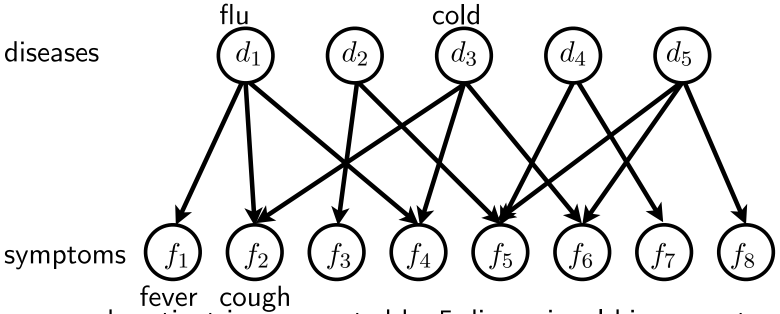
CSE 515 Statistical methods in computer science

- Wed-Fri 11:30am - 12:45pm, LOW 105
- Sewoong Oh (sewoong@cs.washington.edu)
- `https://courses.cs.washington.edu/courses/cse515/20wi`
- 4 homeworks (80%)
  take-home final exam (20%)

Topics

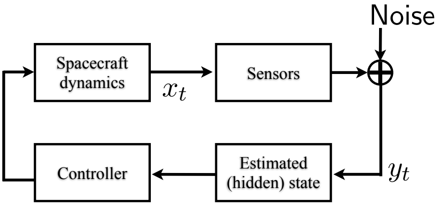
- Provide a unifying framework for **inference** tasks in complex systems
- Graphical models: random variables sit on vertices
- Probability distributions: that can be ‘decomposed’ or ‘factorized’
- Inference tasks : draw a conclusion based on the distribution
- Applications: Images, error-correcting codes, machine learning, etc.

Example: translate information in Quick Medical Reference into a graphical model

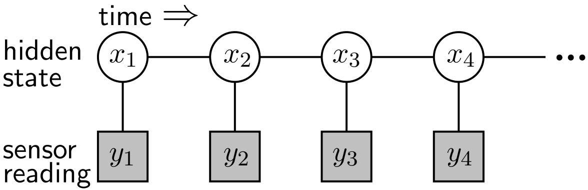


- each patient is represented by 5-dimensional binary vector $d = [d_1, \ldots, d_5] \in \{0,1\}^5$ representing which diseases are present (value 1) in the patient
- $f_j \in \{0,1\}$: symptoms are determined by diseases
- Graphical model: Bayesian network (e.g. $\mathbb{P}(f_2 = 1 | d_1 = 0, d_3 = 1)$)
- Inference task: Given symptoms (e.g. $f = 01010010$), what disease is likely ($\arg\max_d \mathbb{P}(d|f)$)?

Example: Navigation



Navigating Spacecrafts (e.g. lunar landing, guiding shuttles)



- Linear system as Gaussian graphical models (e.g. $\mathbb{P}(x_2|x_1, u_1)$)
- Inference task: Given noisy sensor readings, what is the current state? (compute $\mathbb{P}(x_4|y_1, \ldots, y_4)$)
- Kalman filtering

Example: Image processing

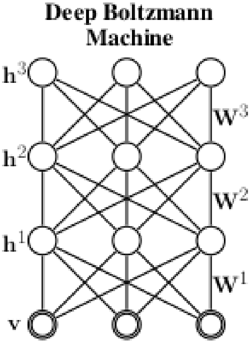
Can computers generate/classify handwritten letters/numbers?
[R.Salakhutdinov, G. Hinton, 2009 AISTATS]

$10,000$ Training data

Reconstruction by sampling

28×28 Pixel images

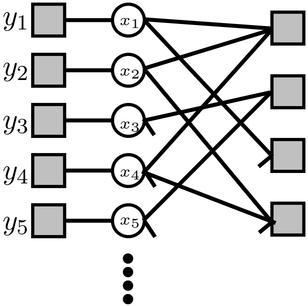
Example: Image processing



- Pairwise Markov random fields (deep Boltzmann machines)
- Gibbs sampling

Example: Communication

Error-correcting codes (e.g. Low-Density Parity Check codes)



- Factor graphs
- (loopy) Belief propagation
- Inference task: Received y_1^n, what x_1^n is most likely? $(\arg\max_x \mathbb{P}(x|y))$

General theme

Probability distribution over $X = (X_1, X_2, \ldots, X_n)$
given observations $Y = (Y_1, \ldots, Y_m)$

$$\mu_y(x) \;=\; \mathbb{P}_{X_1,\ldots,X_n|Y_1,\ldots,Y_m}(x_1, x_2, \ldots, x_n|y_1, y_2, \ldots, y_m)$$

from a set $x_i \in \mathcal{X}$ and $y_j \in \mathcal{Y}$, typically $|\mathcal{X}| < \infty$

- Finding the most probable realization

$$\hat{x} \in \arg\max_{x \in \mathcal{X}^n} \mu_y(x)$$

- Calculate marginals

$$\mu_y(x_1) \;=\; \sum_{x_2,\ldots,x_n} \mu_y(x)$$

- Sampling

Key challenge: $n \gg 1$

computational complexity is $O(|\mathcal{X}|^n)$ and there is no efficient method for general distributions

Structure

Suppose the variables are independent

$$\mu_y(x) = \mu_1(x_1)\mu_2(x_2)\cdots\mu_n(x_n)$$

then, computational complexity is only $|\mathcal{X}| \cdot n$

- Finding the most probable realization

$$\hat{x}_i \in \arg\max_{x_i \in \mathcal{X}} \mu_i(x_i)$$

- Calculate marginals

$$\mu_y(x_1) \quad = \quad \mu_1(x_1)$$

- Sampling: $X_1, X_2, \ldots, X_n$ independently

When the probability distribution factorizes,
we can achieve huge computational gains

Graphical models

- Undirected pairwise graphical models
- Factor graphs
- Bayesian networks

Topics include

- Representing inference tasks using graphical models
- General and powerful framework for efficient inference
- Belief propagation
- Hidden Markov models, Kalman filtering
- Plenty of math: convex analysis, random processes, Markov chains, etc.