CS 583: Probabilistic Graphical Models

Topic: MAP Inference

Mustafa Bilgic

http://www.cs.iit.edu/~mbilgic

https://twitter.com/bilgicm

Querying a distribution

- All variables: $\mathcal{X}$, evidence variables: $\boldsymbol{E}$
- **Probability query**
  - $P(\boldsymbol{Y} \mid \boldsymbol{e})$
- **MAP query**
  - $\boldsymbol{W} = \mathcal{X} \setminus \boldsymbol{E}$ (i.e., all the non-evidence variables)
  - $\text{MAP} \mid \boldsymbol{e} = \text{argmax}_{\boldsymbol{w}} P(\boldsymbol{w}, \boldsymbol{e})$
- **Marginal MAP query**
  - $\text{MAP}(\boldsymbol{Y} \mid \boldsymbol{e}) = \text{argmax}_{\text{y}} P(\boldsymbol{y} \mid \boldsymbol{e})$
  - Let $\boldsymbol{Z} = \mathcal{X} \setminus \boldsymbol{E} \cup \boldsymbol{Y}$
  - $\text{MAP}(\boldsymbol{Y} \mid \boldsymbol{e}) = \text{argmax}_{\boldsymbol{y}} \sum_{\boldsymbol{z}} P(\boldsymbol{z}, \boldsymbol{y} \mid \boldsymbol{e})$

Variable elimination for MAP

- Variable elimination works by multiplying and summing
- It's also called *sum-product* algorithm
- We can use the same technique for MAP, if we replace the sum operator with a max operator
- The algorithm is called *max-product*
- A few differences
  - Sum is replaced with max
  - All variables (except evidence) are eliminated
  - We need to trace back our steps to find the MAP assignment

EXAMPLES

- Two variables
- Student network
- Markov network

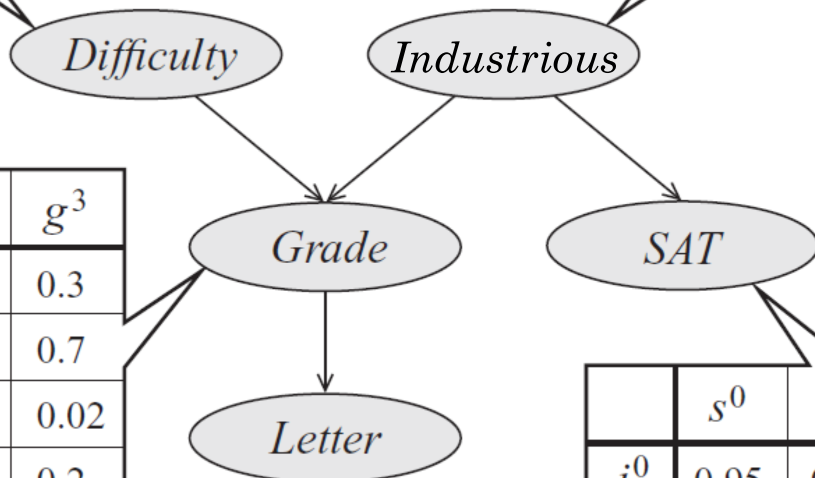
Two variables

- Network
  - $A \rightarrow B$
- Parameters
  - $\mathrm{P}(A) = [0.4; 0.6]$
  - $\mathrm{P}(B \mid A{=}t) = [0.8; 0.2]$
  - $\mathrm{P}(B \mid A{=}f) = [0.3; 0.7]$
- *MAP ?*
- *MAP|A=t ?*
- *MAP|B=t ?*

MAP ON THE STUDENT NETWORK

d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3



	g^1	g^2	g^3
i^0, d^0	0.3	0.4	0.3
i^0, d^1	0.05	0.25	0.7
i^1, d^0	0.9	0.08	0.02
i^1, d^1	0.5	0.3	0.2

	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

	l^0	l^1
g^1	0.1	0.9
g^2	0.4	0.6
g^3	0.99	0.01

1. MAP ?
2. MAP | $G=g^1$?

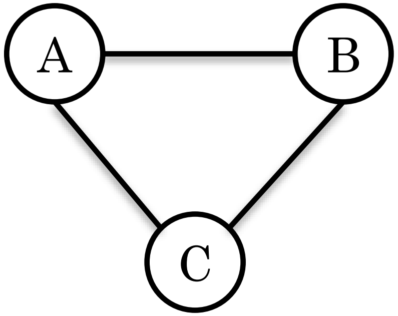
What if MAP is not unique?

- Network
  - $A \rightarrow B$
- Parameters – I
  - $\mathrm{P}(A) = [0.5; 0.5]$
  - $\mathrm{P}(B \mid A{=}t) = [0.4; 0.6]$
  - $\mathrm{P}(B \mid A{=}f) = [0.6; 0.4]$
- Parameters – II
  - $\mathrm{P}(A) = [0.4; 0.6]$
  - $\mathrm{P}(B \mid A{=}t) = [0.25; 0.75]$
  - $\mathrm{P}(B \mid A{=}f) = [0.5; 0.5]$

MAP on a Markov network

A	B	$\phi(A,B)$
T	T	5
T	F	1
F	T	1
F	F	5

A	$\phi(A)$
T	2
F	1



B	$\phi(B)$
T	1
F	4

A	C	$\phi(A,C)$
T	T	6
T	F	1
F	T	1
F	F	6

C	$\phi(C)$
T	1
F	8

B	C	$\phi(B,C)$
T	T	1
T	F	10
F	T	10
F	F	1

Complexity

- All the arguments from the sum-product algorithm carry over
- Same variable ordering heuristics can be applied
- Message passing on a junction-tree structure can be done

MARGINAL MAP

- **Marginal MAP query**
  - $\mathrm{MAP}(\boldsymbol{Y} \mid \boldsymbol{e}) = \mathrm{argmax}_y \mathrm{P}(\boldsymbol{y} \mid \boldsymbol{e})$
  - Let $\boldsymbol{Z} = \mathcal{X} \setminus \boldsymbol{E} \cup \boldsymbol{Y}$
  - $\mathrm{MAP}(\boldsymbol{Y} \mid \boldsymbol{e}) = \mathrm{argmax}_y \sum_z \mathrm{P}(\boldsymbol{z}, \boldsymbol{y} \mid \boldsymbol{e})$
- We need to sum out $\boldsymbol{Z}$
- Max-sum-product algorithm
  - First, eliminate $\boldsymbol{Z}$ using sum-product algorithm
  - Then, find MAP for $\boldsymbol{Y}$ using max-product algorithm
- Unfortunately, max and sum cannot be interleaved
- The variables are partitioned into three disjoint sets: $\boldsymbol{E}$, $\boldsymbol{Z}$, $\boldsymbol{Y}$
- This partitioning limits which orders we can choose, as we can order only within $\boldsymbol{Z}$ and within $\boldsymbol{Y}$

$$\max_y \sum_z \text{joint}$$

$$\neq \sum_z \max_y$$