Probability-Foundations

Sargur Srihari
srihari@cedar.buffalo.edu

Topics

- Random Variables
- Joint and Conditional Distributions
- Independence and Conditional Independence

Random Variable

- We have a population of students
  - We want to reason about their grades
  - Random variable: *Grade*
  - *P(Grade)* associates a probability with each outcome $Val(Grade)=\{ A, B, C \}$
- If $k=|Val\{X\}|$ then

$$\boxed{\sum_{i=1}^{k} P(X = x^i) = 1}$$

  - Distribution is referred to as a *multinomial*
  - If $Val\{X\}=\{false,true\}$ then it is a *Bernoulli* distribution
- $P(X)$ is known as the marginal distributionof X

Joint Distribution

- We are interested in questions involving several random variables
  - Example event: $Intelligence=high$ and $Grade=A$
  - Need to consider joint distributions
  - Over a set $\chi=\{X_1,..,X_n\}$ denoted by $P(X_1,..,X_n)$
  - We use ξ to refer to a full assignment to variables χ, i.e. $\xi \varepsilon Val(\chi)$
- Example of joint distribution
  - And marginal distributions

		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

Conditional Probability

- $P(Intelligence|Grade=A)$ describes the distribution over events describable by Intelligence given the knowledge that student's grade is A
- It is not the same as the marginal distribution

		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

$P(Intelligence=high)=0.3$

$$P(Intelligence=high|Grade=A)$$
$$=0.18/0.25$$
$$=0.72$$

Independent Random Variables

- We expect $P(\alpha|\beta)$ to be different from $P(\alpha)$
  - i.e., β is true changes our probability over α
- Sometimes equality can occur, i.e, $P(\alpha|\beta)=P(\alpha)$
  - i.e., learning that β occurs did not change our probability of α
  - We say event α is independent of event β denoted $P \rightarrow (\alpha \perp \beta)$ if $P(\alpha|\beta)=P(\alpha)$ or if $P(\beta)=0$
- A distribution P satisfies $(\alpha \perp \beta)$ if and only if $P(\alpha \wedge \beta)=P(\alpha)P(\beta)$

Conditional Independence

- While independence is a useful property, we don't often encounter two independent events
- A more common situation is when two events are independent given an additional event
  - Reason about student accepted at Stanford or MIT
    - These two are not independent
      - If student admitted to Stanford then probability of MIT is higher
      - If both based on GPA and we know the GPA to be A
      - Then the student being admitted to Stanford does not change probability of being admitted to MIT
      - $P(MIT|Stanford,Grade\ A)=P(MIT|Grade\ A)$
      - i.e., MIT is conditionally independent of Stanford given Grade A

Querying Joint Probability Distributions

Query Types

1. Probability Queries
   - Given L and S give distribution of I
2. MAP Queries
   - Maximum a posteriori probability
   - Also called MPE (*Most Probable Explanation*)
     - What is the most likely setting of D,I, G, S, L
   - Marginal MAP Queries
     - When some variables are known

Probability Queries

- Most common type of query is a probability query
- Query has two parts
  - *Evidence*: a subset E of variables and their instantiation e
  - *Query Variables*: a subset Y of random variables in network
- Inference Task: $P(Y|E=e)$
  - *Posterior probability distribution* over values y of Y
  - *Conditioned* on the fact $E=e$
  - Can be viewed as Marginal over Y in distribution we obtain by conditioning on e
- Marginal Probability Estimation

$$P(Y = y_i \mid E = e) = \frac{P(Y = y_i, E = e)}{P(E = e)}$$

MAP Queries (Most Probable Explanation)

- Finding a high probability assignment to some subset of variables
- Most likely assignment to all non-evidence variables $W=\chi - Y$

$$MAP(W \mid e) = \arg\max_{w} P(w,e)$$

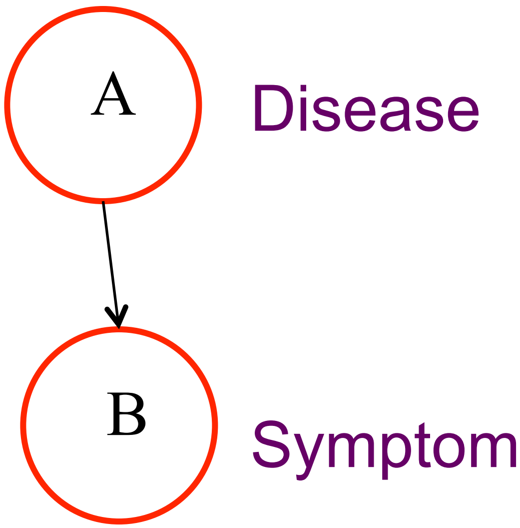
Value of w for which $P(w,e)$ is maximum

- Difference from probability query
  - Instead of a probability we get the most likely value for all remaining variables

Example of MAP Queries

P(Diseases)

a^0	a^1
0,4	0.6



P(Symptom|Disease)

$P(B\|A)$	b^0	b^1
a^0	0.1	0.9
a^1	0.5	0.5

Medical Diagnosis Problem

- Diseases (A) cause Symptoms (B)
- Two possible diseases
  - Mono and Flu
- Two possible symptoms
  - Headache, Fever

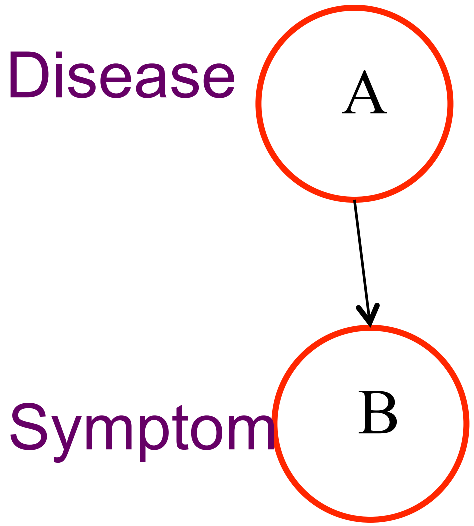
Q1: Most likely disease $P(A)$?
A1: Flu (Trivially determined for root node)

Q2: Most likely disease and symptom $P(A,B)$?
A2:

Q3: Most likely symptom P(B)?
A3:

Example of MAP Query

a^0	a^1
0,4	0.6



$P(B\|A)$	b^0	b^1
a^0	0.1	0.9
a^1	0.5	0.5

$$MAP(A) = \arg \max_{a} A = a^1$$

A1: Flu

$$MAP(A,B) = \arg \max_{a,b} P(A,B) = \arg \max_{a,b} P(A)P(B \mid A)$$

$$= \arg \max_{a,b} \{0.04, 0.36, 0.3, 0.3\} = a^0, b^1$$

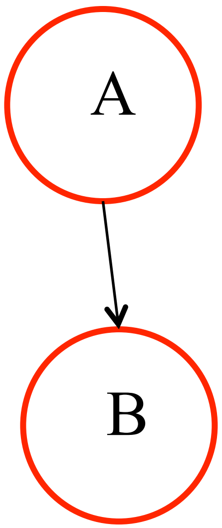
A2: Mono and Fever

Note that individually most likely value a^1 is not in the most likely joint assignment

Marginal MAP Query

Diseases

a^0	a^1
0,4	0.6



Symptoms

A	b^0	b^1
a^0	0.1	0.9
a^1	0.5	0.5

We looked for highest joint probability assignment of disease and symptom

Can look for most likely assignment of disease variable only

Query is not all remaining variables but a subset of them

Y is query, evidence is $E=e$

Task is to find most likely assignment to Y:

$MAP(Y|e)=arg\ max\ P(Y|e)$

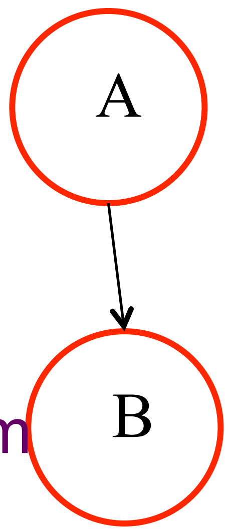
If $Z=X-Y-E$

$$MAP(Y \mid e) = \arg\max_Y \sum_Z P(Y, Z \mid e)$$

Example of Marginal MAP Query

a^0	a^1
0,4	0.6

Disease



Symptom

P(B|A)	b^0	b^1
a^0	0.1	0.9
a^1	0.5	0.5

$$MAP(A,B) = \arg\max_{a,b} P(A,B) = \arg\max_{a,b} P(A)P(B \mid A)$$

$$= \arg\max_{a,b} \{0.04, 0.36, 0.3, 0.3\} = a^0, b^1$$

A2: Mono and Fever

$$MAP(B) = \arg\max_{b} P(B) = \arg\max_{b} \sum_{A} P(A,B)$$

$$= \arg\max_{b} \{0.34, 0.66\} = b^1$$

A3: Fever

P(A,B)	b^0	b^1
a^0	0.04	0.36
a^1	0.3	0.3

Marginal MAP Assignments

- They are not monotonic
- Most likely assignment $MAP(Y_1|e)$ might be completely different from assignment to Y_1 in $MAP(\{Y_1,Y_2\}|e)$
  - Q1: Most likely disease $P(A)$?
  - A1: Flu
  - Q2: Most likely disease and symptom $P(A,B)$?
  - A2: Mono and Fever
- Thus we cannot use a MAP query to give a correct answer to a marginal map query

Marginal MAP more Complex than MAP

- Contains both summations (like in probability queries) and maximizations (like in MAP queries)

$$MAP(B) = \arg\max_{b} P(B) = \arg\max_{b} \sum_{A} P(A,B)$$

$$= \arg\max_{b} \{0.34, 0.66\} = b^1$$