Probabilistic Reasoning

Russell Chap 13, 14, 15

Luger Chap. 5, 9

Managing Uncertainty

- Abduction
  - If $P \Rightarrow Q$, Q then P → Not sound
- Most diagnostic rules are abduction
  - True implication (causal relation):
    - *Problem $\Rightarrow$ Symptom*

      Ex> Problem in battery $\Rightarrow$ Light gone
  - Expert system rule:
    - *Symptom $\Rightarrow$ Problem*

      Ex> Light gone $\Rightarrow$ Problem in battery

  → Not 100% true

  → Need measure of uncertainty

Certainty Factor

- Certainty Theory
  - Confidence measure (expert heuristic) + combining rules

    E(evidence) → H(hypothesis) : Confidence?
  - Certainty Factor(CF) of H(hypothesis) given E(evidence)
    - CF(H | E) = MB(H | E) - MD(H | E)
    - MB: Measure of belief of H given E
    - MD: Measure of disbelief of H given E
  - Inference with CF
    - Rule: if P then Q (CF = 0.9)
    - Fact: P (CF = 0.9)

    ⇒ Q (CF = 0.9*0.9 = 0.81)
  - Combining CF
    - CF(P1 and P2) = min (CF(P1), CF(P2))
    - CF(P1 or P2) = max (CF(P1), CF(P2))

Certainty Factor

- Example
  - Rule: (P1 and P2) or P3 $\rightarrow$ R1(0.7) and R2(0.3)
  - During production cycle, it is found that

    CF(P1) = 0.6

    CF(P2) = 0.4

    CF(P3) = 0.2
  - Then

    CF((P1 and P2) or P3) = max (min (0.6, 0.4), 0.2)

    = 0.4

    $\therefore$CF(R1) = 0.4 * 0.7 = 0.28

    CF(R2) = 0.4 * 0.3 = 0.12

MYCIN

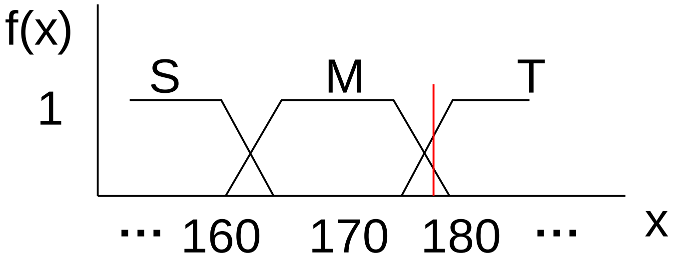
- Purpose
  - Expert system: determines the organism that causes meningitis and provides therapy
- Inference
  - Start from a hypothesis(organism), perform backward chaining
  - Use certainty factor
- Example rule
  - ```
    If   ((infection is primary-bacteremia) AND
          (portal-of-entry is gastrointestinal-tract))
    Then (conclude bacteroid 0.7)
    ```
    ```

Fuzzy Logic

- Fuzzy set
  - Set membership is a function to [0, 1]

    Ex> S = set of small integer

    fs(1) = 1.0, fs(5) = 0.7, fs(100) = 0.001
  - Membership function
    - Sets: S, M, T



$f_T(178) = 0.7$, $f_M(178) = 0.3$

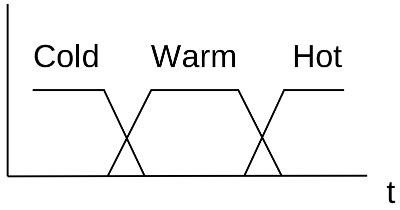
Fuzzy Logic

- Fuzzy logic
  - If $f_A(x) = a$, $f_B(x) = b$ then

    $f_{A \cap B}(x) = \min(a, b)$

    $f_{A \cup B}(x) = \max(a, b)$
- Fuzzy rules
  - If (speed = high) and (no. of cars = many) then (push brake)
  - If (temp = hot) then (fan speed = high)

Reasoning with Fuzzy Logic



1. Define fuzzy sets
   - Temperature Hot, Warm, Cold, …
2. Build logical rules with the fuzzy sets
   - If (temp is Hot) then (fan_speed is High)
3. Perform inference
   1. Fuzzification : real value (30° C) → fuzzy set (Hot)
   2. Apply rules : use fuzzy logic
   3. Defuzzification : fuzzy set (High) → real_value (4500 RPM)

Reasoning with Fuzzy Logic

- Example: Inverted Pendulum
  - Input: $x1 = \theta$, $x2 = d\theta/dt$
  - Output: u = movement M

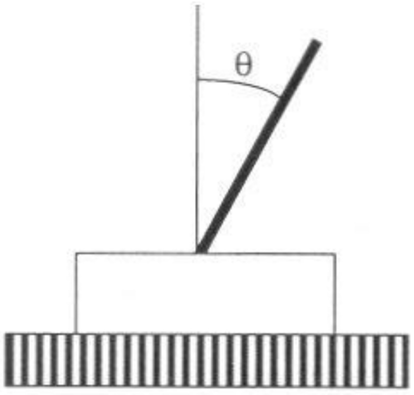


Figure 7.15 The inverted pendulum and the angle θ and dθ/dt input values.

Reasoning with Fuzzy Sets

- Fuzzy set
  - x1, x2: Z(zero), P(positive), N(negative)
  - u: Z, P, PB(positive big), N, NB(negative big)
- Rules
  1. If (x1 = P and x2 = Z) then u = P
  2. If (x1 = P and x2 = N) then u = Z
  3. If (x1 = Z and x2 = Z) then u = Z
  4. If (x1 = Z and x2 = N) then u = N
  5. . . .

Reasoning with Fuzzy Sets

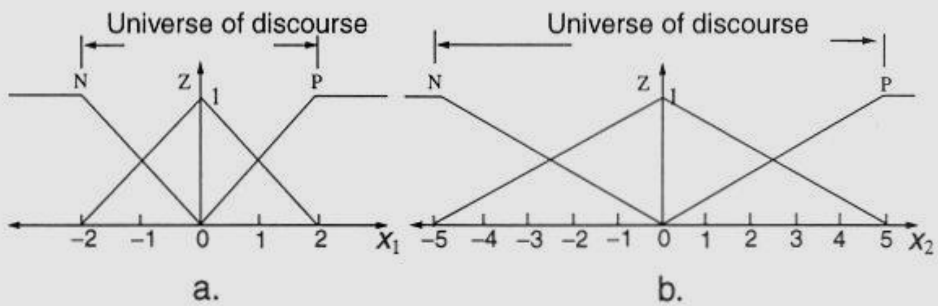


Figure 7.16 The fuzzy regions for the input values θ a. and dθ/dt b.

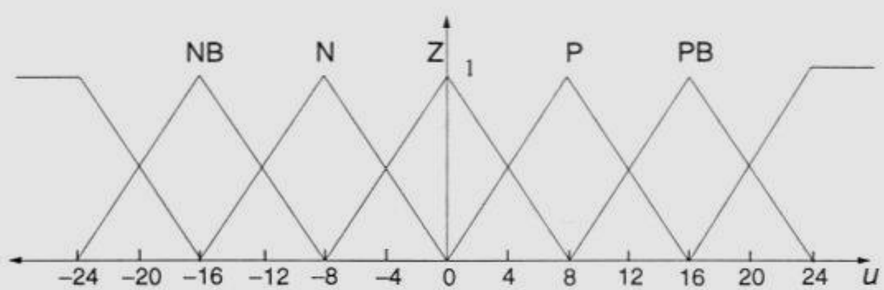


Figure 7.17 The fuzzy regions of the output value *u*, indicating the movement of the pendulum base.

Reasoning with Fuzzy Sets

- When x1 = θ = 1, x2 = dθ/dt = -4

  1. from fuzzy set membership

     x1: 0.5P, 0.5Z

     x2: 0.8N, 0.2Z

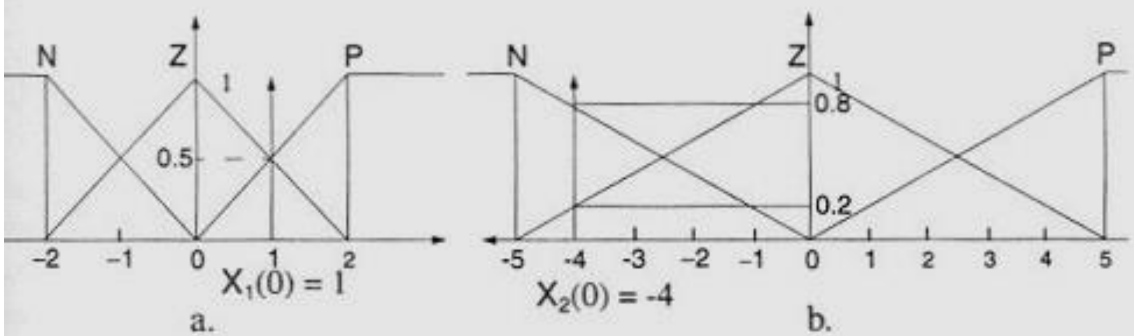
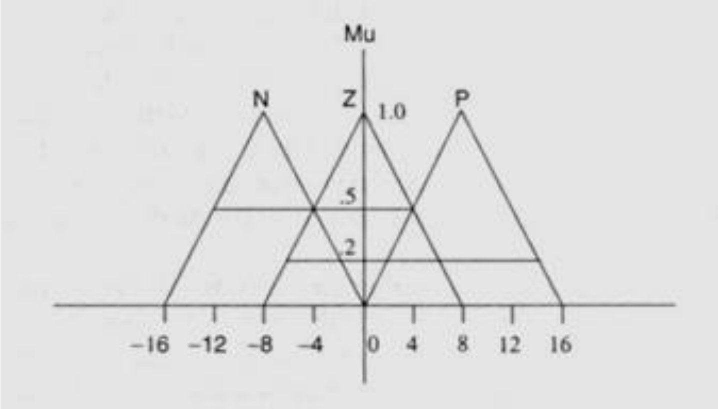


Figure 7.18 The fuzzification of the input measures x_1=1, x_2 = -4.

Reasoning with Fuzzy Sets

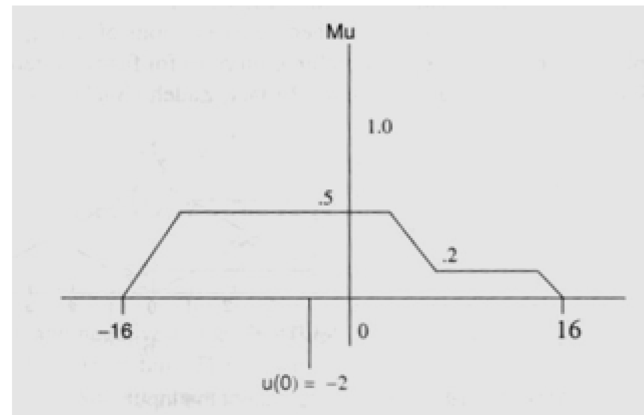
2. From fuzzy rules

u = 0.2P (1. If x1 = P(0.5) and x2 = Z(0.2) then u = P(min(0.5,0.2))
u = 0.5Z (2. If x1 = P(0.5) and x2 = N(0.8) then u = Z(min(0.5,0.8))
u = 0.2Z (3. If x1 = Z(0.5) and x2 = Z(0.2) then u = Z(min(0.5,0.2))
u = 0.5N (4. If x1 = Z(0.5) and x2 = N(0.8) then u = N(min(0.5,0.8))



Reasoning with Fuzzy Sets

3. From fuzzy set membership, computing centroid of the union
   → u = -2



X1 = 1,
X2 = -4

Fuzzification →

P, Z, N

Fuzzy rules

P, Z, N

← *Defuzzification*

U = -2

Probability

- For an experiment that produces outcomes
  - Sample space S: set of all outcomes
  - Event E: a subset of S
  - Probability P(E) = |E|/|S| (if outcomes are equally likely)

- Probability distribution
  - Function P: x → [0, 1] (x: random variable)

S	x	P(x)	y(odd)	P(y)
1	1	1/6	1	1/2
2	2	1/6	0	1/2
…	…	…	…	
6	6	1/6	0	

S	x(# of H)	P(x)
HH	2	1/4
HT	1	1/2
TH	1	
TT	0	1/4

Probability

- Joint probability distribution
  - P(x,y) : represents joint probability distribution of x and y
  - Example
    - Gas{true,false}, Meter{empty, full}, Start{yes, no}
    - P(G,M,S):

Gas	Meter	Start	P(G,M,S)
false	empty	no	0.1386
false	empty	yes	0.0014
false	full	no	0.0594
false	full	yes	0.0006
true	empty	no	0.0240
true	empty	yes	0.0560
true	full	no	0.2160
true	full	yes	0.5040

Probability

Gas	Meter	Start	P(G,M,S)
false	empty	no	0.1386
false	empty	yes	0.0014
false	full	no	0.0594
false	full	yes	0.0006
true	empty	no	0.0240
true	empty	yes	0.0560
true	full	no	0.2160
true	full	yes	0.5040

- P(Gas=false, Meter=empty, Start=no) = 0.1386
- P(Start=yes) = 0.5620

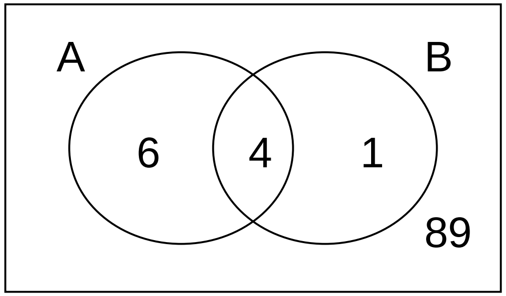
  ⇨ *Prior probability*

- P(Start=yes), given that Meter=empty ?

  ⇨ *Posterior probability (conditional probability)*

Conditional Probability

- Conditional probability
  - $P(A|B) = P(A \wedge B) / P(B)$
  - Example
    - $P(A \wedge B) = 4 / 100 = 0.04$
    - $P(B) = 5 / 100 = 0.05$
    - $P(A | B) = 4 / 5 = 0.80$



- Independence
  - $P(A|B) = P(A)$
  - $P(A \wedge B) = P(A|B)\ P(B)$

    $= P(A)\ P(B)$ if A, B are independent

Bayesian Reasoning

- Let H: Hypothesis (Problem)
  E: Evidence (Symptom)
- Then the probability of E → H is P(H | E)

$$P(H \cap E) = P(H \mid E) * P(E)$$
$$= P(E \mid H) * P(H)$$

$$\therefore \quad \boxed{P(H \mid E) = \frac{P(E \mid H) * P(H)}{P(E)}}$$

Bayesian Reasoning

- If we want to compare P(H1 | E), P(H2 | E), …

$$P(H1 \mid E) = \frac{P(E \mid H1) * P(H1)}{P(E)}$$

$$P(H2 \mid E) = \frac{P(E \mid H2) * P(H2)}{P(E)}$$

same

$$\therefore \quad \boxed{P(Hi \mid E) = \alpha\, P(E \mid Hi)\, P(Hi)}$$

Bayesian Reasoning

- For multiple, conditionally independent evidences

$$P(E1, E2, \ldots En \mid Hi) = P(E1 \mid Hi)P(E2 \mid Hi) \ldots P(En \mid Hi)$$

$$\therefore P(Hi \mid E1, E2, \ldots En)$$
$$= \alpha\, P(E1, E2, \ldots En \mid Hi)\, P(Hi)$$
$$= \alpha\, P(E1 \mid Hi)\, P(E2 \mid Hi) \ldots P(En \mid Hi)\, P(Hi)$$
$$= \alpha\, P(Hi) \underbrace{\prod_{i=1..n} P(Ei \mid Hi)}$$

These probabilities are obtained from sample data

Bayesian Classifier Example

No.	age	income	student	credit_rating	buys_computer
1	<=30	high	no	fair	no
2	<=30	high	no	excellent	no
3	31…40	high	no	fair	yes
4	>40	medium	no	fair	yes
5	>40	low	yes	fair	yes
6	>40	low	yes	excellent	no
7	31…40	low	yes	excellent	yes
8	<=30	medium	no	fair	no
9	<=30	low	yes	fair	yes
10	>40	medium	yes	fair	yes
11	<=30	medium	yes	excellent	yes
12	31…40	medium	no	excellent	yes
13	31…40	high	yes	fair	yes
14	>40	medium	no	excellent	no

Bayesian Classifier Example

E X: (age=“<30”, income=“medium”,
student=“yes”, credit=“fair”) → yes / no ? H

- P(Yes | X) = α P(Yes) P(X|Yes)
  = α P(Yes) P(<30|Yes) P(m|Yes) P(y|Yes) P(f|Yes)
  = α x 9/14 x 2/9 x 4/9 x 6/9 x 6/9 = 0.028 α
- P(No | X) = α P(No) P(X|No)
  = α P(No) P(<30|No) P(m|No) P(y|No) P(f|No)
  = α x 5/14 x 3/5 x 2/5 x 1/5 x 2/5 = 0.007 α

⇒ X is classified to **yes**

Bayesian Classifier Example

- Category C1 : # of sample documents – 60
  Keywords - information (2), network (12), algorithm (6), system (10)
- Category C2 : # of sample documents – 40
  Keywords - information (10), database (8), algorithm (4), system (8)
- New document **D**: has keywords - **information, system**

P(C1 | information, system)
 = k * P(information | C1) * P(system | C1) * P(C1)
 = k * 2/30 * 10/30 * 60/100 = 0.013 k
P(C2 | information, system)
 = k * P(information | C2) * P(system | C2) * P(C2)
 = k * 10/30 * 8/30 * 40/100 = 0.036 k

→ D is classified to C2

Bayesian Classifier Example

- Breast cancer classification
  - Compute P(recur | age=40-49, menop=ge40, …)
  - Based on P(recur), P(age=40-49 | recur), P(menop=ge40 | recur), …

Viewer

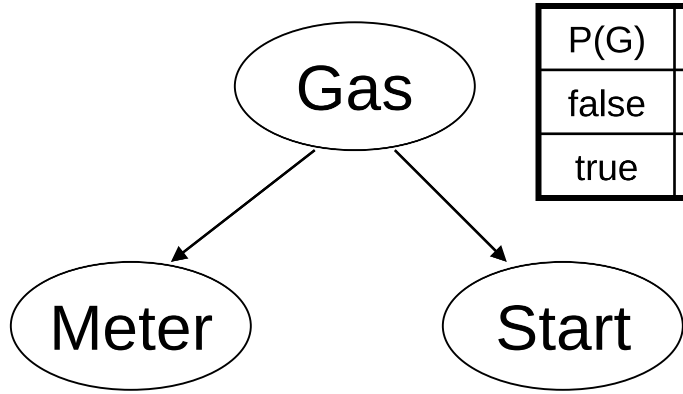
Relation: breast-cancer

No.	age Nominal	menopause Nominal	tumor-size Nominal	inv-nodes Nominal	node-caps Nominal	deg-malig Nominal	breast Nominal	breast-quad Nominal	irradiat Nominal	Class Nominal
1	40-49	premeno	15-19	0-2	yes	3	right	left_up	no	recurrence-events
2	50-59	ge40	15-19	0-2	no	1	right	central	no	no-recurrence-events
3	50-59	ge40	35-39	0-2	no	2	left	left_low	no	recurrence-events
4	40-49	premeno	35-39	0-2	yes	3	right	left_low	yes	no-recurrence-events
5	40-49	premeno	30-34	3-5	yes	2	left	right_up	no	recurrence-events
6	50-59	premeno	25-29	3-5	no	2	right	left_up	yes	no-recurrence-events
7	50-59	ge40	40-44	0-2	no	3	left	left_up	no	no-recurrence-events
8	40-49	premeno	10-14	0-2	no	2	left	left_up	no	no-recurrence-events
9	40-49	premeno	0-4	0-2	no	2	right	right_low	no	no-recurrence-events
10	40-49	ge40	40-44	15-17	yes	2	right	left_up	yes	no-recurrence-events
11	50-59	premeno	25-29	0-2	no	2	left	left_low	no	no-recurrence-events
12	60-69	ge40	15-19	0-2	no	2	right	left_up	no	no-recurrence-events
13	50-59	ge40	30-34	0-2	no	1	right	central	no	no-recurrence-events
14	50-59	ge40	25-29	0-2	no	2	right	left_up	no	no-recurrence-events
15	40-49	premeno	25-29	0-2	no	2	left	left_low	yes	recurrence-events
16	30-39	premeno	20-24	0-2	no	3	left	central	no	no-recurrence-events
17	50-59	premeno	10-14	3-5	no	1	right	left_up	no	no-recurrence-events
18	60-69	ge40	15-19	0-2	no	2	right	left_up	no	no-recurrence-events
19	50-59	premeno	40-44	0-2	no	2	left	left_up	no	no-recurrence-events
20	50-59	ge40	20-24	0-2	no	3	left	left_up	no	no-recurrence-events

Undo OK Cancel

Bayesian Belief Network

- Nodes: Random variables
- Edges: Direct influence
- Each node x stores P(x | parents(x))



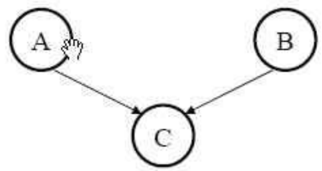
P(G)	
false	0.2
true	0.8

P(M|G)	G = f	G = t
empty	0.7	0.1
full	0.3	0.9

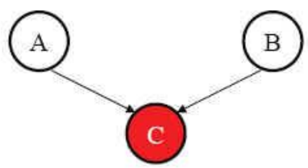
P(S|G)	G = f	G = t
no	0.99	0.30
yes	0.01	0.70

$\Rightarrow$ P(G,M,S) = P(G) P(M | G) P(S | G)

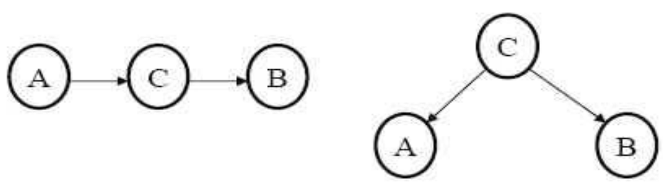
Dependency



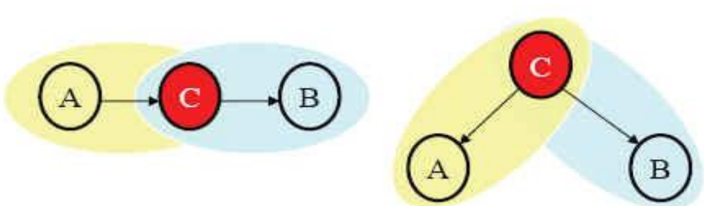
- A, B are independent
  - P(B|A) = P(B)



- A, B are *conditionally* dependent
  - P(B|A,C) $\neq$ P(B|C)



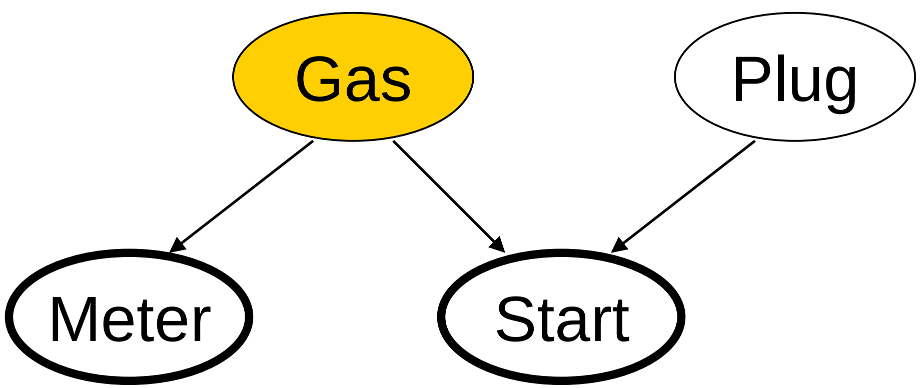
- A, B are dependent
  - P(B|A) $\neq$ P(B)



- A, B are *conditionally* independent
  - P(B|A,C) = P(B|C)

Dependency

- Gas and Start are dependent
- Gas and Plug are independent
- Gas and Plug are conditionally dependent given Start
- Meter and Start are conditionally independent given Gas
  - P(S | M, G) = P(S | G)



Chain Rule and Independence

- $P(A,B) = \frac{P(A,B)}{P(A)} P(A) = P(B|A)\ P(A)$
- $P(A,B,C) = \frac{P(A,B,C)}{P(A,B)} \frac{P(A,B)}{P(A)} P(A) = P(C|B,A)\ P(B|A)\ P(A)$
- $P(A,B,C,D) = P(D|C,B,A)\ P(C|B,A)\ P(B|A)\ P(A)$
- $P(G,M,S) = P(S \mid G,M)\ P(M \mid G)\ P(G)$ (Chain rule)

  $= P(S \mid G)\ P(M \mid G)\ P(G)$

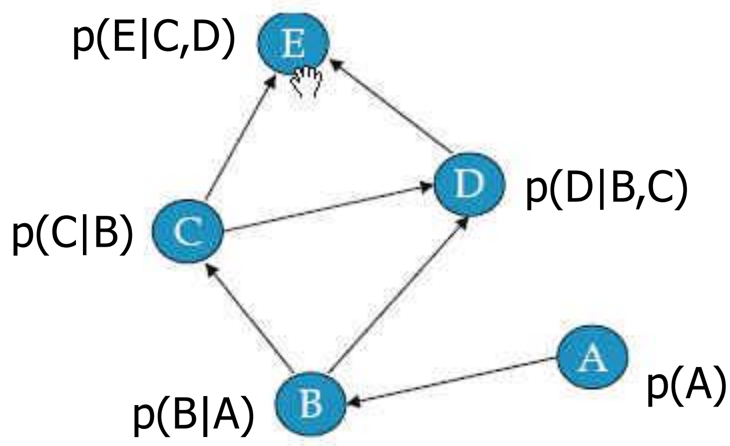
  (If S, M are conditionally independent given G)

  $= P(S)\ P(M)\ P(G)$

  (If S, M, G are all independent)

Computing Joint Distribution from Bayesian Network

- In general,
  Joint probability ← *product of conditional probabilities*
- Ex>



⇨ $P(A,B,C,D,E) = P(A)P(B|A)P(C|A,B)P(D|A,B,C)P(E|A,B,C,D)$
$= P(A)P(B|A)P(C|B) \quad P(D|B,C) \quad P(E|C,D)$

Bayesian Reasoning with Full Probability Distribution

Gas	Meter	Start	P(G,M,S)
false	empty	no	0.1386
false	empty	yes	0.0014
false	full	no	0.0594
false	full	yes	0.0006
true	empty	no	0.0240
true	empty	yes	0.0560
true	full	no	0.2160
true	full	yes	0.5040

- P(Gas=false, Meter=empty, Start=no) = 0.1386
- P(Gas=false) = 0.2
- P(Start=yes | Meter=full) = P(S=yes, M=full) / P(M=full)
  = (0.5040+0.0006) / (0.5040+0.0006+0.2160+0.0594)
  = 0.6469

Bayesian Reasoning with Full Probability Distribution

- P(Start=yes | Meter=full)

Gas	Meter	Start	P(G,M,S)
false	full	no	0.0594
false	full	yes	0.0006
true	full	no	0.2160
true	full	yes	0.5040

Select M=full

Gas	Start	P(G,M,S)
ALL	no	0.2754
ALL	yes	0.5046

Sum G

Start	P(G,M,S)
no	0.3531
yes	0.6469

Normalize

Reasoning with Bayesian Belief Network

- Inference for P(H | E)
  - From the product of probability table,
  1. Remove all rows except E
  2. Compute product
  3. Sum over irrelevant variables
  4. Normalize
- Example
  - P(S=yes | M=full)

Reasoning with Bayesian Belief Network

Remove M=empty

- p(S,G,M)

P(G)	
false	0.2
true	0.8

P(M|G)	G = f	G = t
~~empty~~	~~0.7~~	~~0.1~~
full	0.3	0.9

P(S|G)	G = f	G = t
no	0.99	0.30
yes	0.01	0.70

- p(S,G | M=f)

S	G = f	G = t
no	0.0594	0.2160
yes	0.0006	0.5040

Product

- p(S | M=f)

S	
no	0.2754
yes	0.5046

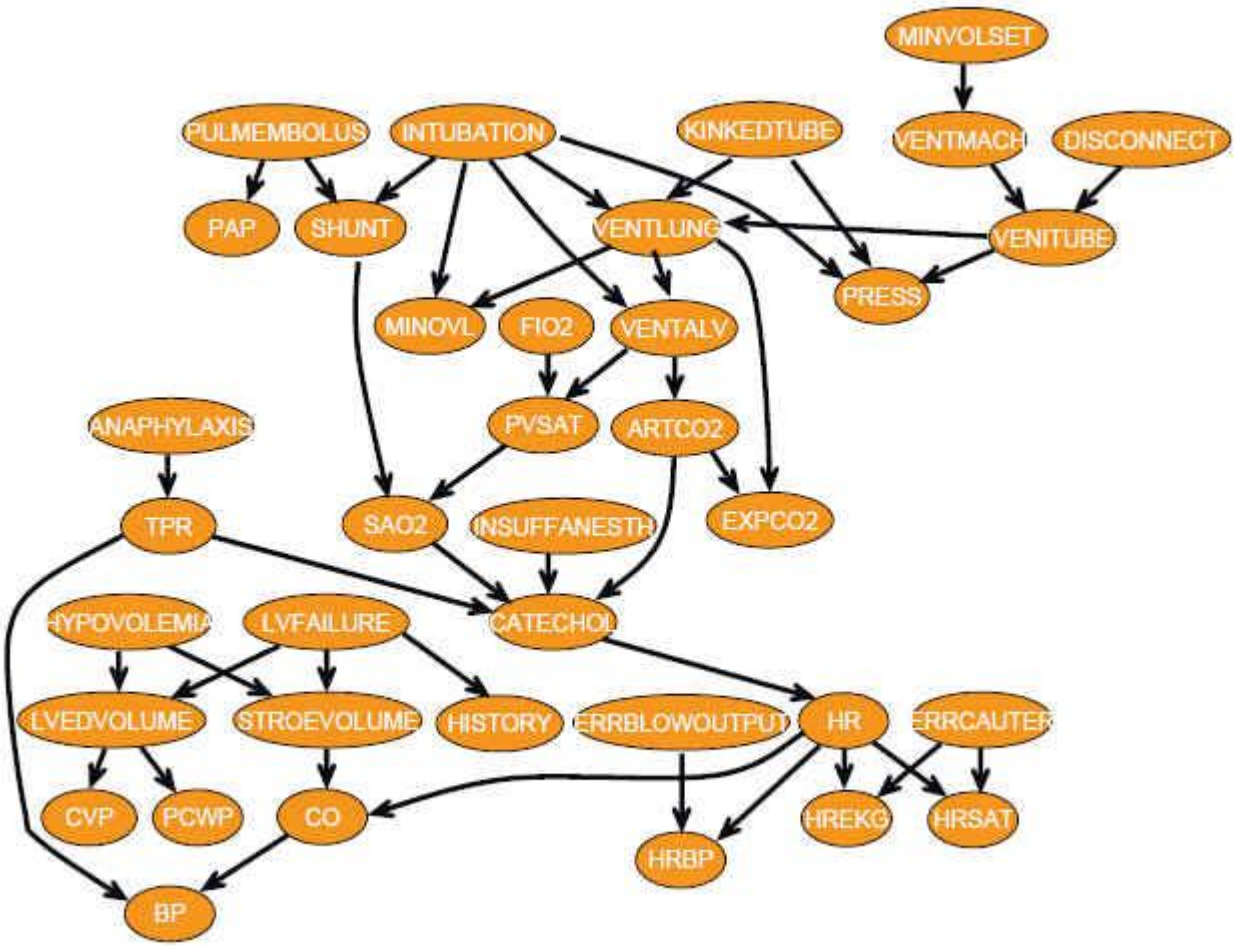
Sum over G

- p(S | M=f)

S	
no	0.3531
yes	**0.6469**

Normalize

Advantage of Reasoning with Bayesian Belief Network



Advantage of Reasoning with Bayesian Belief Network

- Assume 20 boolean variables: 19 E → 1 H
- Compute P(H | E1, E2, …, E19)

  $$= \frac{P(E1, E2, \ldots E19 \mid H)\, P(H)}{P(E1, E2, \ldots, E19)}$$

1) From full joint distribution

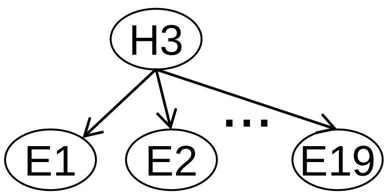
P(H, E1, E2, … E19)

→ We need to know 2^{20} = 1,048,576 prob.

⇨ Almost impossible

H	E1	E2	…	P(H,E1,E2,…)
T	T	T	…	0.xxxx
			…	…
F	F	F	…	0.xxxx

Advantage of Reasoning with Bayesian Belief Network

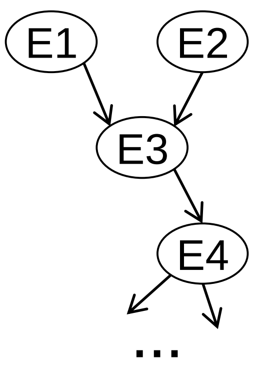


2) Assume complete independences (Naïve Bayesian)

P(E1 | H) **P**(E2 | H) … **P**(E19 | H) **P**(H)

→ 4*19 + 2 = 78 prob.

But the evidences are not actually independent



2) Consider actual dependences (Bayesian Network)

P(E1) **P**(E2) **P**(E3 | E1, E2) …

(assume less than 2 parents in Bayesian Network)

→ less than 8*20 = 160 prob.

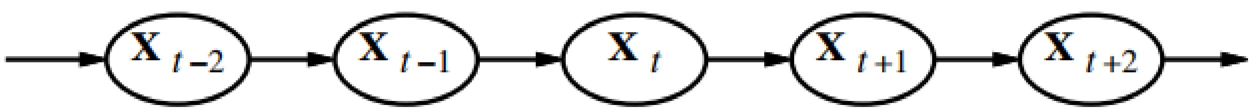
Actual dependences are considered,
yet need small # of probabilities.

Markov Model

- Markov process (Markov chain)
  - Probability of a state at time t depends on its previous n states

    $P(X_t \mid X_{t-1}, X_{t-2}, X_{t-3}, \ldots, X_{t-n})$
- First-order Markov process
  - Probability of a state at time t depends on its previous 1 state

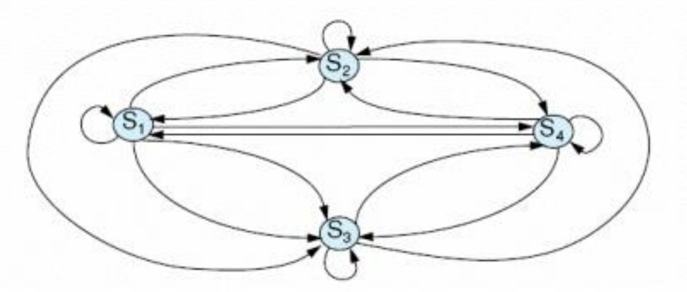
    $P(X_t \mid X_{t-1})$



Markov Model

- Example
  - 4 states: S1(sunny), S2(cloudy), S3(Foggy), S4(Rainy)

	s_1	s_2	s_3	s_4
s_1	0.4	0.3	0.2	0.1
s_2	0.2	0.3	0.2	0.3
s_3	0.1	0.3	0.3	0.3
s_4	0.2	0.3	0.3	0.2

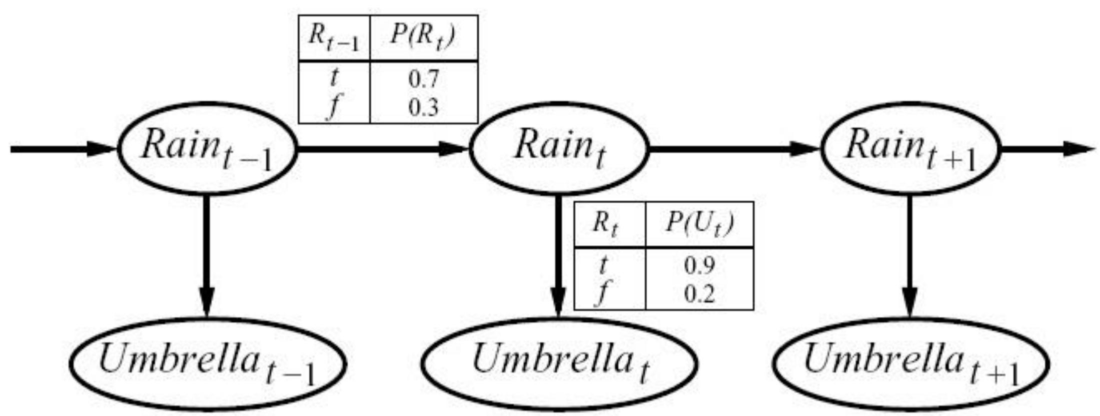


  - Today is sunny. Prob. of next 2 days are rainy?

$$P(S1, S4, S4) = P(S1)\ P(S4 \mid S1)\ P(S4 \mid S1,S4)$$
$$= P(S1)\ P(S4 \mid S1)\ P(S4 \mid S4)$$
$$= 1 * 0.1 * 0.2 = 0.02$$

Hidden Markov Model (HMM)

- HMM
  - States are “hidden”
  - Probability of observation is given. P(Oj | Si)



Hidden Markov Model (HMM)

$P(s1 \ldots sn \mid o1 \ldots on)$

$$= \frac{P(o1 \ldots on \mid s1 \ldots sn)\, P(s1 \ldots sn)}{P(o1 \ldots on)}$$

$= \boldsymbol{\alpha}\, P(o1 \ldots on \mid s1 \ldots sn)\, P(s1 \ldots sn)$

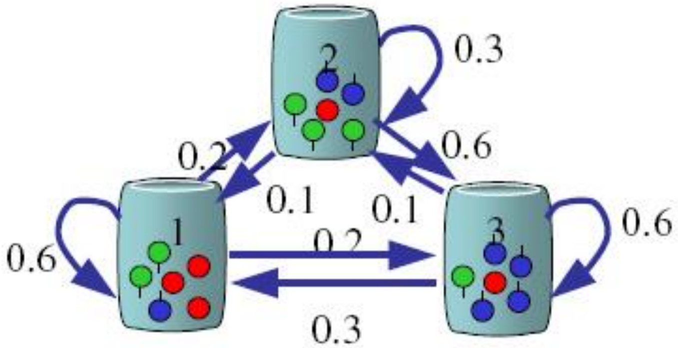
$= \boldsymbol{\alpha}\, P(o1 \mid s1) P(o2 \mid s2) \ldots P(on \mid sn)\, P(s1 \ldots sn)$

$= \boldsymbol{\alpha}\, P(o1 \mid s1) P(o2 \mid s2) \ldots P(on \mid sn)\, P(s1) P(s2 \mid s1) \ldots P(sn \mid sn\text{-}1)$

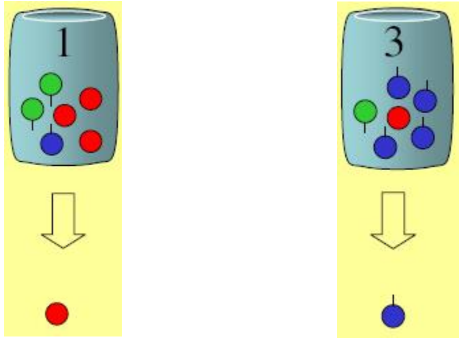
$= \prod_{i=1..n} P(oi \mid si)\, P(si \mid si\text{-}1)$

Hidden Markov Model (HMM)

- Example
  - Markov process

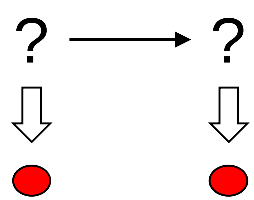


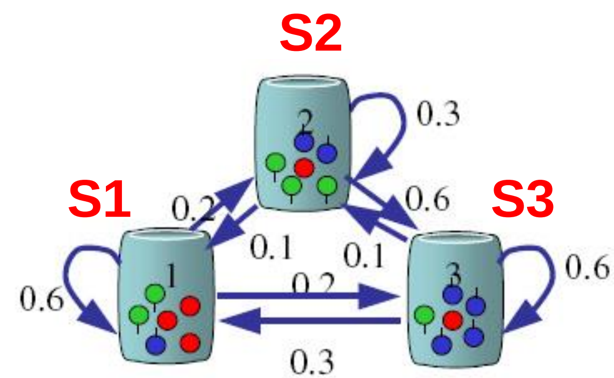
  - Output process



Hidden Markov Model (HMM)

- Which cup is selected? → hidden
- Only output sequence is observed





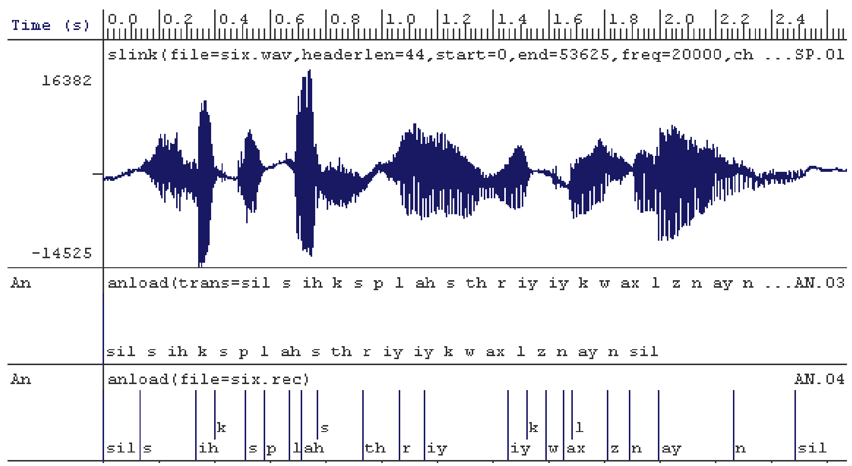
- Most likely sequence

  $= \text{argmax}_{X}\ P(X_0, X_1, \ldots, X_t \mid E_1, E_2, \ldots, E_t)$

  $= \text{argmax}_{X1,X2}$ (1/3*(**S1**)*3/6 * 0.6(**S1**)*3/6 ,
  1/3*(**S1**)*3/6 * 0.2(**S2**)*1/6 ,
  1/3*(**S1**)*3/6 * 0.2(**S3**)*1/6 ,
  1/3*(**S2**)*1/6 * 0.1(**S1**)*3/6 , **…**) = **S1, S1**

Speech Recognition

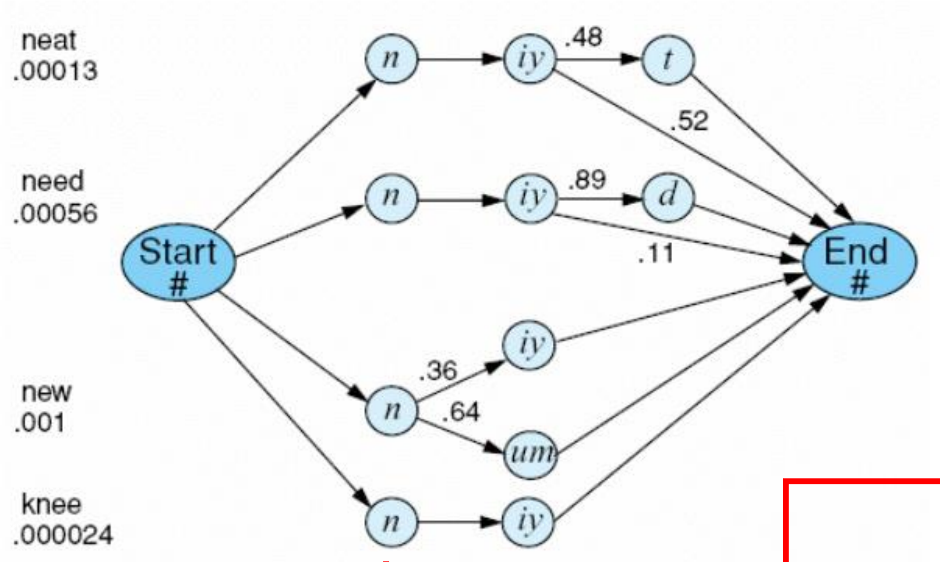
- The problem
  - Observed: sequence of acoustic signals
  - Determine: which phoneme? → which word?



➡ Compute P(phoneme | signal) by using HMM

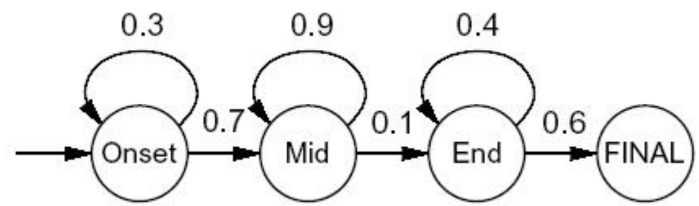
Find argmax P(*p1, p2, p3, ..* | *o1, o2, o3, ..*)
= argmax $\prod_{i=1..n}$ P(*oi* | *pi*) P(*pi* | *pi-1*)

Speech Recognition



state
(hidden)

Phone HMM for [m]:



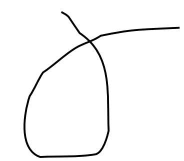
Output probabilities for the phone HMM:

Onset:	Mid:	End:
C1: 0.5	C3: 0.2	C4: 0.1
C2: 0.2	C4: 0.7	C6: 0.5
C3: 0.3	C5: 0.1	C7: 0.4

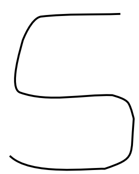
observation

Handwriting Recognition

- Hand-written character recognition



α or *6* ?



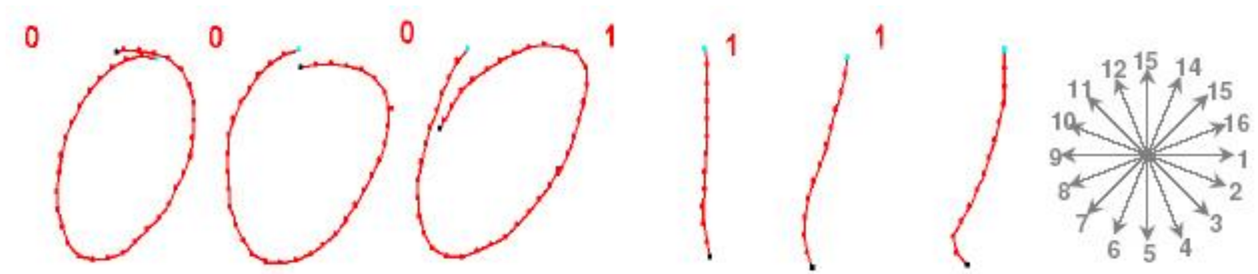
5 or *S* ?

Handwriting Recognition

- The problem
  - Observed: sequence of moving directions (d1.. dn)
  - Determine: which character? - states (s1.. sn)

$$\text{Find } \arg\max P(s1 \ldots sn \mid d1 \ldots dn)$$

$$= \arg\max \prod_{i=1..n} P(di \mid si)\, P(si \mid si\text{-}1)$$

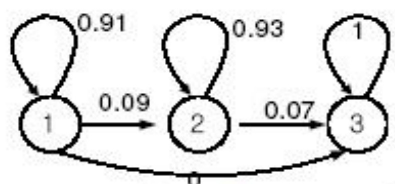


data0{1} = [9 8 8 7 7 7 6 6 6 5 5 5 4 4 3 2 1 16 15 15 15 15 14 14 14 13
13 13 12 12 11 10 9 9 8]

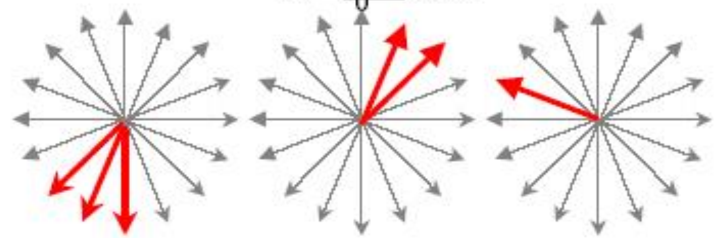
data1{2} = [5 6 6 6 6 6 6 6 6 5 5 4]

Handwriting Recognition

- Example
  - Writing: [8, 8, 7, 7, 7, 6, 6, 5, 5, …]
  - P(States of ‘zero’ | 8, 8, 7, 7, 7, 6, 6, 5, 5, …) >> P(States of ‘one’ | 8, 8, 7, 7, 7, 6, 6, 5, 5, …)
  - Markov process for ‘0’:



  - Output process:



0.00	0.00	0.00	0.00	0.30	0.33	0.21	0.12	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.09	0.07	0.07	0.11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.28	0.28	0.11
0.00	0.00	0.00	0.00	0.00	0.06	0.06	0.16	0.23	0.13	0.10	0.10	0.16	0.00	0.00	0.00