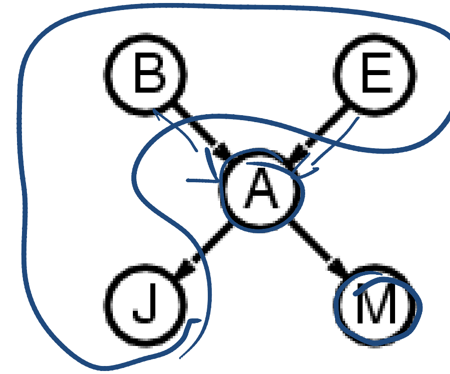
Reasoning Under Uncertainty: More on BNets structure and construction

Computer Science cpsc322, Lecture 28

(Textbook Chpt 6.3)

Nov, 15, 2013

Belief networks Recap



- By considering **causal dependencies**, we order variables in the joint.
- **Apply**…chain rule…………… and **simplify**

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

why M indep(B, E, J) given A $\quad P(M, B, E, J, A)$

- **Build a directed acyclic graph** (DAG) in which the parents of each var *X* are those vars on which *X* directly depends.
- **By construction**, a var is independent form it non-descendant given its parents. why?

Belief Networks: open issues

- **Independencies:** Does a BNet encode more independencies than the ones specified by construction? Yes

n vars

- **Compactness**: We reduce the number of probabilities from $O(2^n)$ to $O(n2^k)$ k parents

In some domains we need to do better than that!

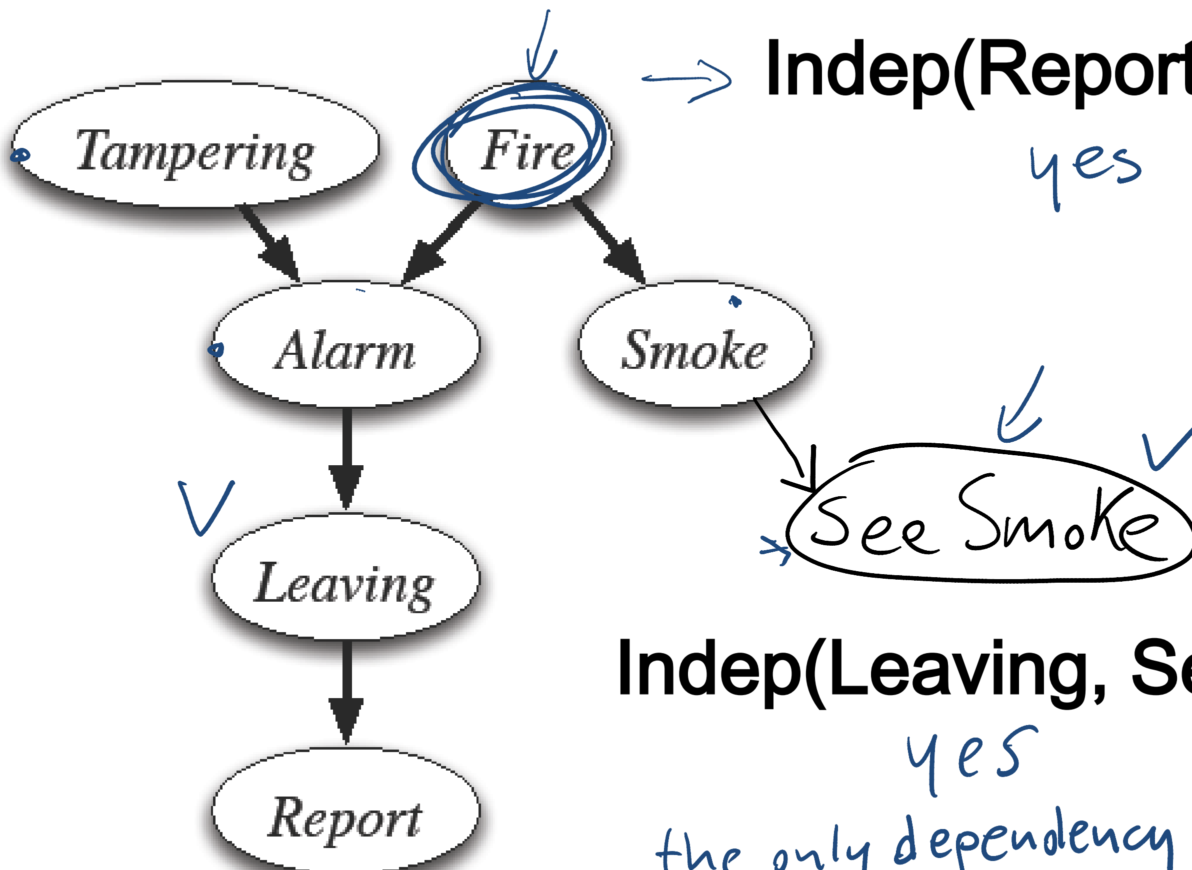
- Still too many and often there are no data/experts for accurate assessment

Solution: Make stronger (approximate) independence assumptions

Lecture Overview

- Implied Conditional Independence relations in a Bnet
- Compactness: Making stronger Independence assumptions
  - Representation of Compact Conditional Distributions
  - Network structure(Naïve Bayesian Classifier)

Bnets: Entailed (in)dependencies



Indep(Report, Fire,{Alarm})?

yes

If you know the value of Alarm, knowledge of the value of Fire does not affect your belief in the value of Report

Indep(Leaving, SeeSmoke,{Fire})?

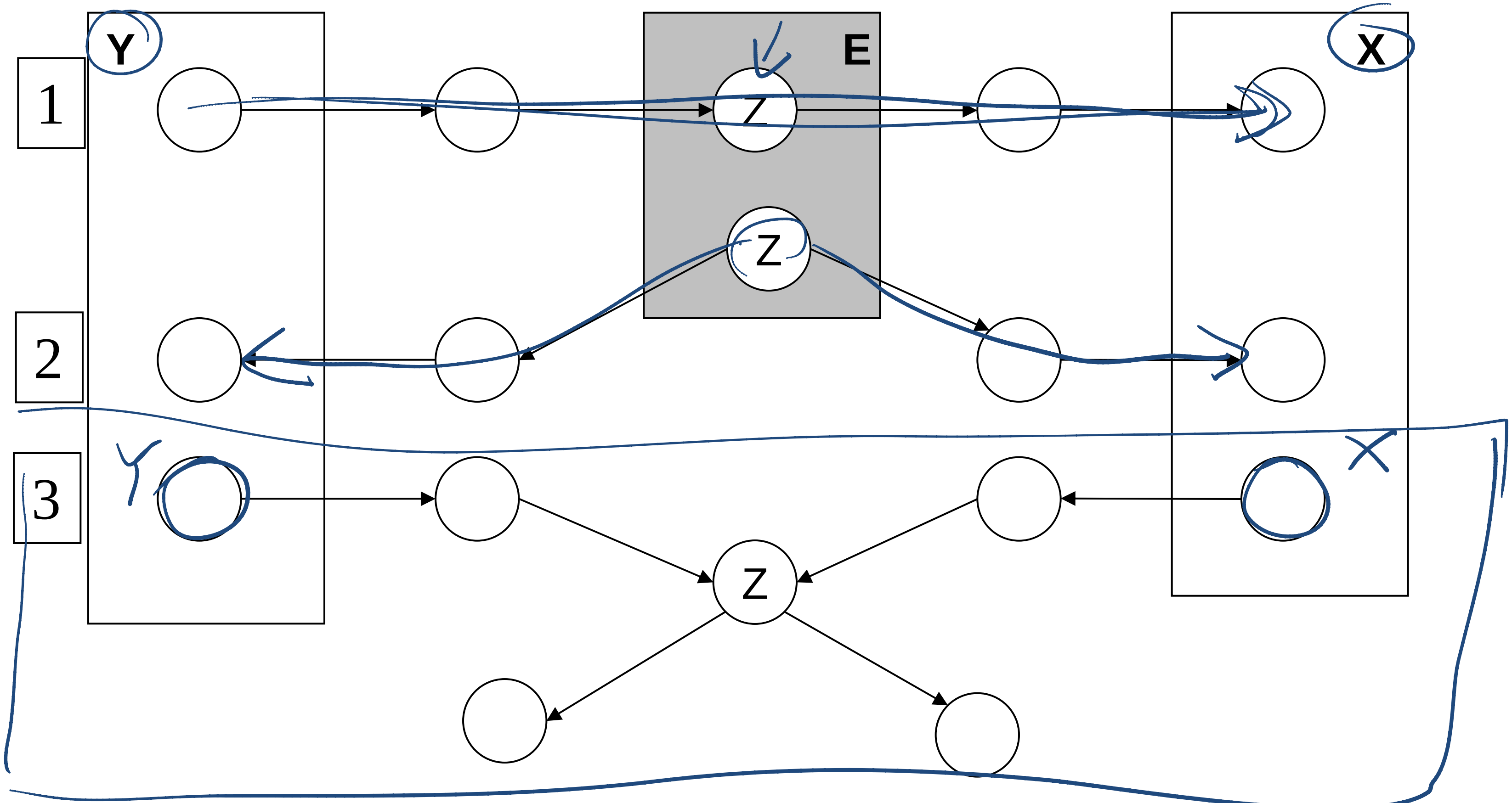
yes

the only dependency between Leaving and See Smoke is through Fire

If you know the value of Fire the two become independent

Conditional Independencies

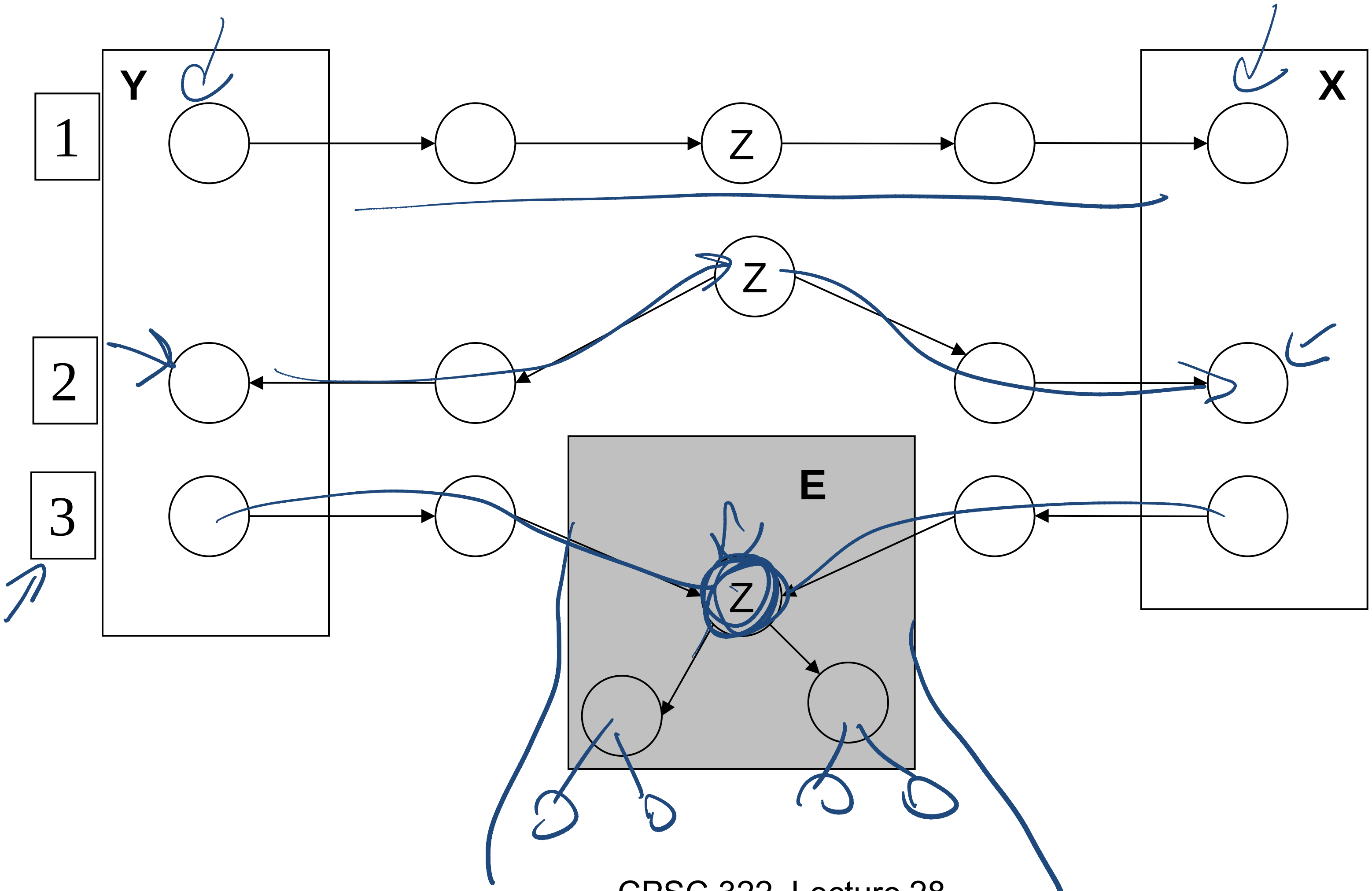
Or, blocking paths for probability propagation. Three ways in which a path between X to Y can be blocked, (1 and 2 given evidence E)



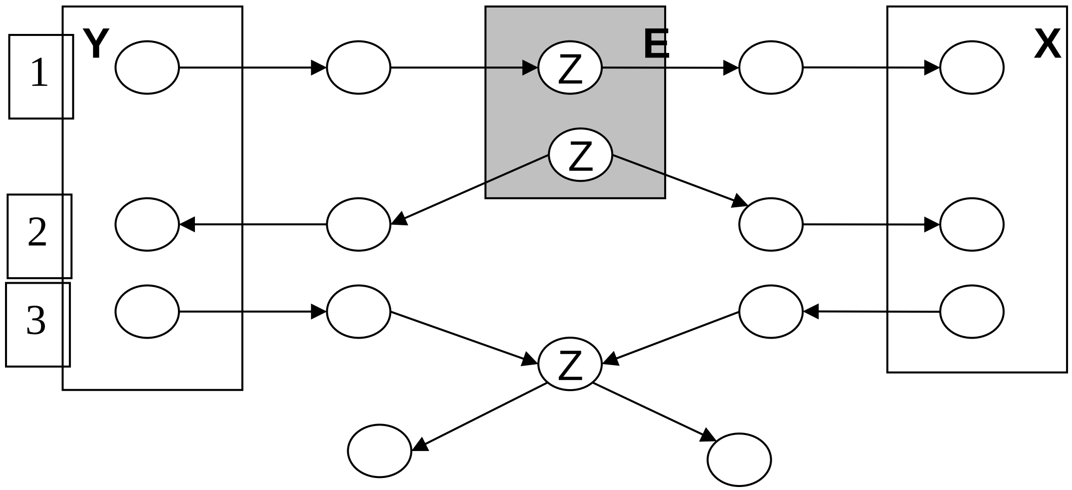
Note that, in 3, X and Y become dependent as soon as I get evidence on Z or on *any of its descendants*

Or ….Conditional Dependencies

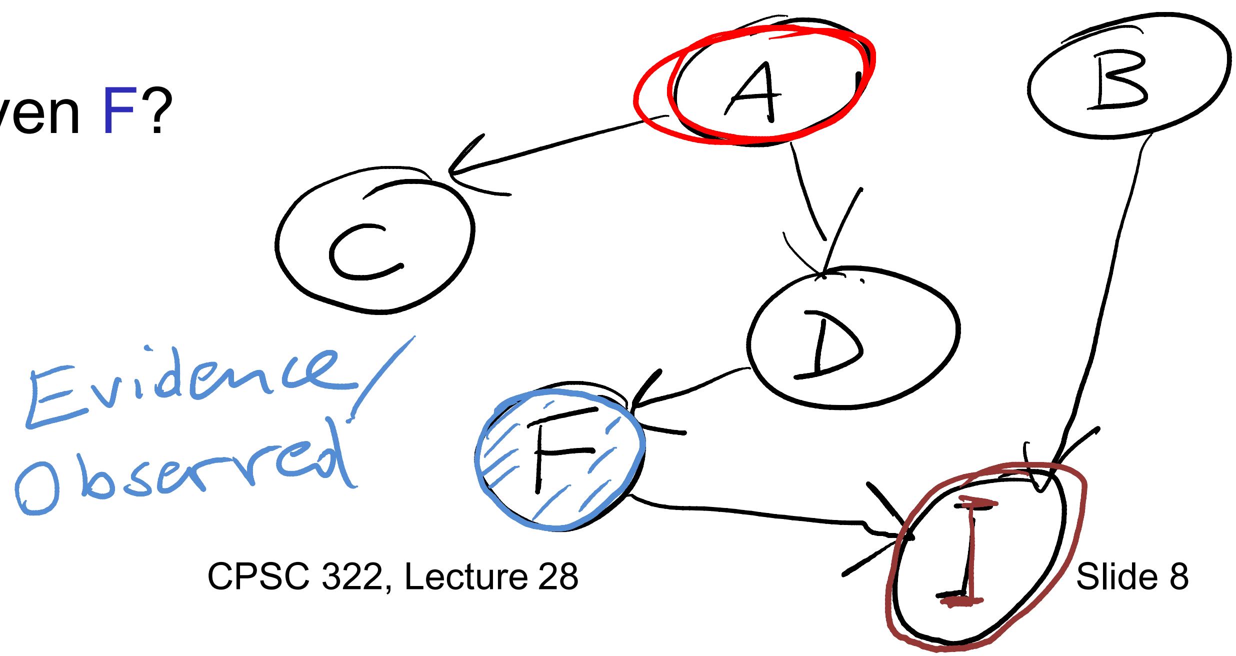
In 1, 2, 3 X Y are dependent



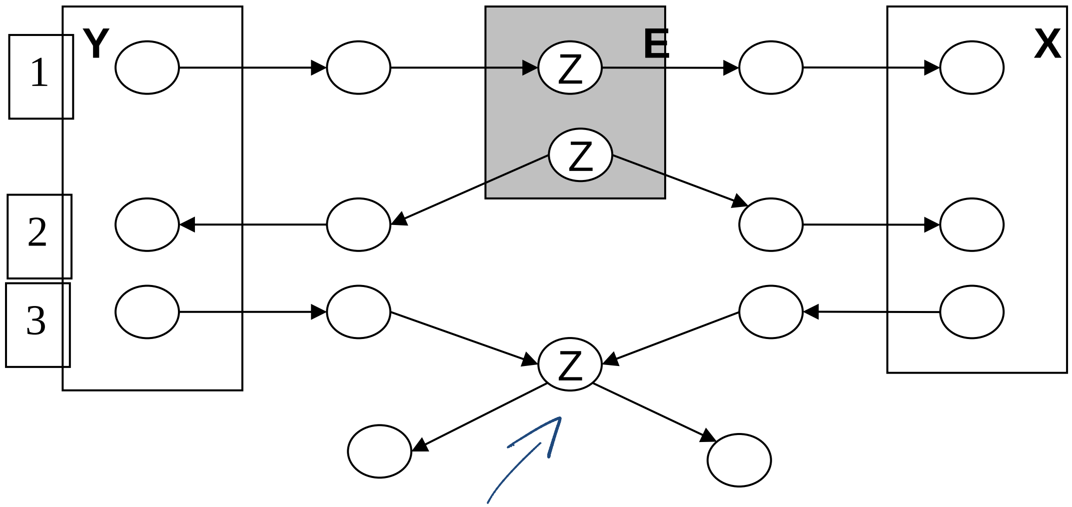
In/Dependencies in a Bnet : Example 1



Is A conditionally independent of I given F?

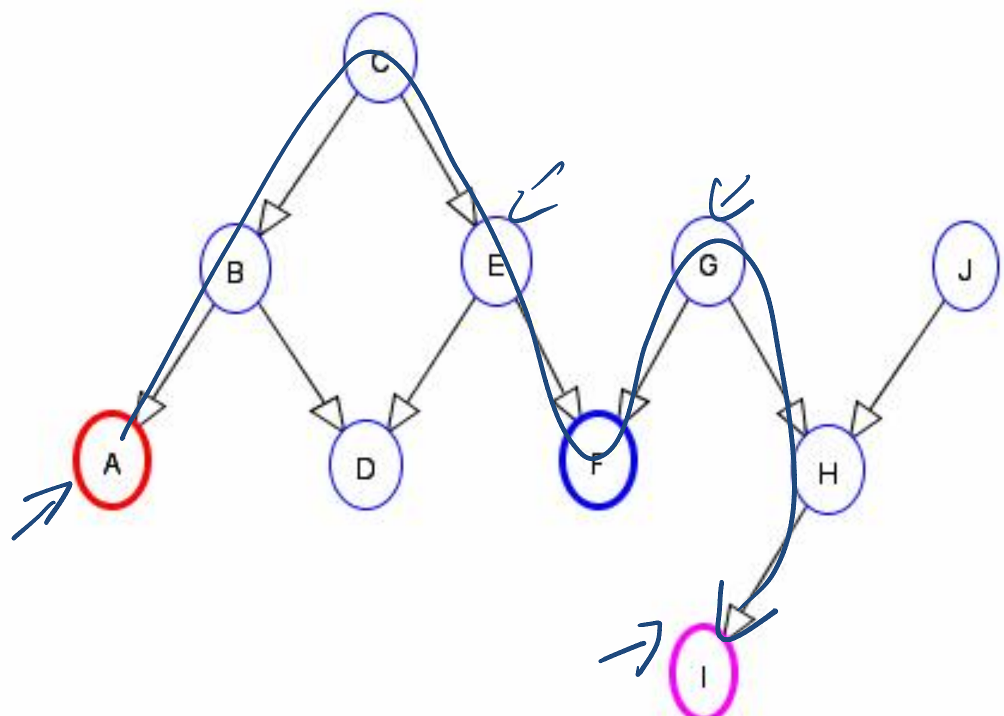


In/Dependencies in a Bnet : Example 2



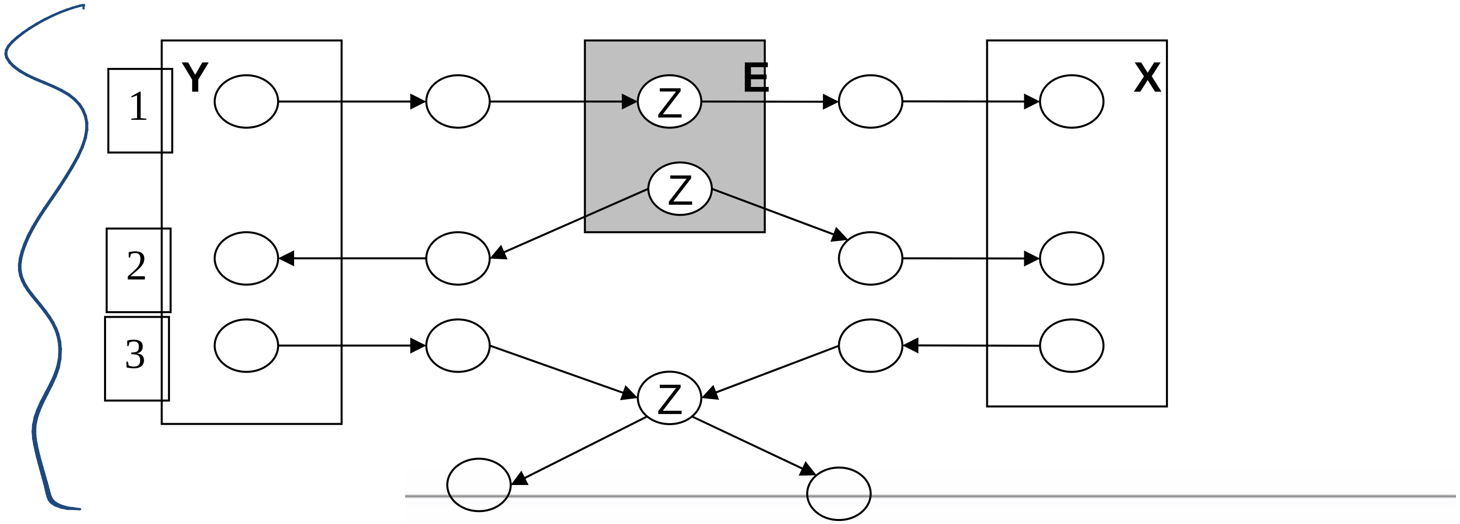
Is A conditionally independent of I given F?

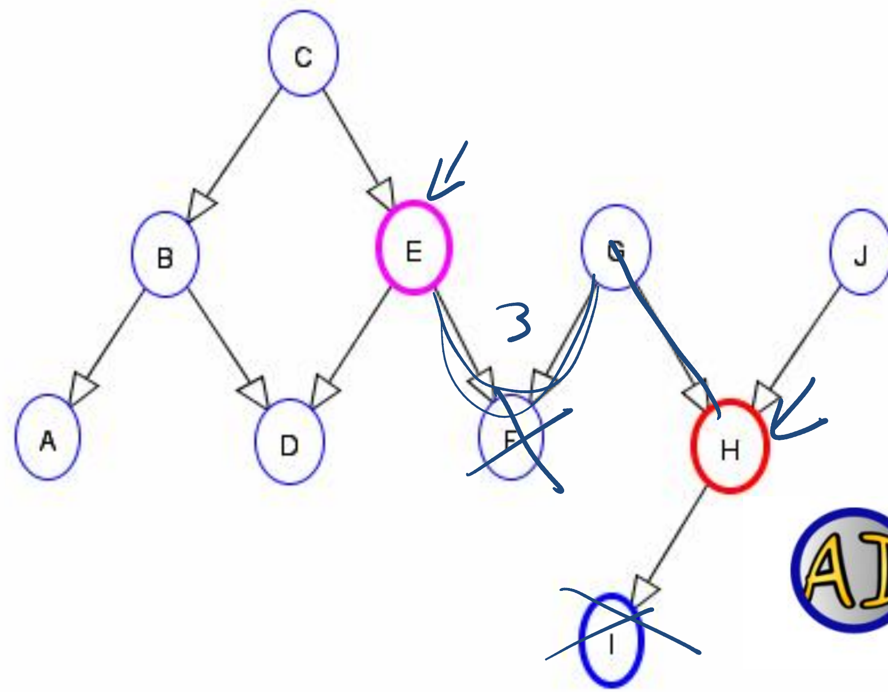
false



i>clicker

In/Dependencies in a Bnet : Example 3





Is H conditionally independent of E given ~~I~~?

true

Lecture Overview

- Implied Conditional Independence relations in a Bnet
- Compactness: Making stronger Independence assumptions
  - Representation of Compact Conditional Distributions
  - Network structure(Naïve Bayesian Classifier)

More on Construction and Compactness: Compact Conditional Distributions

Once we have established the topology of a Bnet, we still need to specify the conditional probabilities

How?

- **From Data**
- **From Experts**

To facilitate acquisition, we aim for compact representations for which data/experts can provide accurate assessments

More on Construction and Compactness: Compact Conditional Distributions

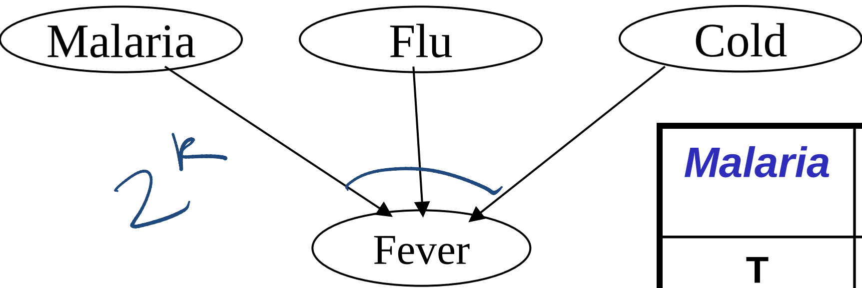
From JointPD 2^n to $n2^k$

But still, CPT grows exponentially with number of parents

In **realistic model of internal medicine** with 448 nodes and 906 links 133,931,430 values are required!

And often there are no data/experts for accurate assessment

Effect with multiple non-interacting causes



2^k

What do we need to specify?

Malaria	*Flu*	*Cold*	*P(Fever=T | ..)*	*P(Fever=F|..)*
T	T	T		
T	T	F		
T	F	T		
T	F	F		
F	T	T		
F	T	F		
F	F	T		
F	F	F		

What do you think data/experts could easily tell you?

when only one cause is present

More difficult to get info to assess more complex conditioning….

Solution: Noisy-OR Distributions

- Models multiple non interacting causes
- Logic OR with a probabilistic twist.
  - Logic OR Conditional Prob. Table.

Malaria	*Flu*	*Cold*	*P(Fever=T | ..)*	*P(Fever=F|..)*
T	T	T	1	0
T	T	F	1	0
T	F	T	1	0
T	F	F	1	0
F	T	T	1	0
F	T	F	1	0
F	F	T	1	0
F	F	F	0	1

Solution: Noisy-OR Distributions

The **Noisy-OR model** allows for **uncertainty in the ability of each cause to generate the effect** (e.g.. one may have a cold without a fever)

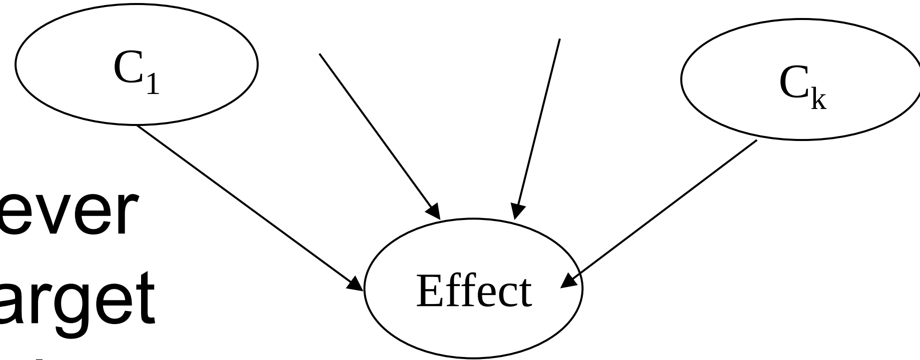
u on zeros

Malaria	*Flu*	*Cold*	*P(Fever=T | ..)*	*P(Fever=F|..)*
T	T	T		
T	T	F		
T	F	T		
T	F	F		
F	T	T		
F	T	F		
F	F	T		
F	F	F	0	1

Two assumptions

1. All possible causes a listed
2. For each of the causes, whatever inhibits it to generate the target effect is independent from the inhibitors of the other causes

Noisy-OR: Derivations



For each of the causes, whatever inhibits it to generate the target effect is independent from the inhibitors of the other causes

Independent Probability of failure q_i for each cause alone:

- P(Effect=F | C_i = T, *and no other causes*) = q_i
- P(Effect=F | C_1 = T,.. C_j = T, C_{j+1} = F,., C_k = F)=

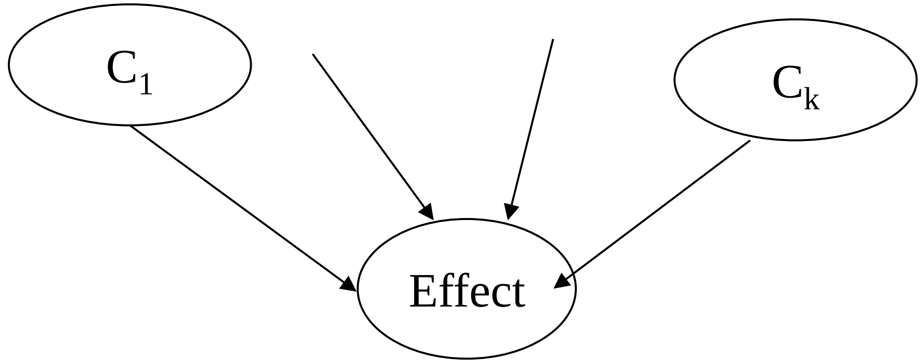
A. $\sum_{i=1}^{J} q_i$

B. $\prod_{i=1}^{J} q_i$

C. $1 - q_i$

D. None of those

Noisy-OR: Derivations



For each of the causes, whatever inhibits it to generate the target effect is independent from the inhibitors of the other causes

Independent Probability of failure q_i for each cause alone:

- P(Effect=F | C_i = T, *and no other causes*) = q_i
- P(Effect=F | C_1 = T,.. C_j = T, C_{j+1} = F,., C_k = F) = $\prod_{i=1}^{j} q_i$
- P(Effect=T | C_1 = T,.. C_j = T, C_{j+1} = F,., C_k = F) = $1 -$

qi

Noisy-OR: Example

P(Fever=F| Cold=T, Flu=F, Malaria=F) = 0.6

P(Fever=F| Cold=F, Flu=T, Malaria=F) = 0.2

P(Fever=F| Cold=F, Flu=F, Malaria=T) = 0.1

Model of internal medicine

$\underbrace{133{,}931{,}430}$ → $\underbrace{8{,}254}$

using Noisy-ORs

- P(Effect=F | C_1 = T,.. C_j = T, C_{j+1} = F,., C_k = F)= $\prod_{i=1}^{j} q_i$

Malaria	*Flu*	*Cold*	*P(Fever=T | ..)*	*P(Fever=F|..)*
→ **T**	**T**	**T**	.988	0.1 x 0.2 x 0.6 = **0.012**
→ **T**	**T**	**F**	→ .98	0.2 x 0.1 = **0.02**
T	**F**	**T**	.94	0.6 x 0.1=**0.06**
→ **T**	**F**	**F**	**0.9**	**0.1** ←
F	**T**	**T**	.88	0.2 x 0.6 = **0.12**
→ **F**	**T**	**F**	**0.8**	**0.2** ←
→ **F**	**F**	**T**	**0.4**	**0.6** ←
F	**F**	**F**	0	**1.0**

required

3 in this example

- Number of probabilities linear in ….. K

Lecture Overview

- Implied Conditional Independence relations in a Bnet
- Compactness: Making stronger Independence assumptions
  - Representation of Compact Conditional Distributions
  - **Network structure (Naïve Bayesian Classifier)**

Naïve Bayesian Classifier

A very simple and successful Bnets that allow to classify entities in a set of classes C, given a set of attributes

Example:

- Determine whether an email is spam (only two classes spam=T and spam=F)
- Useful attributes of an email ? words contained in the email

Assumptions

- The value of each attribute depends on the classification
- **(Naïve)** The attributes are independent of each other given the classification

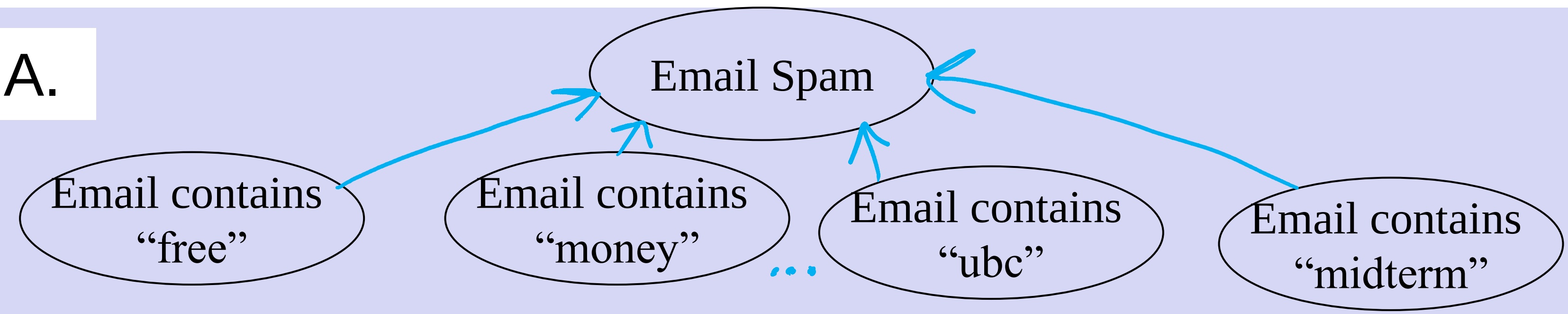
P("bank" | "account" , spam=T) ≠ P("bank" | spam=T)

What is the structure?

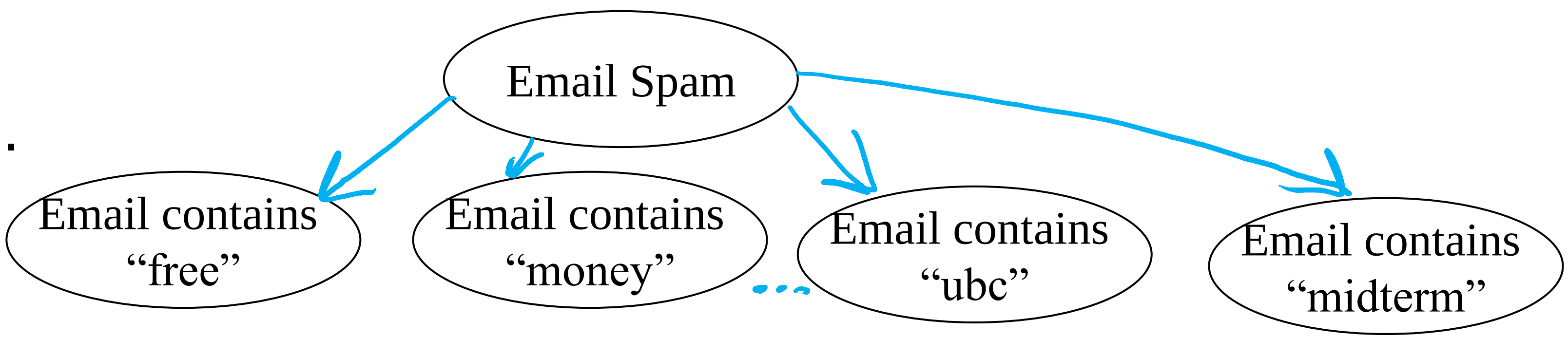
Assumptions

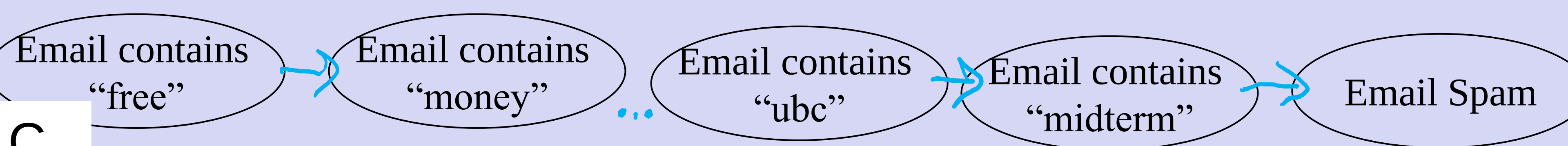
- The value of each attribute depends on the classification
- **(Naïve)** The attributes are independent of each other given the classification

A.



B.



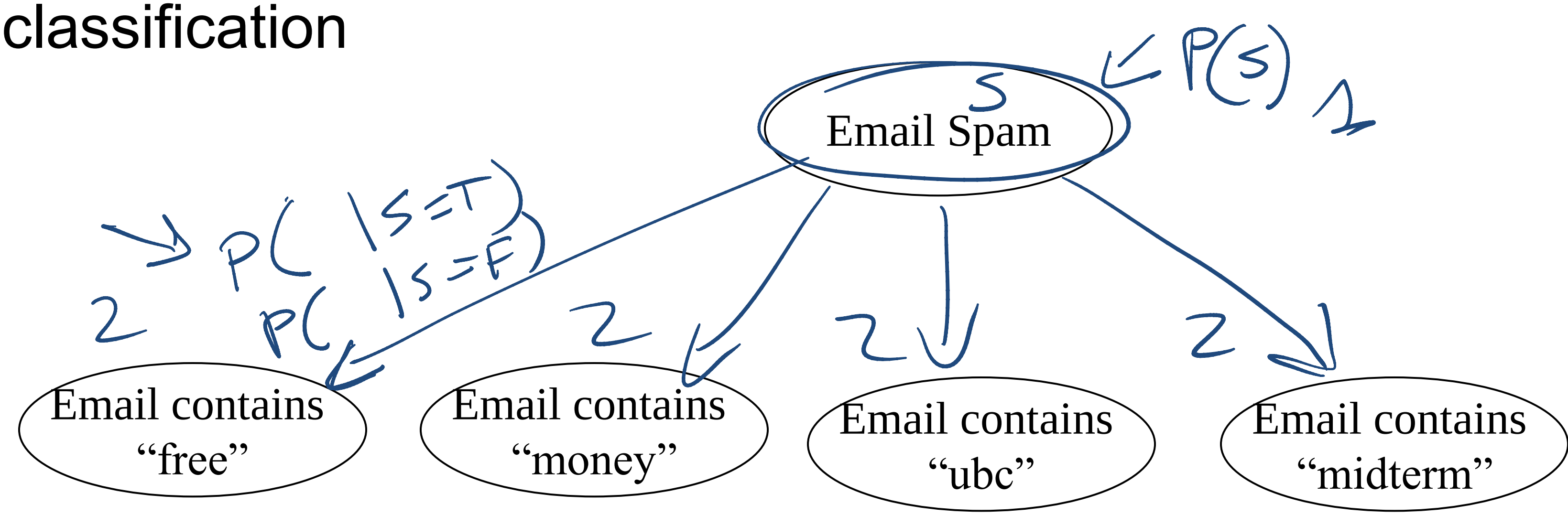


C.

Naïve Bayesian Classifier for Email Spam

Assumptions

- The value of each attribute depends on the classification
- **(Naïve)** The attributes are independent of each other given the classification



words

10^5 in English

Number of parameters?

$2 \cdot 10^5$ vs 2^{10^5} Joint

Easy to acquire?

If you have a large collection of emails for which you know if they are spam or not……

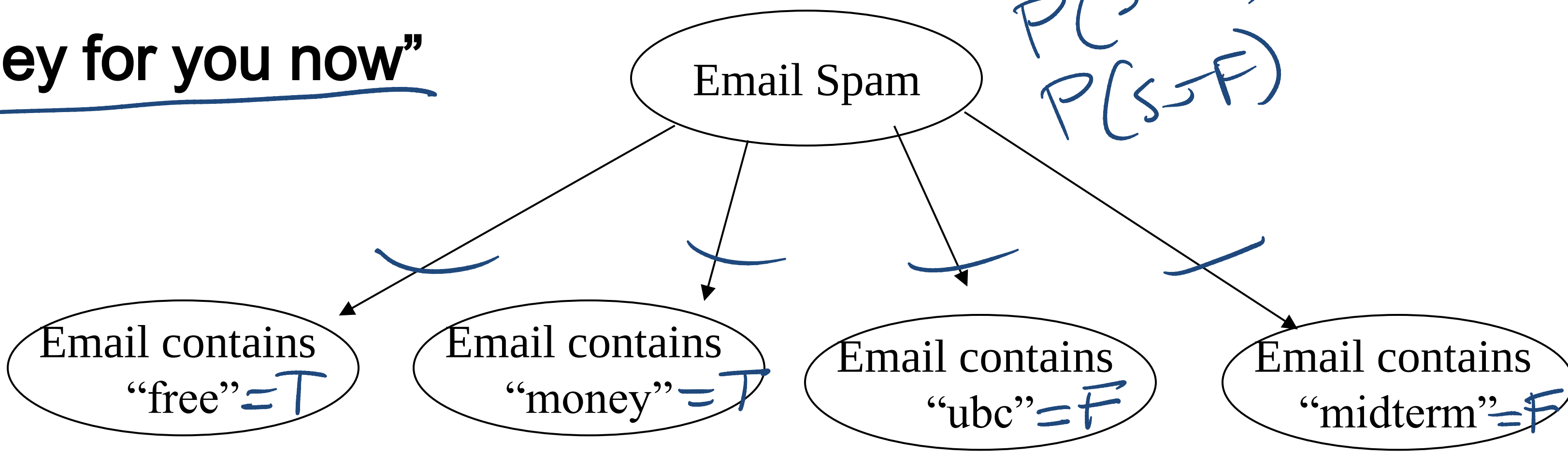
NB Classifier for Email Spam: Usage

Most likely class given set of observations

Is a given Email *E* spam?

“free money for you now”

$P(S=T)$
$P(S=F)$



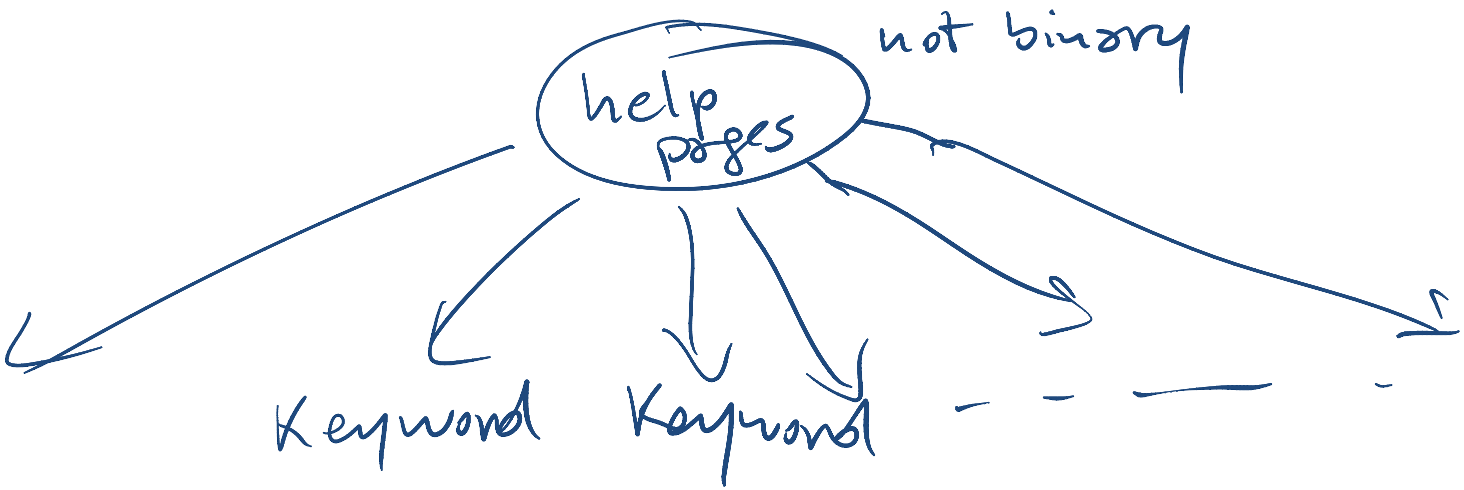
Email is a spam if....... $P(S=T) > P(S=F)$

after the two probs are updated in light of the evidence (words in email are set to T)

For another example of naïve Bayesian Classifier

See textbook ex. 6.16

help system to determine what **help page a user is interested in** based on **the keywords they give in a query** to a help system.



Learning Goals for today's class

You can:

- Given a Belief Net, determine whether one variable is conditionally independent of another variable, given a set of observations.

- Define and use **Noisy-OR** distributions. Explain assumptions and benefit.

- Implement and use a **naïve Bayesian classifier**. Explain assumptions and benefit.

Next Class

Bayesian Networks Inference: **Variable Elimination**

Course Elements

- Work on Practice Exercises 6A and 6B
- Assignment 3 is due on Wed the 20th !
- Assignment 4 will be available on Thur and due on Nov the 29th (last class).