

Probabilistic Graphical Models

Naïve Bayes, Ancestral
Sampling

Probabilistic Graphical Models

Directed graphical models

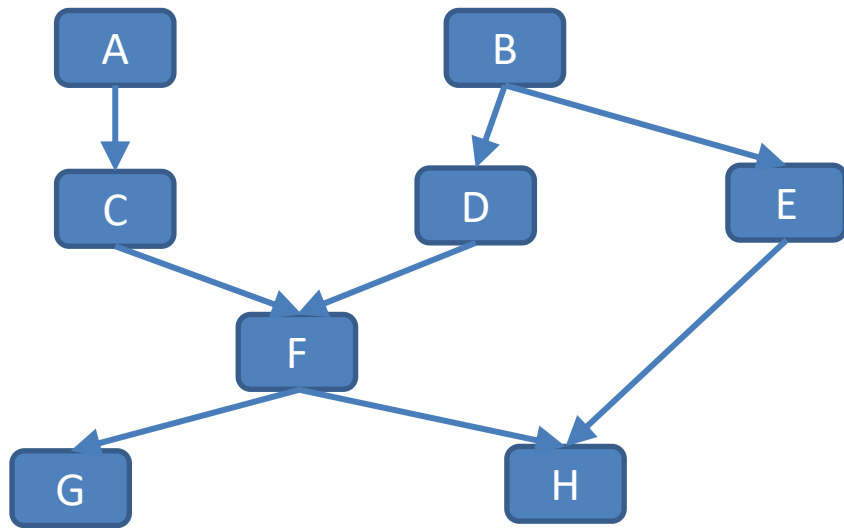
- Bayes Nets
- Conditional dependence

Undirected graphical models

- Markov random fields (MRFs)
- Factor graphs

From Markov Condition to Factorization

A Directed Acyclic Graph



A joint Probability Distribution

$$P(A, B, C, D, E, F, G, H)$$

$$P(A, \dots, H)$$

$$= \prod_{V \in \{A, \dots, H\}} P(V | Pa_G(V))$$

Markov Condition:

Every variable is independent of its non-descendants given its parents (in the graph)

Summary

BN: DAG + Distribution

The distribution factorizes according to the graph based on the Markov condition: Every variable is independent from its non-descendants (in the graph) based on its parents (in the graph)

D-separation allows us to read the independencies from the graph.
sound (dsep->ind) and
complete (dcon->dep in some distribution that factorizes according to G)

If $I(G) \subseteq I(P)$ then G is an I-Map for P

Summary

BN: DAG + Distribution

The distribution factorizes according to the graph based on the Markov condition: Every variable is independent from its non-descendants (in the graph) based on its parents (in the graph)

D-separation allows us to read the independencies from the graph.
sound (dsep->ind) and
complete (dcon->dep in some distribution that factorizes according to G)

Summary

BN: DAG + Distribution

If $I(G) \subseteq I(P)$ then G is an I-Map for P

If G is an I-Map for P and every G' that stems from removing an edge from G is not an I-Map for P , G is minimal I-Map for P

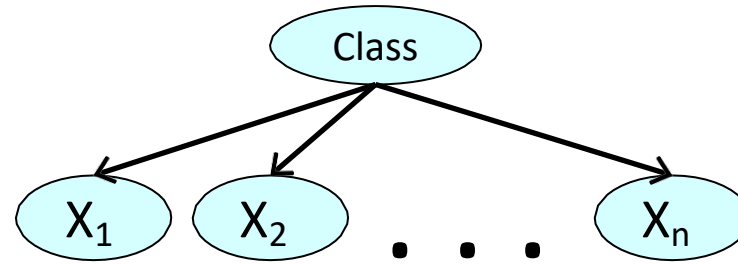
If $I(G) = I(P)$ then G is a perfect map for P

If $I(G) = I(G')$, G and G' are Markov Equivalent (I-Equivalent)

The Markov Boundary of Y is the set of Parents, Children and Spouses of G

Naïve Bayes Model

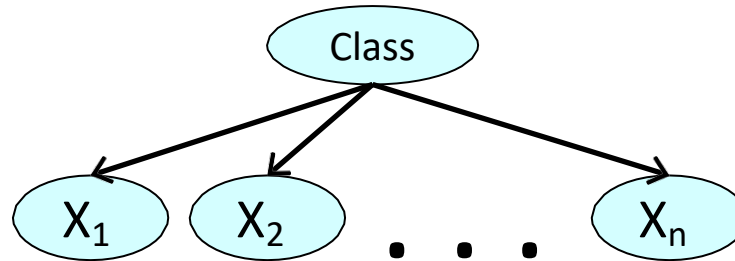
Features are independent given the class



$(X_i \perp X_j | C)$ for all X_i, X_j

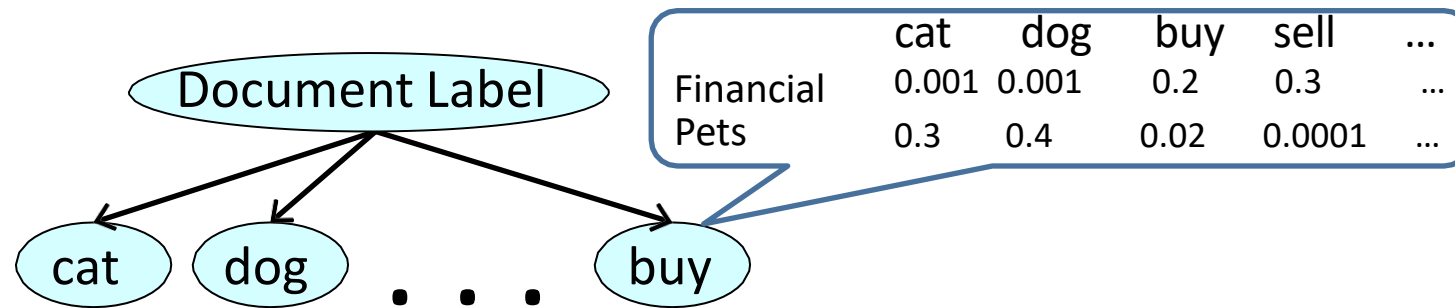
$$P(C, X_1, \dots, X_n) = P(C) \prod_{i=1}^n P(X_i | C)$$

Naïve Bayes Classifier



$$\frac{P(C = c^1 | x_1, \dots, x_n)}{P(C = c^2 | x_1, \dots, x_n)} = \frac{P(C = c^1)}{P(C = c^2)} \prod_{i=1}^n \frac{P(x_i | C = c^1)}{P(x_i | C = c^2)}$$

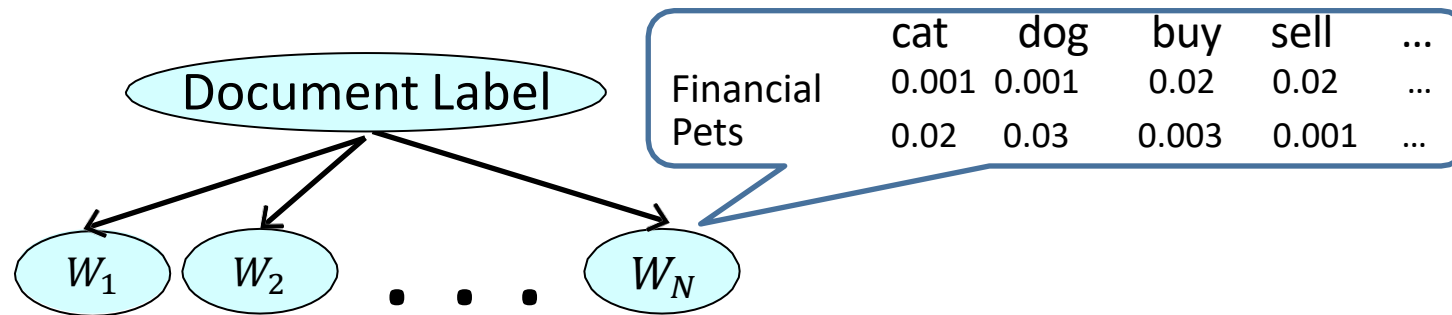
Bernoulli Naïve Bayes for Text



$P(\text{"cat" appears} \mid \text{Label})$

$$\frac{P(C = c^1 \mid x_1, \dots, x_n)}{P(C = c^2 \mid x_1, \dots, x_n)} = \frac{P(C = c^1)}{P(C = c^2)} \prod_{i=1}^n \frac{P(x_i \mid C = c^1)}{P(x_i \mid C = c^2)}$$

Multinomial Naïve Bayes for Text



P(word i is "cat" | Label)

$$\frac{P(C = c^1 | x_1, \dots, x_n)}{P(C = c^2 | x_1, \dots, x_n)} = \frac{P(C = c^1)}{P(C = c^2)} \prod_{i=1}^n \frac{P(x_i | C = c^1)}{P(x_i | C = c^2)}$$

Summary

- Simple approach for classification
 - Computationally efficient
 - Easy to construct
- Surprisingly effective in domains with many weakly relevant features
- Strong independence assumptions reduce performance when many features are strongly correlated

Bayesian Networks as Generative models

How do we generate samples from this distribution?

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x+1)^2}$$

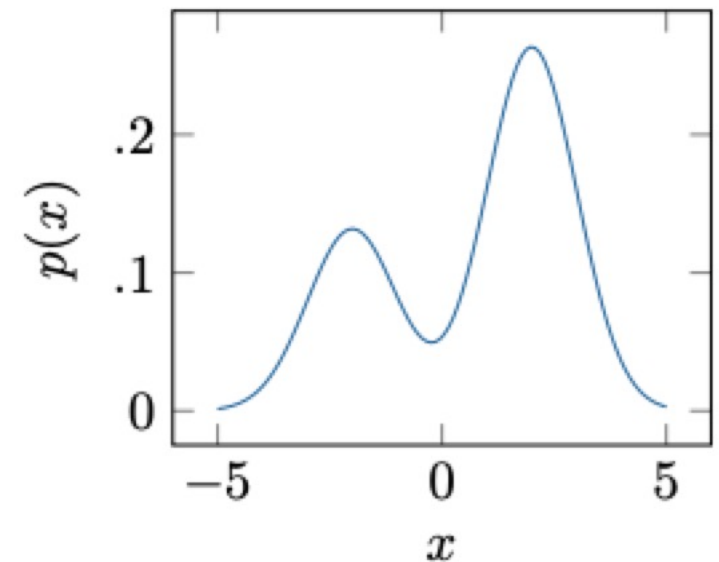
How do we generate samples from this distribution?

$$p(x) = \frac{1}{3\sqrt{2\pi}} e^{-\frac{1}{2}(x+1)^2} + \frac{2}{3\sqrt{2\pi}} e^{-\frac{1}{2}(x-1)^2}$$

See it as the sum of two normal distributions

1. Choose one component
2. Draw from that component

Ancestral sampling: sample in order in a BN



Bayesian Networks as Generative models

How do we generate samples from this distribution?

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x+1)^2}$$

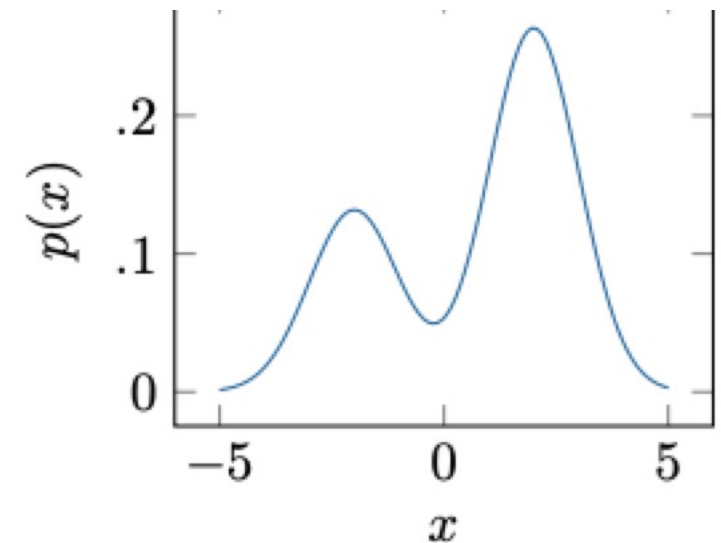
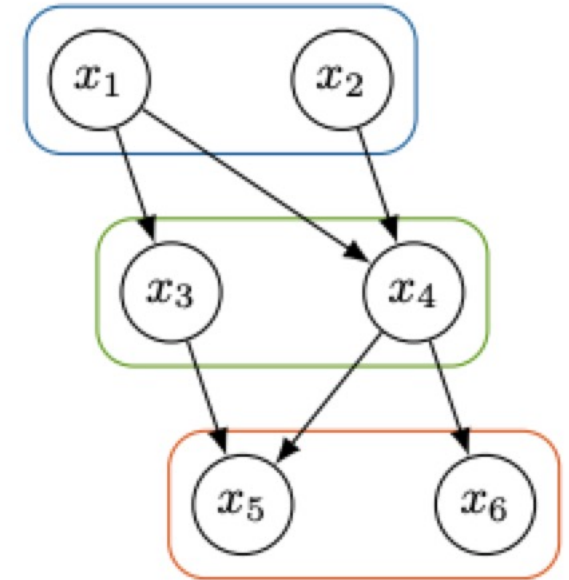
How do we generate samples from this distribution?

$$p(x) = \frac{1}{3\sqrt{2\pi}} e^{-\frac{1}{2}(x+1)^2} + \frac{2}{3\sqrt{2\pi}} e^{-\frac{1}{2}(x-1)^2}$$

See it as the sum of two normal distributions

1. Choose one component
2. Draw from that component

Ancestral sampling: sample in order in a BN

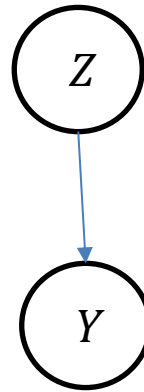


Example: Gaussian Mixture Model

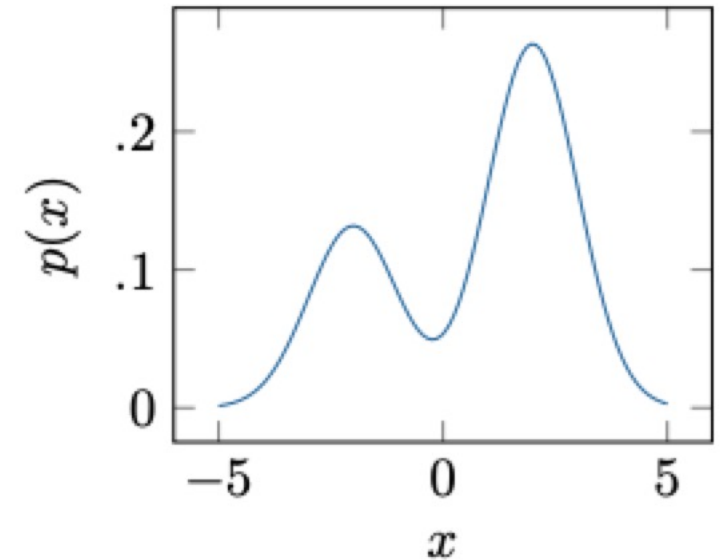
Probability Model

$$Z \sim \text{Bern}\left(\frac{1}{3}\right)$$
$$Y|Z = 0 \sim N(-1, 1)$$
$$Y|Z = 1 \sim N(1, 1)$$

Bayes Net



Joint Sample



Sample all nodes with no parents, then children, etc., to terminals. Can sample nodes at same level in parallel.

Medical Diagnosis: Pathfinder (1992)

- Help pathologist diagnose lymph node pathologies (60 different diseases)
- Pathfinder I: Rule-based system
- Pathfinder II used naïve Bayes and got superior performance
- Pathfinder III: Naïve Bayes with better knowledge engineering
- No incorrect zero probabilities
- Better calibration of conditional probabilities
 - $P(\text{finding}_1 | \text{disease}_1)$ to $P(\text{finding}_1 | \text{disease}_2)$
 - Not $P(\text{finding}_1 | \text{disease})$ to $P(\text{finding}_2 | \text{disease})$

Medical Diagnosis: Pathfinder (1992)

- Pathfinder IV: Full Bayesian network
 - Removed incorrect independencies
 - Additional parents led to more accurate estimation of probabilities
- BN model agreed with expert panel in 50/53 cases, vs 47/53 for naive Bayes model
- Accuracy as high as expert that designed the model

Heckerman et al.

Daphne Koller

Medical Diagnosis (Microsoft)

The screenshot shows a web browser window with a Microsoft logo and the text "Applet started". The browser address bar shows "ON STAGE", "ESSENTIALS", "COMMUNICATE", and "FIND". The main content area features the "OnParenting" logo with the dates "May 14 - May 20, 1997". To the right is a "Fidelity Investments" logo with the tagline "Our home on the web [is where] click here". Below these are navigation links: "cover contents news experts fun handbook talk find help feedback".

The main interface is divided into several sections:

- Search Instructions:** "There are two ways to search for specific information in OnParenting. In Find by Word, type the word(s) you want to find and get a list of titles relevant to that word. Find by Symptom will help you get information about children's symptoms. Help has tips to target your search."
- Search Options:** "Find by Word" and "Find by Symptom" (highlighted with a red arrow).
- Form Fields:**
 - Describe the child:** "in the drop-down boxes at the right. Relevant information will appear below."
 - Age:** "Toddler" (dropdown menu)
 - Sex:** "Female" (dropdown menu)
 - Complaint:** "Abdominal pain" (dropdown menu)
- Localized pain:** "Can the child localize, or point to, the site of the pain?"
 - No, unable to localize
 - Below the navel to the child's left
 - Above the child's navel
 - Either of the child's sides
 - Below the navel to the child's right
 - Above the navel to the child's right
 - Above the navel to the child's left
 - Don't Know
- Results so far:**

Disorder	Relevance
Viral gastroenteritis	<div style="width: 25%;"></div>
Psychosomatic pain	<div style="width: 25%;"></div>
Urinary tract infection	<div style="width: 10%;"></div>
Other	<div style="width: 5%;"></div>
- Navigation Buttons:** "Start Over", "Review", "Next>>", "Finish".

Thanks to: Eric Horvitz, Microsoft Research

Daphne Koller

Fault Diagnosis

- Many examples:
 - Microsoft troubleshooters
 - Car repair
- Benefits:
 - Flexible user interface
 - Easy to design and maintain