CSE 250A. Principles of AI

Probabilistic Reasoning and Decision-Making

Lecture 3 – Belief networks

Lawrence Saul
Department of Computer Science and Engineering
University of California, San Diego

Fall 2021

Outline

Review

- **Types of probabilities:**

$$\begin{array}{rl} P(X,Y) & \text{joint} \\ P(Y|X) & \text{conditional} \\ P(X) & \text{unconditional (or marginal)} \end{array}$$

- **Useful rules:**

$$\begin{array}{rcll} P(A,B,C,\ldots) & = & P(A)\,P(B|A)\,P(C|A,B)\ldots & \textbf{product rule} \\ P(X|Y) & = & P(Y|X)P(X)/P(Y) & \textbf{Bayes rule} \\ P(X) & = & \sum_y P(X,Y=y) & \textbf{marginalization} \end{array}$$

- **Conditioning on background evidence E:**

$$\begin{array}{rcl} P(A,B,C,\ldots|E) & = & P(A|E)\,P(B|A,E)\,P(C|A,B,E)\ldots \\ P(X|Y,E) & = & P(Y|X,E)P(X|E)/P(Y|E) \\ P(X|E) & = & \sum_y P(X,Y=y|E) \end{array}$$

Marginal and conditional independence

- **Marginal independence**

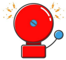
$$\left.\begin{array}{rcl} P(X|Y) & = & P(X) \\ P(Y|X) & = & P(Y) \\ P(X,Y) & = & P(X)\,P(Y) \end{array}\right\} \quad \textit{Each of these implies the other two.}$$

- **Conditional independence**

$$\left.\begin{array}{rcl} P(X|Y,E) & = & P(X|E) \\ P(Y|X,E) & = & P(Y|E) \\ P(X,Y|E) & = & P(X|E)\,P(Y|E) \end{array}\right\} \quad \textit{Each of these also implies the other two.}$$

HW 1

Example of conditional dependence

- **B and E are marginally independent:**

$$\begin{aligned} P(B) &= P(B|E) \\ P(E) &= P(E|B) \\ P(B,E) &= P(B)\,P(E) \end{aligned}$$

- **But B and E are conditionally dependent given A:**

$$\begin{aligned} P(B|A) &\neq P(B|E,A) \\ P(E|A) &\neq P(E|B,A) \\ P(B,E|A) &\neq P(B|A)\,P(E|A) \end{aligned}$$

Motivation

- **Model complexity**

  Suppose $X_i \in \{0,1\}$ are binary random variables.

  Then it requires $O(2^n)$ numbers to specify the joint distribution $P(X_1 = x_1, \ldots, X_n = x_n)$.

- **Conceptual and practical goals**

  Can we develop more *compact representations*?

  Can we develop more *efficient algorithms*?

Outline

1. Review
2. $\boxed{\text{Alarm example}}$
3. Belief networks
4. Homework demo

Alarm example




- **Binary random variables**

$$
\begin{array}{ll}
B \in \{0,1\} & \text{Was there a burglary?} \\
E \in \{0,1\} & \text{Was there an earthquake?} \\
A \in \{0,1\} & \text{Was the alarm triggered?} \\
J \in \{0,1\} & \text{Did John call?} \\
M \in \{0,1\} & \text{Did Mary call?}
\end{array}
$$

Joint distribution

- **Product rule**

$$P(B,E,A,J,M)$$

$$= P(B)\,P(E|B)\,P(A|B,E)\,P(J|B,E,A)\,P(M|B,E,A,J)$$

$\boxed{\textit{Note: the above is true no matter what the variables signify.}}$

- **Domain-specific assumptions**

$$\begin{aligned} P(E|B) &= P(E) && \textbf{marginal independence} \\ P(J|B,E,A) &= P(J|A) && \textbf{conditional independence} \\ P(M|B,E,A,J) &= P(M|A) && \textbf{conditional independence} \end{aligned}$$

Completing the model

- **Joint distribution**

$$
\begin{aligned}
& P(B, E, A, J, M) \\
& \quad = \quad P(B)\, P(E|B)\, P(A|B, E)\, P(J|B, E, A)\, P(M|B, E, A, J) \\
& \quad = \quad P(B)\, P(E)\, P(A|B, E)\, P(J|A)\, P(M|A)
\end{aligned}
$$

- **Conditional probability tables (CPTs)**

$\boxed{P(B=1) = 0.001}$

$\boxed{P(E=1) = 0.002}$

B	E	$P(A=1 \mid B, E)$
0	0	0.001
1	0	0.94
0	1	0.29
1	1	0.95

A	$P(J=1 \mid A)$
0	0.05
1	0.9

A	$P(M=1 \mid A)$
0	0.01
1	0.7

Inference

- **Joint probabilities are easy to compute:**

$$P(B\!=\!1, E\!=\!0, A\!=\!1, J\!=\!1, M\!=\!1)$$
$$= P(B\!=\!1)\, P(E\!=\!0)\, P(A\!=\!1|B\!=\!1, E\!=\!0)\, P(J\!=\!1|A\!=\!1)\, P(M\!=\!1|A\!=\!1)$$
$$= (0.001)(1-0.002)(0.94)(0.9)(0.7)$$

- **Any inference can be expressed in terms of joint probabilities:**

$$P(B\!=\!1, E\!=\!0|M\!=\!1)$$
$$= \frac{P(B\!=\!1, E\!=\!0, M\!=\!1)}{P(M\!=\!1)} \quad \boxed{\textbf{product rule}}$$
$$= \frac{\sum_{a,j} P(B\!=\!1, E\!=\!0, A\!=\!a, J\!=\!j, M\!=\!1)}{\sum_{b',e',a',j'} P(B\!=\!b', E\!=\!e', A\!=\!a', J\!=\!j', M\!=\!1)} \quad \boxed{\textbf{marginalization}}$$

But this approach can be very inefficient!

Efficient inference

How to perform inference most efficiently?

1. Visualize models as directed acyclic graphs.

2. Exploit graph structure to organize and simplify calculations.

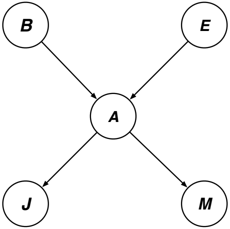
We'll spend today on (1) and next week on (2).

Visualizing the model

- **Joint distribution**

$$
\begin{aligned}
& P(B,E,A,J,M) \\
& \quad = \quad P(B)\,P(E|{\color{red}B})\,P(A|B,E)\,P(J|{\color{red}B},{\color{red}E},A)\,P(M|{\color{red}B},{\color{red}E},A,{\color{red}J}) \\
& \quad = \quad P(B)\,P(E)\,P(A|B,E)\,P(J|A)\,P(M|A)
\end{aligned}
$$

- **Directed acyclic graph (DAG)**

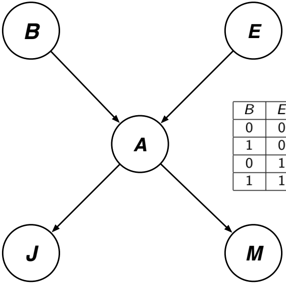


Absent edges encode assumptions of independence.

Alarm belief network

$P(B=1) = 0.001$

$P(E=1) = 0.002$



B	E	$P(A=1 \mid B, E)$
0	0	0.001
1	0	0.94
0	1	0.29
1	1	0.95

A	$P(J=1 \mid A)$
0	0.05
1	0.9

A	$P(M=1 \mid A)$
0	0.01
1	0.7

This visual representation of the joint distribution is called a **belief network** (or a **Bayesian network**).

Outline

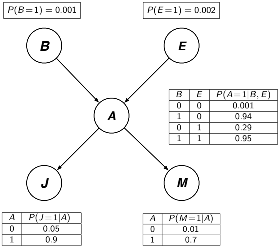
1. Review
2. Alarm example
3. Belief networks
4. Homework demo

Definition

A **belief network** (BN) is a directed acyclic graph (DAG) in which:

1. Nodes represent random variables.
2. Edges represent dependencies.
3. Conditional probability tables (CPTs) describe how each node depends on its parents.

$$\boxed{\textbf{BN} = \textbf{DAG} + \textbf{CPTs}}$$



From distributions to graphs

- **It is always true from the product rule that**

$$\begin{aligned} P(X_1, X_2, \ldots, X_n) &= P(X_1)\, P(X_2|X_1)\, \ldots\, P(X_n|X_1, \ldots, X_{n-1}) \\ &= \prod_{i=1}^{n} P(X_i|X_1, X_2, \ldots, X_{i-1}) \end{aligned}$$

- **But suppose in a particular domain that**

$$P(X_i|X_1, X_2, \ldots, X_{i-1}) = P(X_i|\mathrm{parents}(X_i)),$$

where $\mathrm{parents}(X_i)$ **is a subset of** $\{X_1, \ldots, X_{i-1}\}$**.**

- **Big idea: represent conditional dependencies by a DAG.**

Constructing a belief network

Three steps:

1. Choose your random variables of interest.
2. Choose an ordering of these variables (e.g., $X_1, X_2, \ldots, X_n$).
3. While there are variables left:
   - **(a)** add the node X_i to the network
   - **(b)** set the parents of X_i to be the minimal subset satisfying

$$P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | \mathrm{parents}(X_i)),$$

   - **(c)** define the conditional probability table $P(X_i | \mathrm{parents}(X_i))$

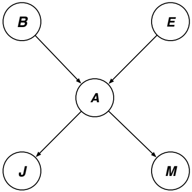
Node ordering

- **Best ordering:**

  Add the “root causes,” then the variables they influence, then the next variables that are influenced, etc.

- **Example:**

  In the alarm world, a natural ordering is (B, E, A, J, M).



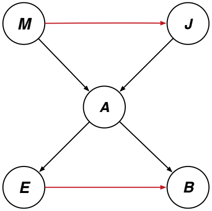
Node ordering

- **What happens if we choose an unnatural ordering?**

  Ex: (M, J, A, E, B)

- **Adding nodes with this ordering:**

$$
\begin{aligned}
& P(M, J, A, E, B) \\
& \quad = \quad P(M)\, P(J|M)\, P(A|J, M)\, P(E|A, J, M)\, P(B|E, A, J, M) \\
& \quad = \quad P(M)\, P(J|M)\, P(A|J, M)\, P(E|A)\, P(B|A, E)
\end{aligned}
$$



This belief network has **two extra edges**.
This DAG does not show $P(B) = P(B|E)$.
This DAG does not show $P(M|A) = P(M|A, J)$.
This belief network has **larger CPTs**.
These CPTs may be more difficult to assess.

Advantages of belief networks

1. **Compact representation of complex models**

   BNs provide a complete but parsimonious representation of joint probability distributions.

2. **Crisp separation of qualitative vs quantitative knowledge**

Qualitative	DAGs encode assumptions of marginal and conditional independence.
Quantitative	CPTs encode numerical influences of some variables on others.

Outline

1. **Review**
2. **Alarm example**
3. **Belief networks**
4. **Homework demo**

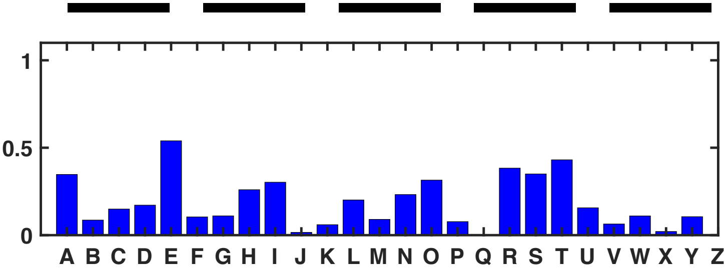
Homework demo

Can you guess the letters in a word?

_____ _____ _____ _____ _____

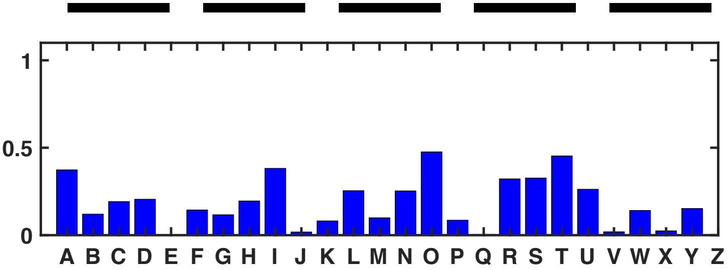
Word 1

Best guess: E



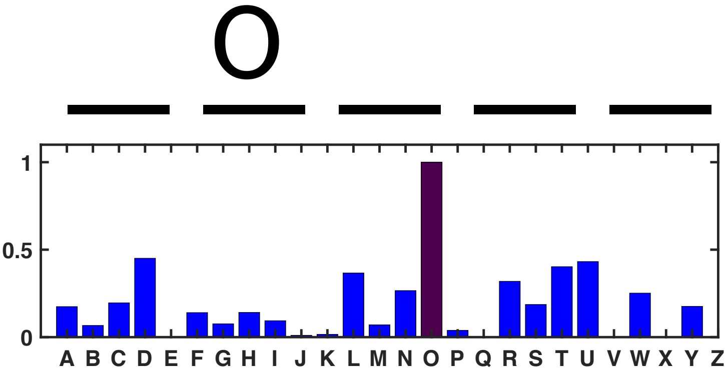
Word 1

Best guess: O



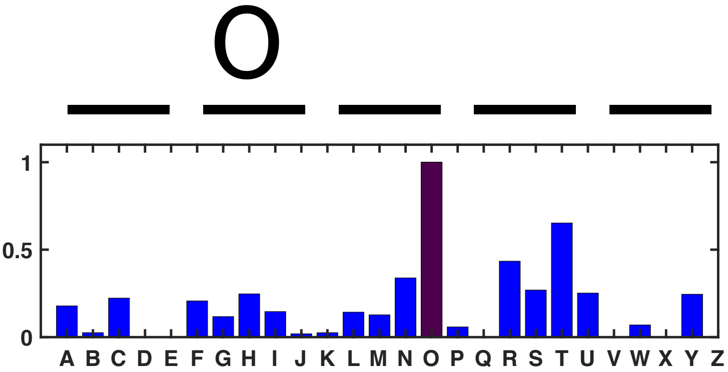
Word 1

Best guess: D



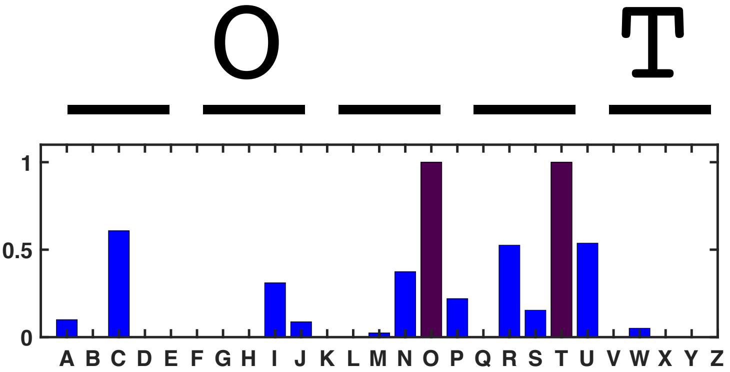
Word 1

Best guess: T



Word 1

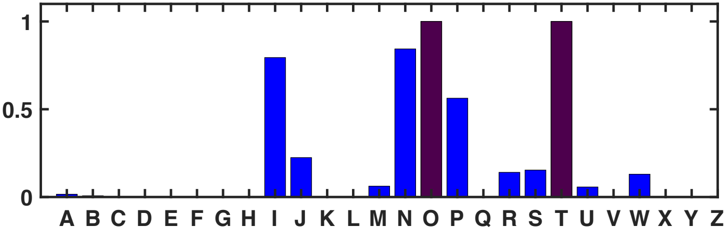
Best guess: C



Word 1

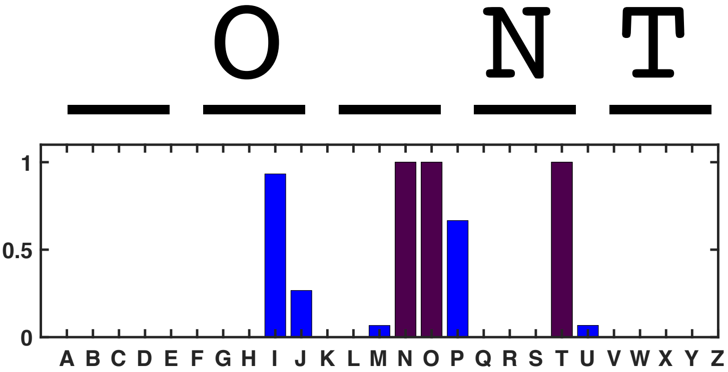
Best guess: N

_ O _ _ T



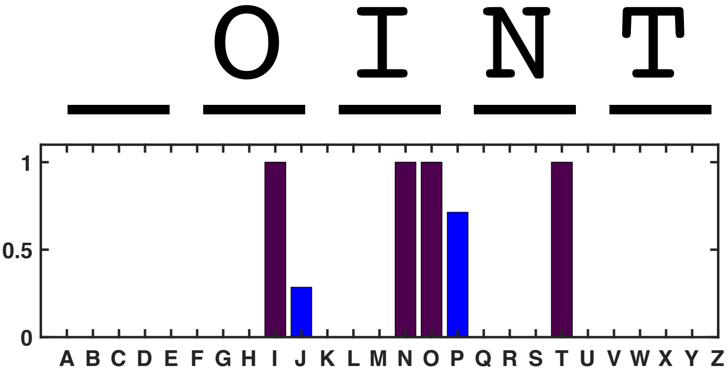
Word 1

Best guess: I



Word 1

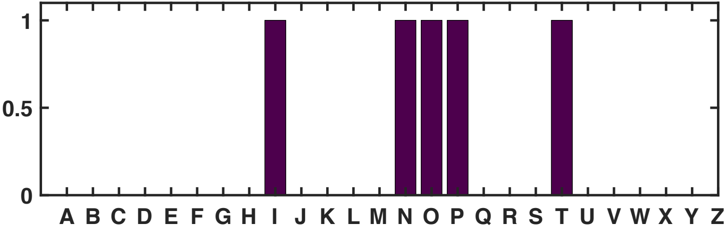
Best guess: P



Word 1

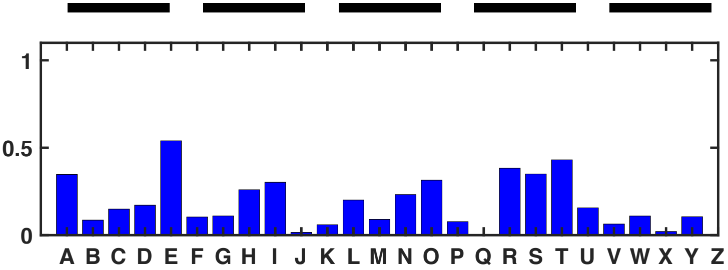
Done!

P O I N T



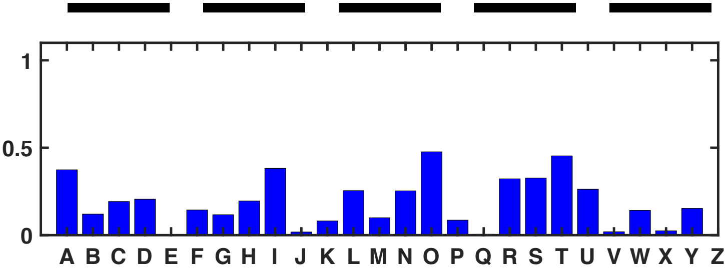
Word 2

Best guess: E



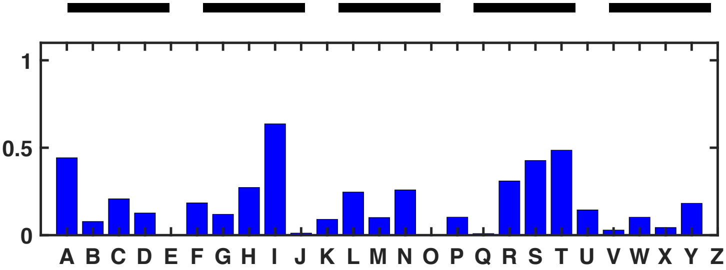
Word 2

Best guess: O



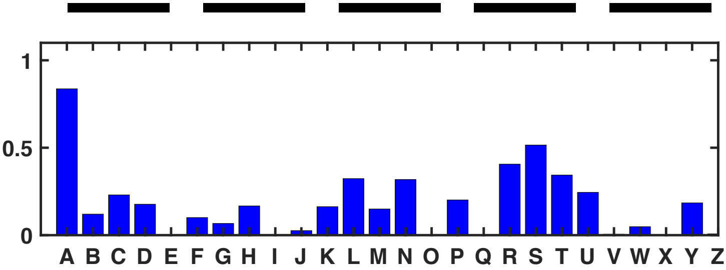
Word 2

Best guess: I



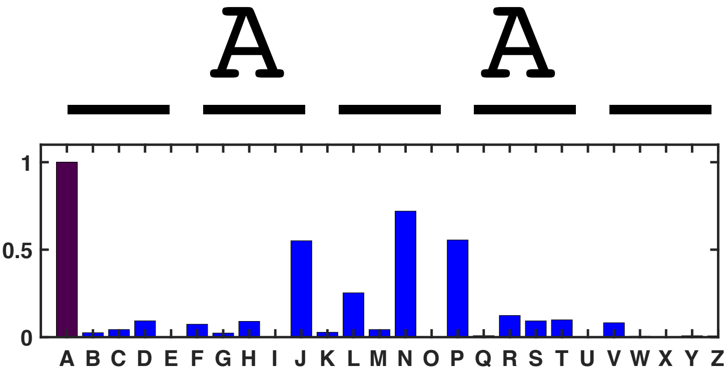
Word 2

Best guess: A



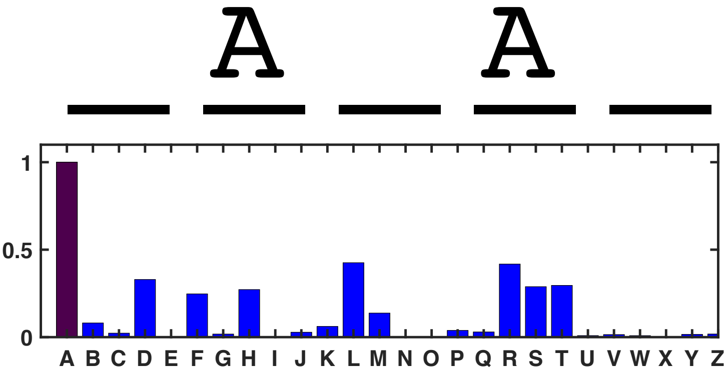
Word 2

Best guess: N



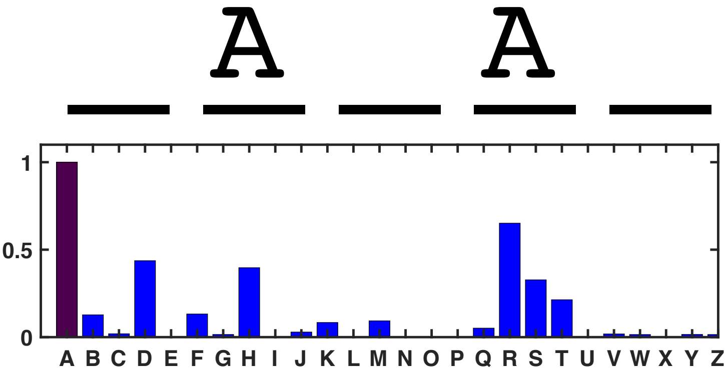
Word 2

Best guess: L



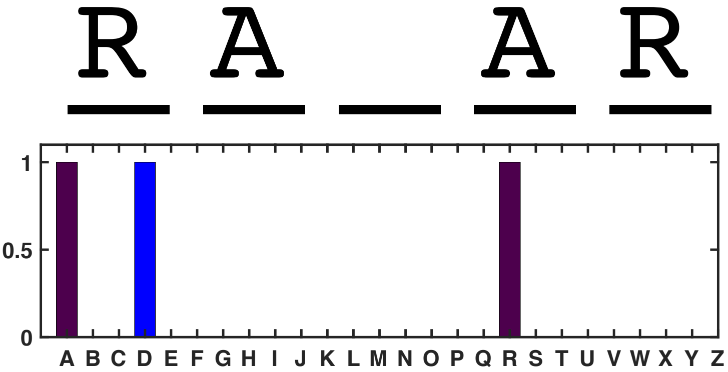
Word 2

Best guess: R



Word 2

Best guess: D



Word 2

Done!

R A D A R

Looking ahead

Homework 1 is due next Tuesday 3:30 pm (PT).

(There is a 48-hour no-questions-asked grace period.)

Next lecture: DAGs and conditional independence.