4
Advanced Approx. Inference
8, 10.3, 11, 12.3-4
Structure Learning
17.3-4, 18.1, 18.3-4, 18.6
Advanced Learning
18.5, 19, 20

Course Style

- Interactive to the extent possible
- Will follow book closely; however, we will skip some details such as proofs of most theorems.
  - Assignments may require more detailed knowledge than covered in class
  - Preliminary version of lecture slides will be posted in advance of the lectures
- We will discuss algorithms at a relatively high level, almost never at the code or data structure levels
  - In English, in pseudo-code, by examples
  - Ability to convert high level descriptions to code is assumed

Reasoning under Uncertainty

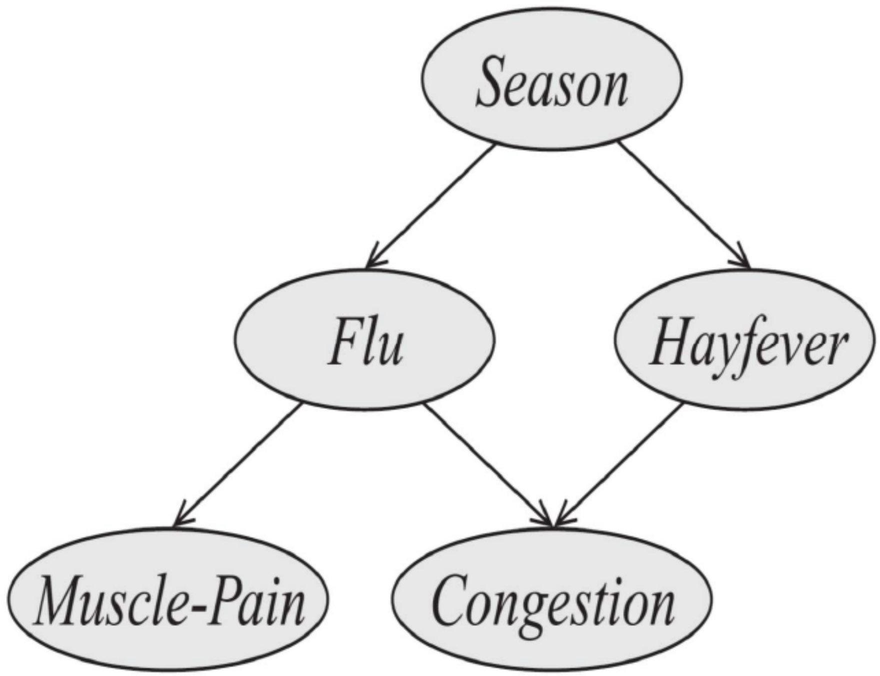
- Uncertainty is almost always present in solving real problems
  - State of the world is not known precisely or not even knowable in principle
  - Some aspects of the state can't be measured directly (*e.g.* cause of some types of sickness)
  - Effects of actions are uncertain
    - *e.g.* What route to take to go to the airport? When to start?
      What courses to take to succeed in a cs career?
- Rational decision (under uncertainty)
  - Consider the relative importance of various goals and the *likelihood* that they will be achieved.
  - Rationality does not *guarantee* success

Possible Application Areas

- Originally developed for “expert systems”
  - Medical diagnosis (Pathfinder), Prospector…
  - Attempt to systematize reasoning about uncertain knowledge
- Now, virtually all aspects of computer science
  - AI, robotics, vision, speech
  - e-commerce, search….
  - Networks, OS, software engineering…
- Outside CS
  - Economics, finance, weather prediction, political science….
- Likely to become as important in CS study as discrete algorithms taught in courses such as cs570/670

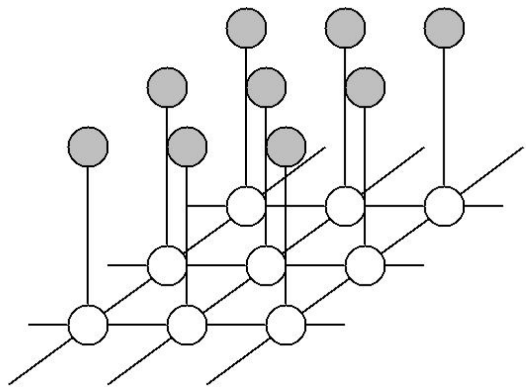
An Example Graph

- Note: only the *structure* of relations is shown
- Methodology attempts to combine rule-based representations with probabilistic reasoning



Denoising an Image

Observed, noisy image

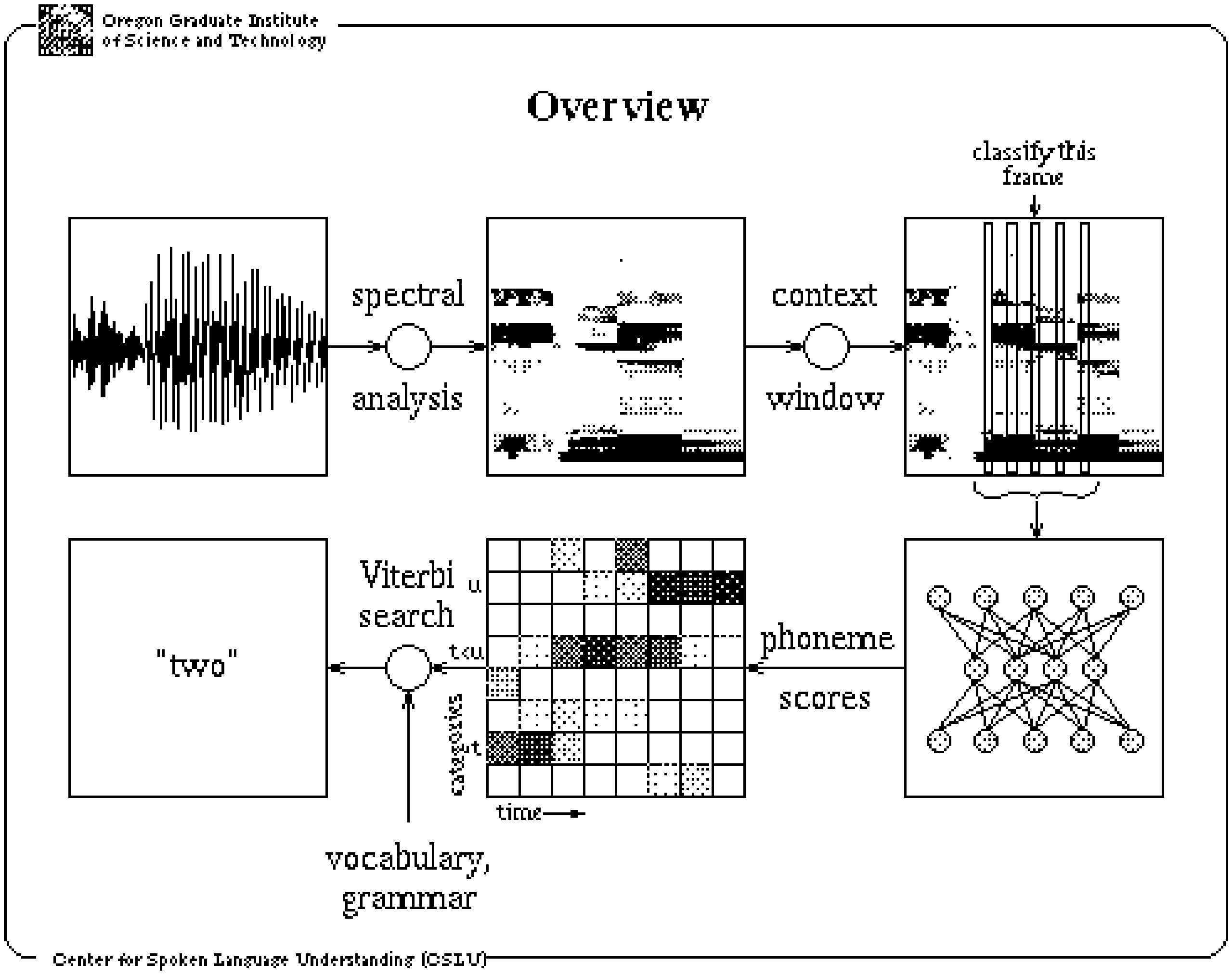


Model

Estimated Image

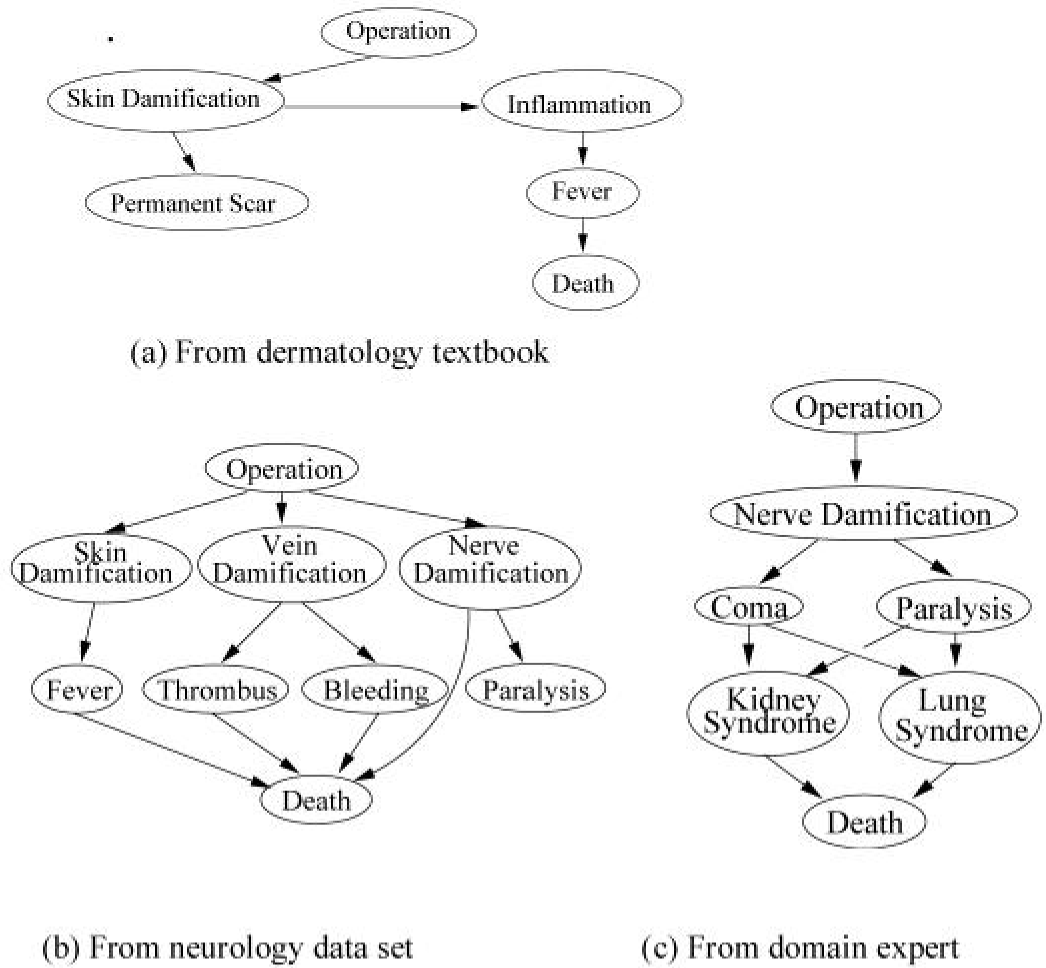
From: http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/AV0809/ORCHARD/

Speech Recognition



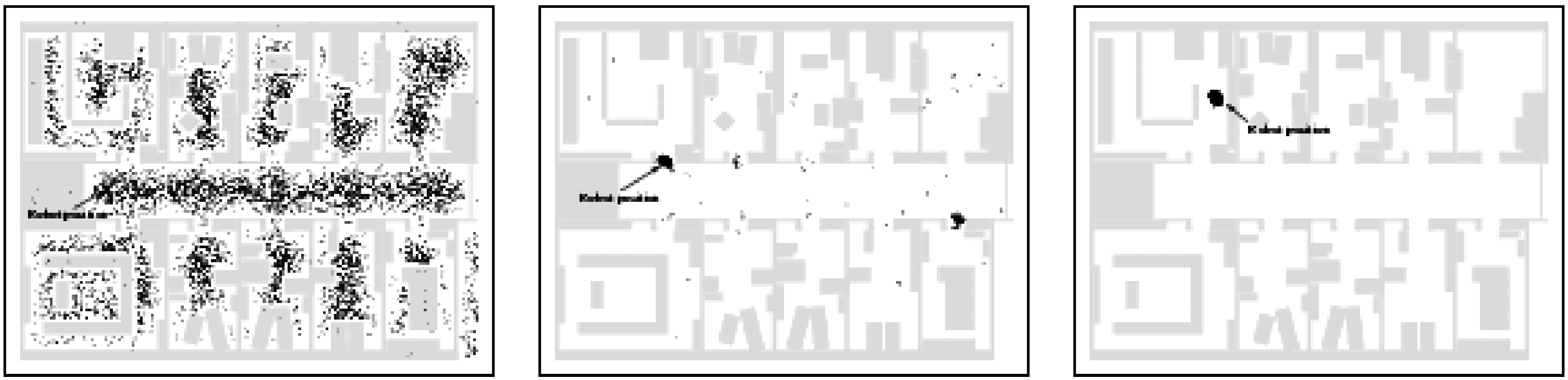
http://www.cslu.ogi.edu/tutordemos/nnet_recog/overview.gif

Medical Diagnosis



From C. Jiang et al: AMIA Annual Proceedings 2005, 370-374

Robot Localization



http://www.cc.gatech.edu/~dellaert/assets/images/autogen/a_sonar.gif

Nature of Study

- Develop a representation for a problem domain
  - We will not study specific domains, just the methodologies
  - Will normally consist of choice of *random variables* and relations between them, represented in a graph
- Given values of some variables, compute probability distributions of others. Some types of *queries*:
  - *Forensic/diagnostic*: causes given evidence
  - *Predictions*: observables given cause
  - *Explanations*: of observed phenomenon
  - *Actions*: optimal actions given a model and observations
- Learning of the parameters (sometimes also the structure) of the graph from examples

Basic Probability: Informal concepts

- *Unconditional* or *prior* probability that a proposition A is true
  – Let P (the event of rain falling on jan 12 in L.A. is True) = 0.2
  – *P* (*fair_coin_toss:head*) = 0.5
- *Conditional* probability P(rain_today|cloudy_today)
- *Joint* probability P(rain_Jan12, cloudy_Jan12)
- *Independence* of variables: coin toss and weather

Probability Theory: Some notations

- Event *space*, space of all possible outcomes, Ω
  - For outcome of roll of dice: Ω = {1,2,3,4,5,6}
  - Each event is a subset of Ω, e.g. {3}, {2,4,6} (even outcome)
- Set of measureable events: *S*
  - We assign probabilities to elements of S
  - Contains *empty event 0* and *trivial event* Ω
  - Closed under union (if α and β ε S, then so is α U β)
  - If α ε *S* , so is Ω - α

Probability Distributions

- Probability distribution P defined over (Ω, S):
  - Mapping from events in S to real values (probabilities)

    $P(\alpha) \geq 0$ for all $\alpha \in S$

    $P(\Omega) = 1$

    If $\alpha, \beta \in S$, and $\alpha \cap \beta = 0$, then $P(\alpha \cup \beta) = P(\alpha) + P(\beta)$
- It follows that $P(0) = 0$ and

    $P(\alpha \cup \beta) = P(\alpha) + P(\beta) - P(\alpha \cap \beta)$
- Common interpretation of probability is as frequency of events
  - Examples: coin toss, dice roll, weather (how many outcomes?)

Conditional Probability, Bayes' Rule

- Conditional probability: P (β | α)
  - Probability that β is true, given that α is true
  - *e.g.* P (rain|cloudy), P(gradeA|high_intell)
  - P (β | α) = P(α ∩ β) / P (α)
- Chain rule: P(α ∩ β) = P (α) P (β | α)

$$P(\alpha_1 \cap \ldots \cap \alpha_k) = P(\alpha_1)P(\alpha_2 \mid \alpha_1) \cdots P(\alpha_k \mid \alpha_1 \cap \ldots \cap \alpha_{k-1}).$$

- Bayes' rule P (α | β) = P (β | α) P (α) / P (β)
  - It may be easier to gather P (β | α) then P (α | β), for some situations (*e.g.* probability of symptoms given disease rather than the disease given symptoms)

$$P(\alpha \mid \beta \cap \gamma) = \frac{P(\beta \mid \alpha \cap \gamma)P(\alpha \mid \gamma)}{P(\beta \mid \gamma)}$$

Random Variables

- Random variables take on different values with probabilities given by a distribution function
  - *e.g.* coin toss, roll of dice, grades in a course *etc*
  - Formally, defined by function that associates a real value to each outcome in Ω.
  - Random variables normally denoted by upper case letters X, Y, Z…
- Discrete random variable: takes one of a finite set of values: $Val(X)$ denotes the set of values of X
- $P\ (X = x^i)$ denotes the probability that X takes value x^i
- Often abbreviated to $P\ (x^i)$
  - $\sum_x$ denotes summation over all possible values of X
  - Sum over all possible values must equal 1 for a discrete variable, *i.e.* $\sum_x P\ (x) = 1$
  - Distribution is called multinomial
    - Binomial or Bernoulli for a binary variable
- Bold letters used to denote a ***set*** of variables, e.g. $\boldsymbol{X}$, $\boldsymbol{Y}$, $\boldsymbol{Z}$…
  - $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$ denote values of variable in these sets

Marginal and Joint Distributions

- *Joint distribution*: probabilities for each combination of values of the random variables.
- $P(x, y)$ is used to denote joint probability of $X = x$ and $Y = y$

Figure 2.1, example of P(Intelligence, grade)

		Intelligence *low*	*high*	
Grade	*A*	0.07	0.18	0.25
	B	0.28	0.09	0.37
	C	0.35	0.03	0.38
		0.7	0.3	1

- *Marginal distribution*: Distribution of one variable, regardless of the values of others;
  - obtained by summing over all other variables from the joint distribution, *e.g.* P(Intelligence = high) or P(grade = A);

Conditional Distribution

- P (X | Y):
  - for each value of Y, assign a distribution over values of X.
- Chain rule: P (X, Y) = P(X) P(Y | X)

$$P(X_1, \ldots, X_k) = P(X_1)P(X_2 \mid X_1) \cdots P(X_k \mid X_1, \ldots, X_{k-1})$$

- Bayes' rule *P* (X | Y) = P(X) P(Y | X)/ P(Y)

$\perp$

Independence

- Event α is independent of event β in P, denoted as $P \models \alpha \perp \beta$,
  if $P(\alpha \mid \beta) = P(\alpha)$, or if $P(\beta) = 0$
- Follows that $P(\alpha \cap \beta) = P(\alpha) P(\beta)$
- Examples: toss two coins; coin toss and weather…
- Full independence is rare, *conditional independence* where two events are independent, given a third event
- Conditional independence
  - P (USC | UCLA, GradeA) = P(USC | GradeA)
    (USC means admitted to USC, similar for UCLA)
  - P (Congestion | Flu, Hayfever, Season) = P (Congestion | Flu, Hayfever)
- Event α is independent of event β in P, given event γ, denoted as $P \models (\alpha \perp \beta \mid \gamma)$ if $P(\alpha \mid \beta \cap \gamma) = P(\alpha \mid \gamma)$, or if $P(\beta \mid \gamma) = 0$
- Follows that $P(\alpha \cap \beta \mid \gamma) = P(\alpha \mid \gamma) P(\beta \mid \gamma)$

Next Class

- Read sections 2.1, 2.2 and 3.1 of the KF book