Bayesian Networks – Representation

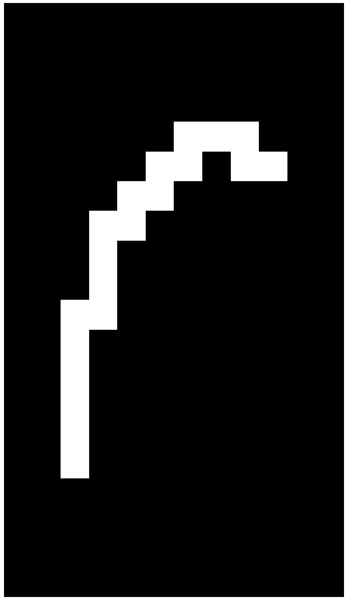
Machine Learning – 10701/15781

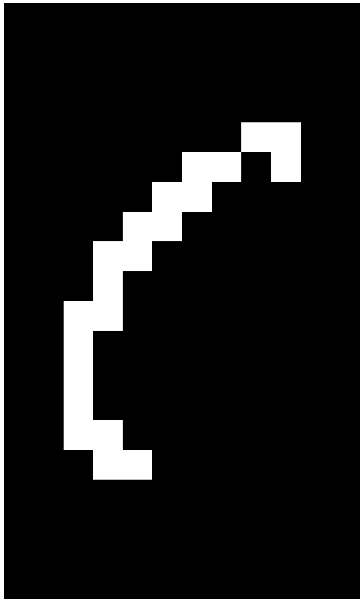
Carlos Guestrin

Carnegie Mellon University

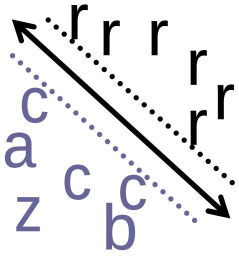
March 16th, 2005

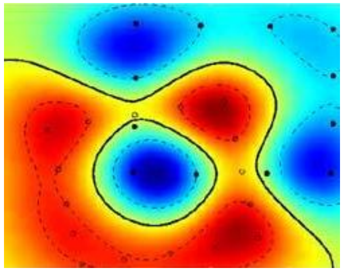
Handwriting recognition





Character recognition, e.g., kernel SVMs





Webpage classification

→

Company home page

vs

Personal home page

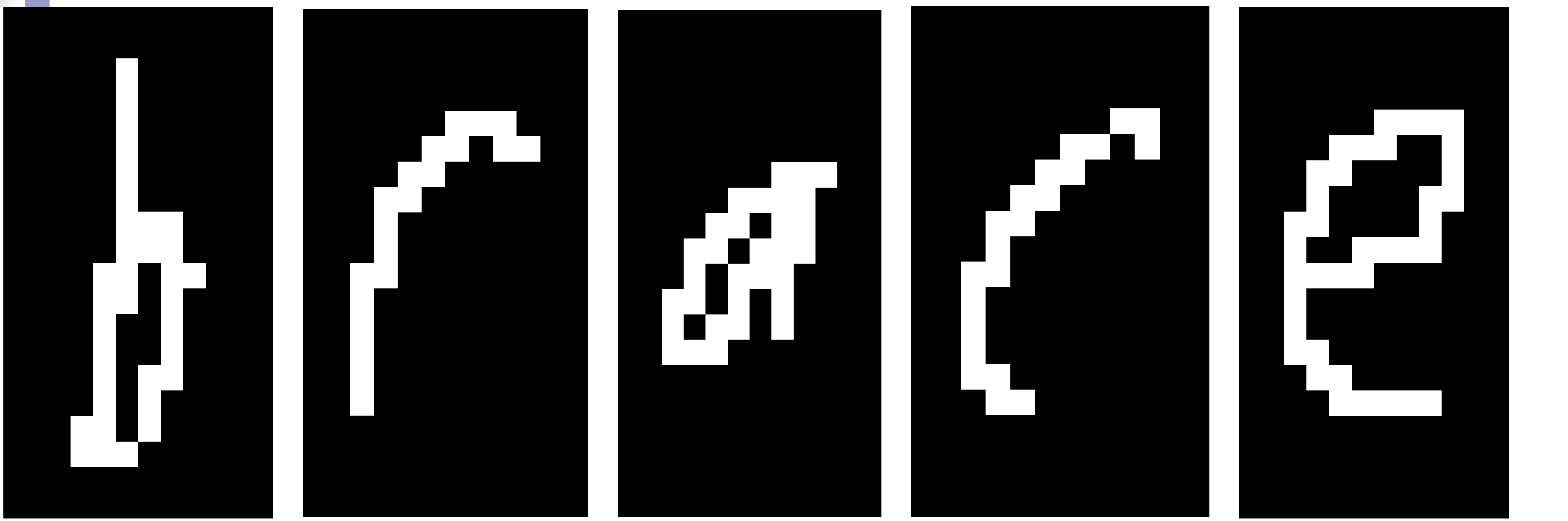
vs

Univeristy home page

vs

…

Handwriting recognition 2



Webpage classification 2

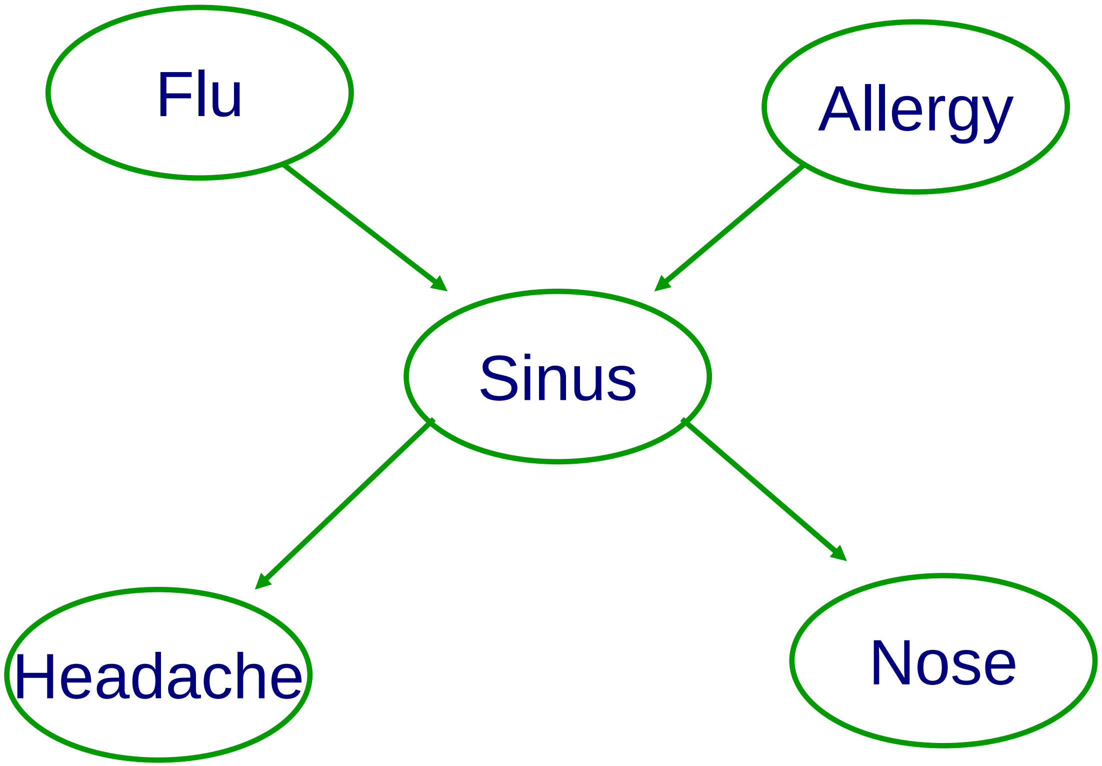
Today – Bayesian networks

- One of the most exciting advancements in statistical AI in the last 10-15 years
- Generalizes naïve Bayes and logistic regression classifiers
- Compact representation for exponentially-large probability distributions
- Exploit conditional independencies

Causal structure

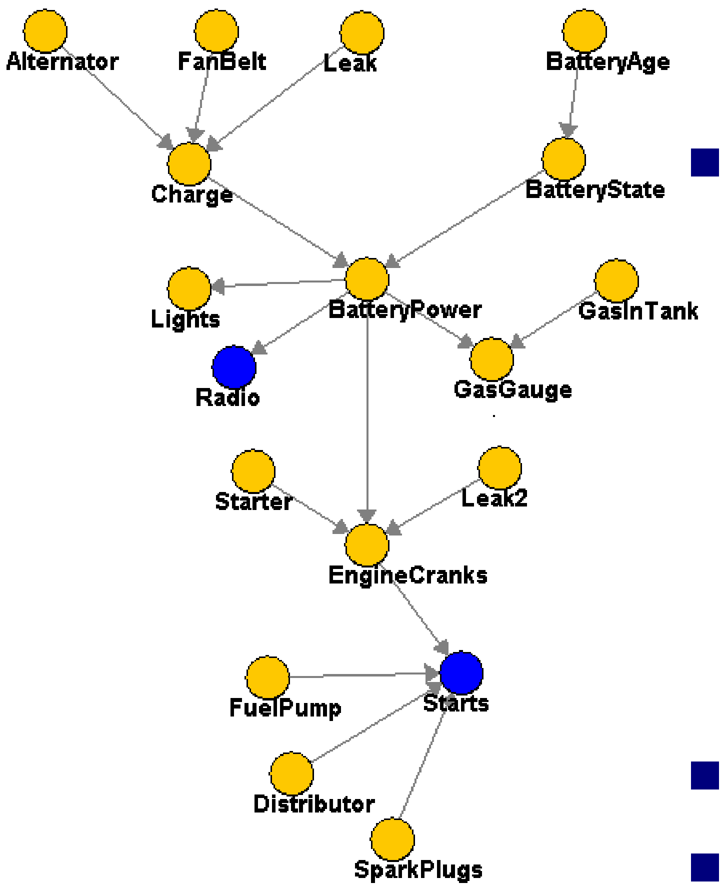
- Suppose we know the following:
  - The flu causes sinus inflammation
  - Allergies cause sinus inflammation
  - Sinus inflammation causes a runny nose
  - Sinus inflammation causes headaches
- How are these connected?

Possible queries



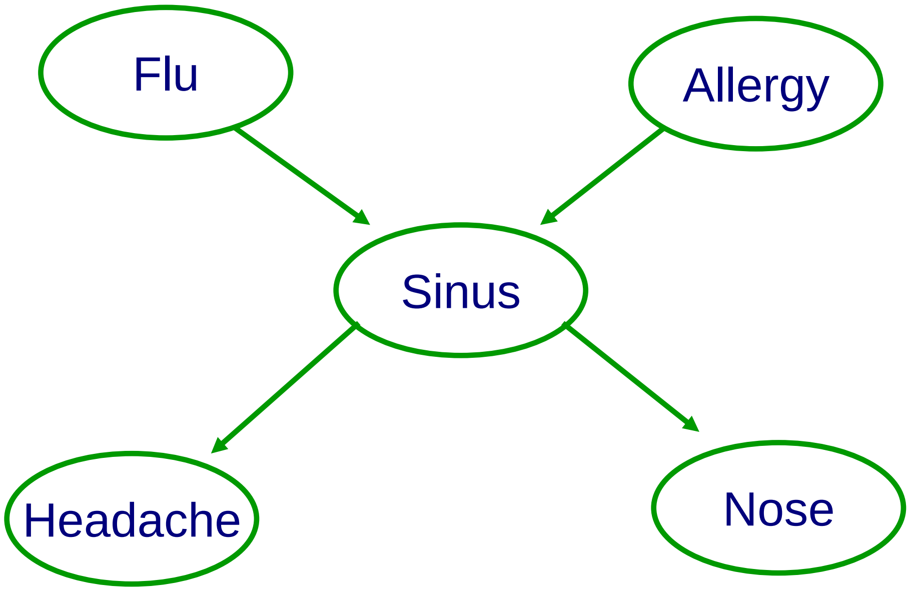
- Inference
- Most probable explanation
- Active data collection

Car starts BN

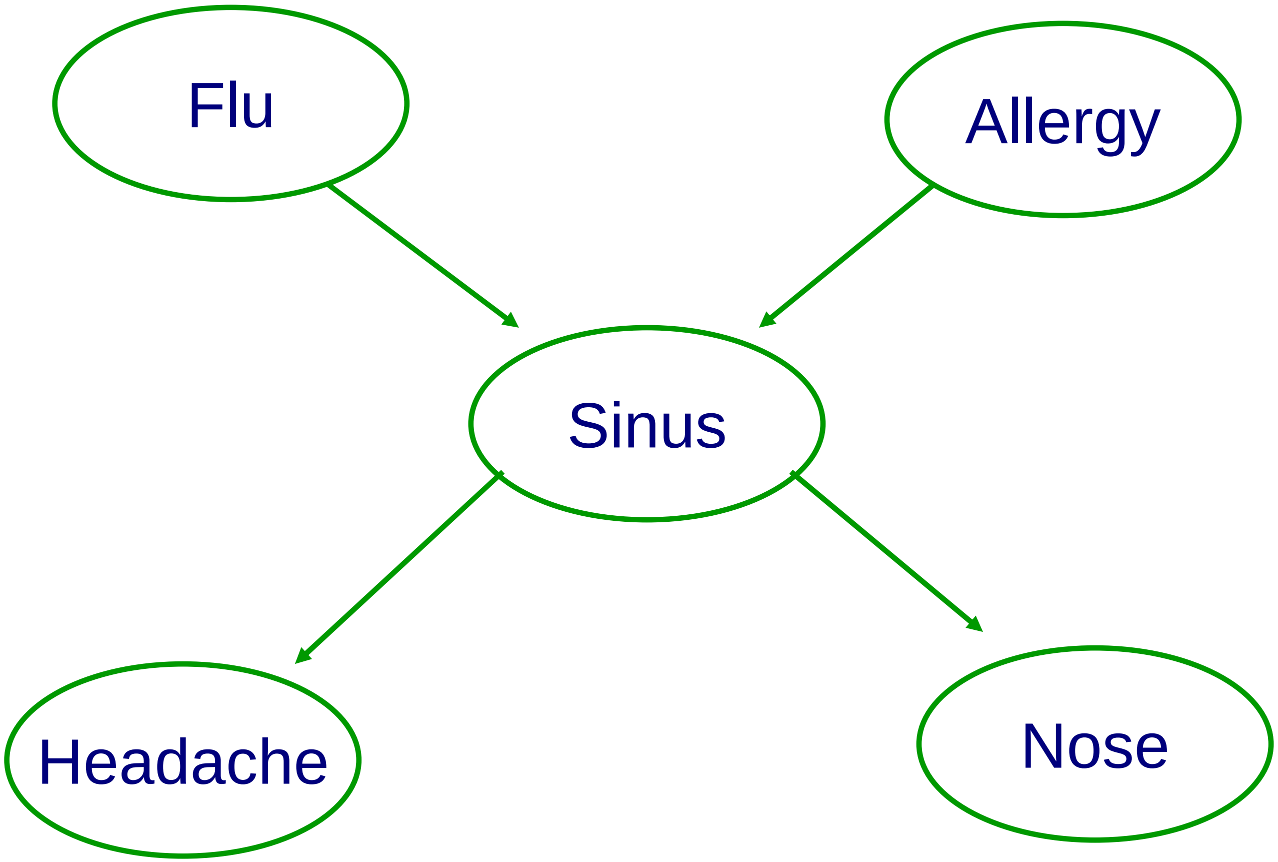


- 18 binary attributes
- Inference
  - P(BatteryAge|Starts=f)
- 2^{18} terms, why so fast?
- Not impressed?
  - HailFinder BN – more than 3^{54} = 58149737003040059690390169 terms

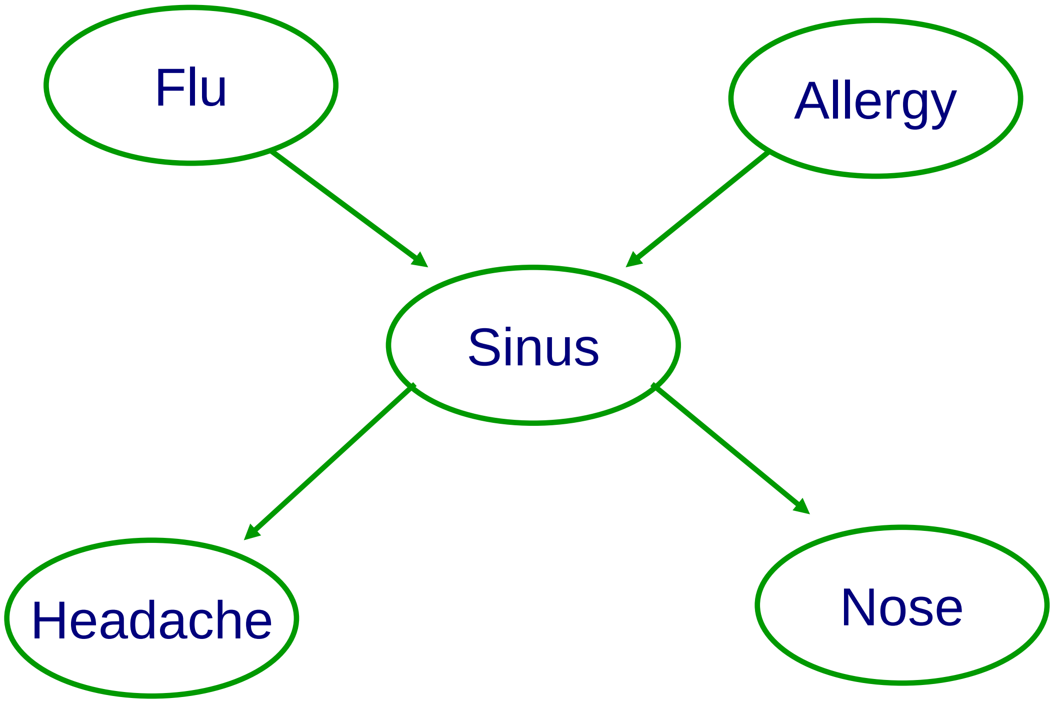
Factored joint distribution - Preview



Number of parameters



Key: Independence assumptions



Knowing sinus separates the variables from each other

(Marginal) Independence

- Flu and Allergy are (marginally) independent

Flu = t	
Flu = f	

Allergy = t	
Allergy = f	

- More Generally:

	Flu = t	Flu = f
Allergy = t		
Allergy = f		

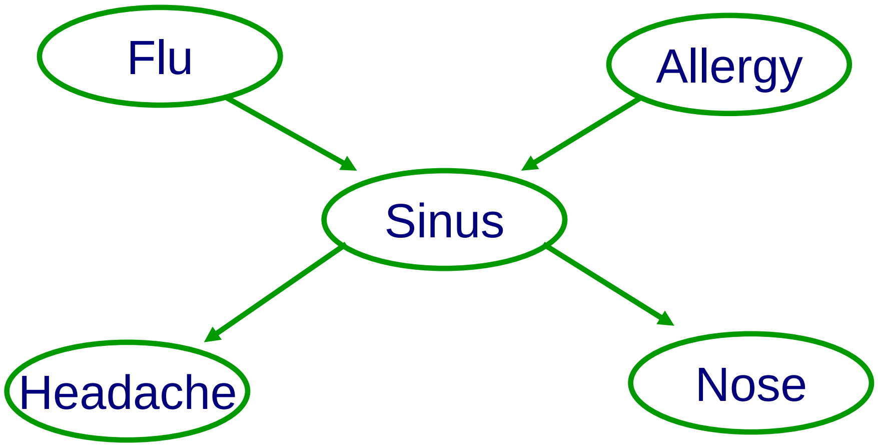
Conditional independence

- Flu and Headache are not (marginally) independent

- Flu and Headache are independent given Sinus infection

- More Generally:

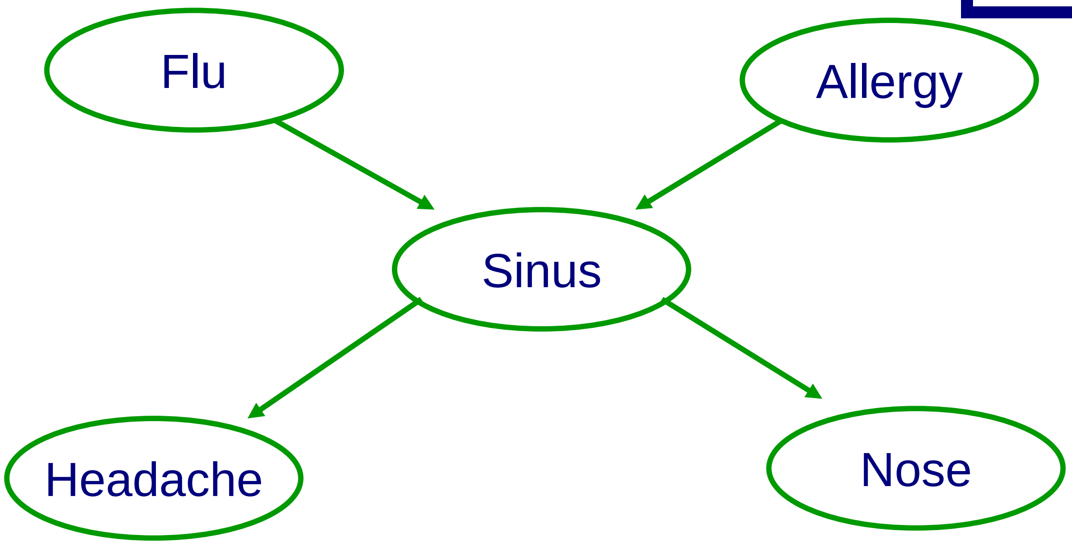
The independence assumption



Local Markov Assumption:
A variable X is independent of its non-descendants given its parents

Explaining away

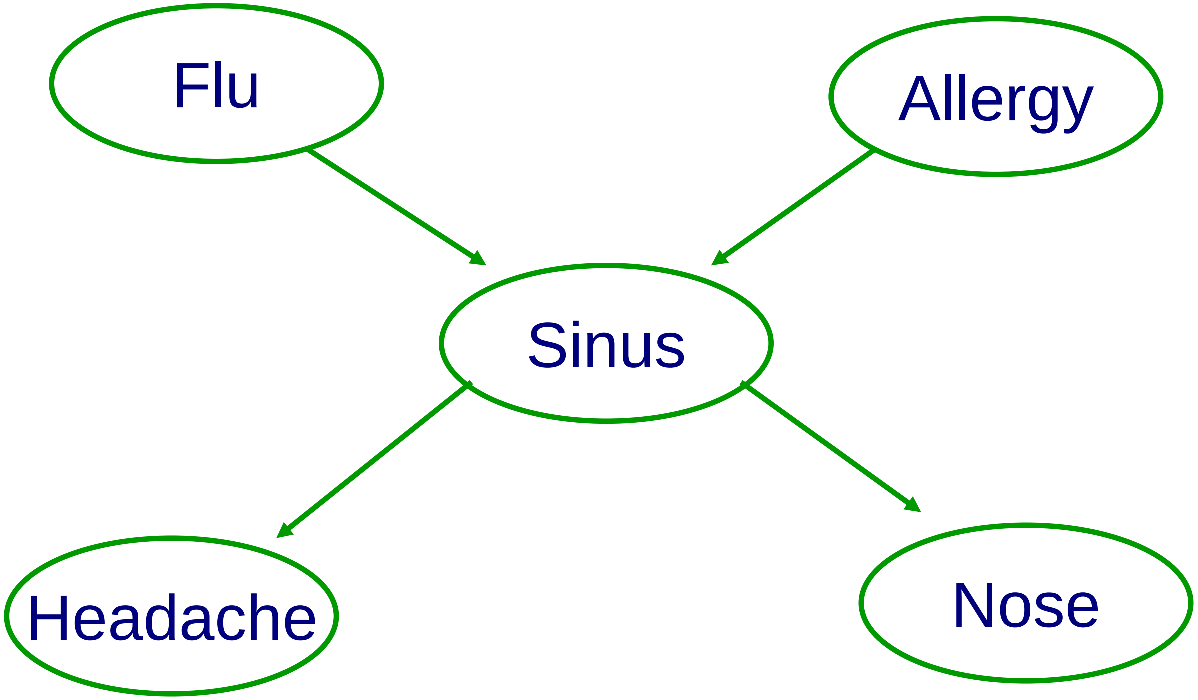
Local Markov Assumption:
A variable X is independent of its non-descendants given its parents



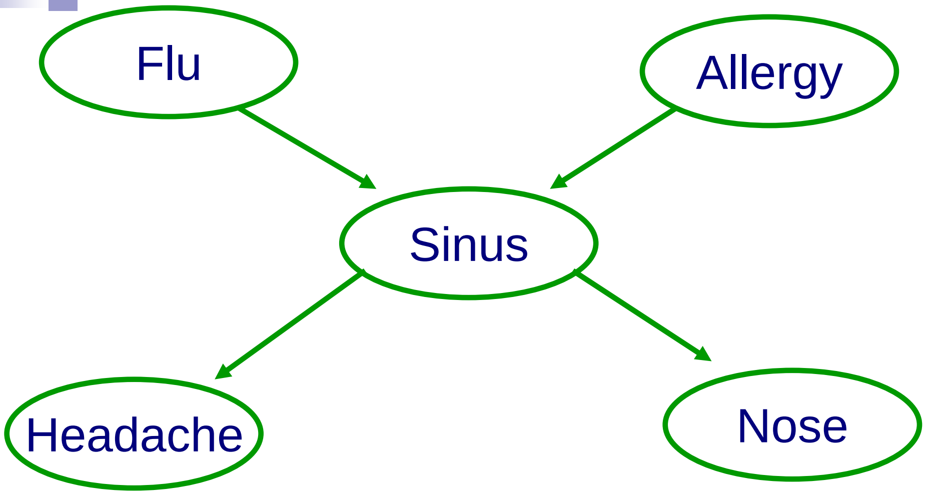
Naïve Bayes revisited

Local Markov Assumption:
A variable X is independent of its non-descendants given its parents

What about probabilities?
Conditional probability tables (CPTs)



Joint distribution



Why can we decompose? Markov Assumption!

Real Bayesian networks applications

- Diagnosis of lymph node disease
- Speech recognition
- Microsoft office and Windows
  - http://www.research.microsoft.com/research/dtg/
- Study Human genome
- Robot mapping
- Robots to identify meteorites to study
- Modeling fMRI data
- Anomaly detection
- Fault dianosis
- Modeling sensor network data

A general Bayes net

- Set of random variables
- Directed acyclic graph
  - Encodes independence assumptions
- CPTs
- Joint distribution:

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P\left(X_i \mid \mathbf{Pa}_{X_i}\right)$$

Another example

- Variables:
  - B – Burglar
  - E – Earthquake
  - A – Burglar alarm
  - N – Neighbor calls
  - R – Radio report
- Both burglars and earthquakes can set off the alarm
- If the alarm sounds, a neighbor may call
- An earthquake may be announced on the radio

Another example – Building the BN

- B – Burglar
- E – Earthquake
- A – Burglar alarm
- N – Neighbor calls
- R – Radio report

Defining a BN

- Given a set of variables and conditional independence assumptions
- Choose an ordering on variables, e.g., $X_1, \ldots, X_n$
- For i = 1 to n
  - Add X_i to the network
  - Define parents of X_i, $\mathbf{Pa}_{X_i}$, in graph as the minimal subset of $\{X_1,\ldots,X_{i-1}\}$ such that local Markov assumption holds – X_i independent of rest of $\{X_1,\ldots,X_{i-1}\}$, given parents $\mathbf{Pa}_{X_i}$
  - Define/learn CPT – $P(X_i | \mathbf{Pa}_{X_i})$

How many parameters in a BN?

- Discrete variables $X_1, \ldots, X_n$
- Graph
  - Defines parents of X_i, $\mathbf{Pa}_{X_i}$
- CPTs – $P(X_i | \mathbf{Pa}_{Xi})$

Defining a BN 2

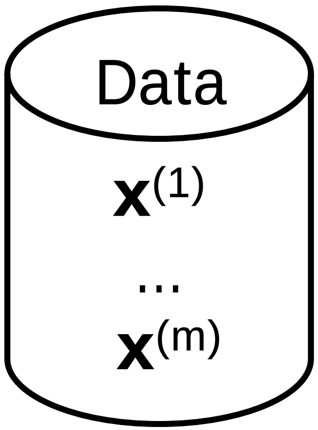
We may not know conditional independence assumptions and even variables

- Given a set of variables and conditional independence assumptions
- Choose an ordering on variables, e.g., $X_1, \ldots, X_n$
- For i = 1 to n
  - Add X_i to the network
  - Define parents of X_i, **P**
    subset of $\{X_1,\ldots,X_{i-1}\}$ s
    assumption holds – X_i independent of rest of $\{X_1,\ldots,X_{i-1}\}$, given parents $\mathbf{Pa}_{Xi}$
  - Define/learn CPT – $P(X_i | \mathbf{Pa}_{Xi})$

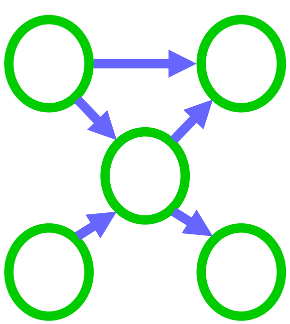
There are good orderings and bad ones – A bad ordering may need more parents per variable $\rightarrow$ must learn more parameters

How???

Learning the CPTs



For each discrete variable X_i



MLE: $$P(X_i = x_i \mid X_j = x_j) = \frac{\text{Count}(X_i = x_i, X_j = x_j)}{\text{Count}(X_j = x_j)}$$

Learning Bayes nets

	Known structure	Unknown structure
Fully observable data		
Missing data		

Queries in Bayes nets

- Given BN, find:
  - Probability of X given some evidence, $P(X|e)$
  - Most probable explanation, $\max_{x_1,\ldots,x_n} P(x_1,\ldots,x_n \mid e)$
  - Most informative query
- Learn more about these next class

What you need to know

- Bayesian networks
  - A compact **representation** for large probability distributions
  - Not an algorithm
- Semantics of a BN
  - Conditional independence assumptions
- Representation
  - Variables
  - Graph
  - CPTs
- Why BNs are useful
- Learning CPTs from fully observable data
- Play with applet!!! ☺

Acknowledgements

- JavaBayes applet
  - http://www.pmr.poli.usp.br/ltd/Software/javabayes/Home/index.html