CS340 Machine learning
Bayesian networks

Conditional independence

- Recall the naïve Bayes assumption

$$X_j \perp X_k | Y$$

- This lets us factorize the class conditional density

$$p(\mathbf{x}|y) = \prod_{j=1}^{n_x} p(x_j|y)$$

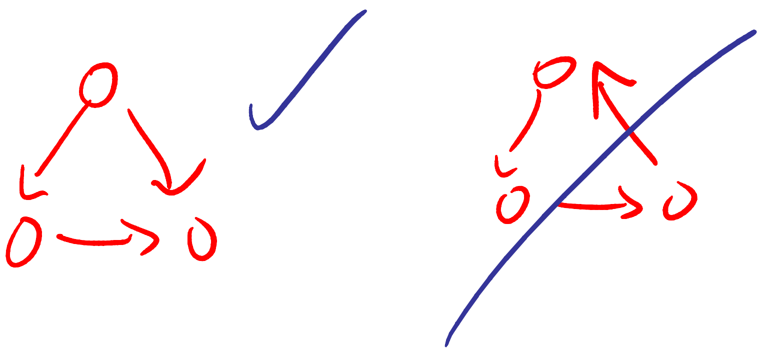
- Hence the joint distribution is

$$p(\mathbf{x}, y) = p(y) \prod_{j=1}^{n_x} p(x_j|y)$$

- Graphical models are ways to represent CI statements pictorially. This provides a compact way to define joint probability distributions.

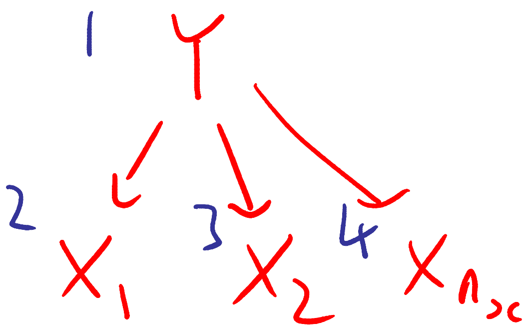
Kinds of graphical models

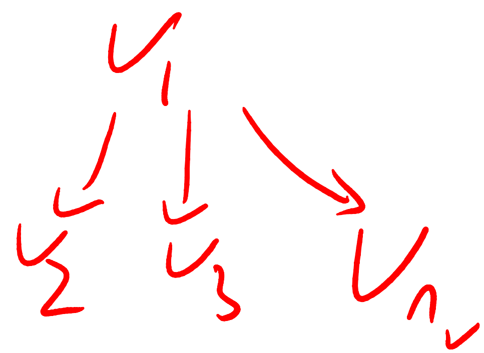
- Undirected graphical models – aka Markov Random fields – see later in class.
- Directed graphical models – aka Bayesian (belief) networks.
  - BNs require that the graph is a DAG (directed acyclic graphs).
  - No directed cycles allowed.



DAGs

- DAGs admit a total ordering (parents before children).
- Local Markov property: A node is independent of its predecssors given its parents.





$$X_j \perp X_{1:j} \mid Y$$

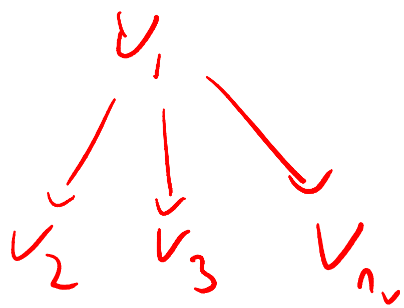
Chain rule

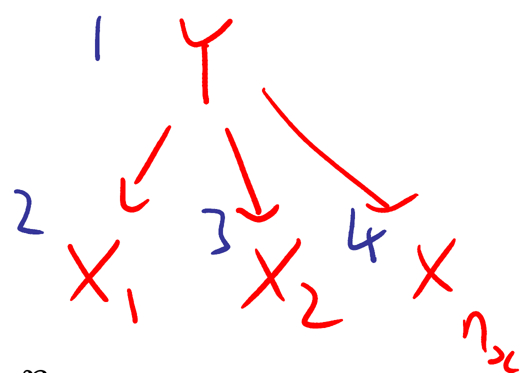
- By the chain rule

$$p(v_{1:n_v}) = p(v_1)p(v_2|v_1)p(v_3|v_1, v_2)\ldots p(v_{n_v}|v_{1:n_v-1}$$

- By the local Markov property

$$p(v_{1:n}) = p(v_1)p(v_2|v_{\pi_2})p(v_3|v_{\pi_3})\ldots p(v_n|\cancel{z}_{\pi_n})$$



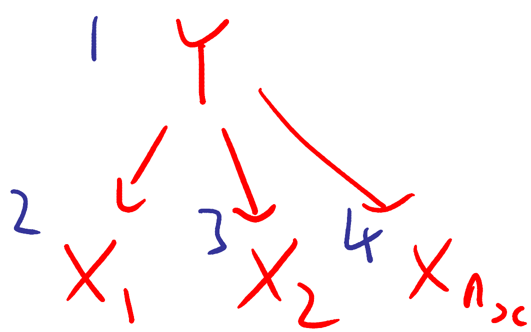


$$p(y, x_{1:n_x}) = p(y)\prod_{j=1}^{n_x} p(x_j|y)$$

Local Markov property is not enough

- NB property is $X_j \perp X_k \mid Y$ for all k, including $k > j$
- But local Markov property only tells us $X_j \perp X_k \mid Y$ for $k < j$
- Want to be able to answer the following for any sets of variables a,b,c: $Z_a \perp Z_b \mid Z_c$?

$V_a \perp V_b \mid V_c$



$X_j \perp X_{1:j} \mid Y$

Global Markov property

- By chaining together local independencies, one can infer global independencies.
- The general definition/ algorithm is complex, so we will break it into pieces.

Chains

- Consider the chain

X → Y → Z

$$p(x, y, z) = p(x)p(y|x)p(z|y)$$

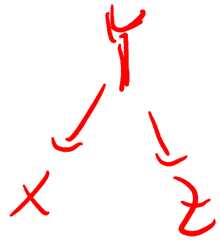
- If we condition and y, x and z are independent

$$\begin{aligned} p(x, z|y) &= \frac{p(x)p(y|x)p(z|y)}{p(y)} \\ &= \frac{p(x, y)p(z|y)}{p(y)} \\ &= p(x|y)p(z|y) \end{aligned}$$

X → Y → Z

Tents

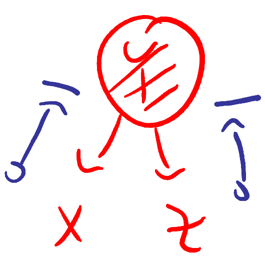
- Consider the "tent"



$$p(x, y, z) = p(y)p(x|y)p(z|y)$$

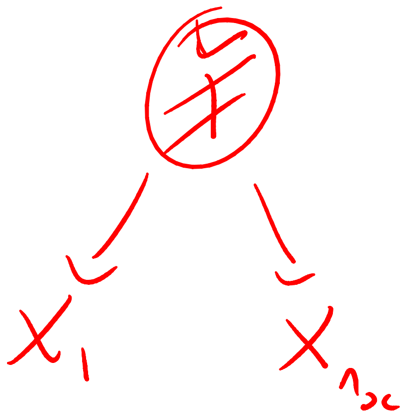
- Conditioning on Y makes X and Z independent

$$\begin{aligned} p(x, z|y) &= \frac{p(x, y, z)}{p(y)} \\ &= \frac{p(y)p(x|y)p(z|y)}{p(y)} = p(x|y)p(z|y) \end{aligned}$$



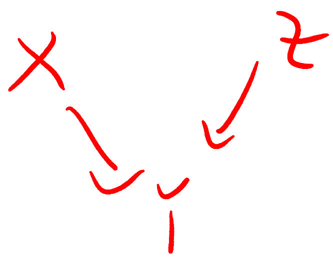
Naïve Bayes assumption

- Conditional on class, features are independent



V-structure

- Consider the v-structure



$$p(x, y, z) = p(x)p(z)p(y|x, z)$$

- X and Z are unconditionally independent

$$p(x, z) = \int p(x, y, z)dy = \int p(x)p(z)p(y|x, z)dy = p(x)p(z)$$

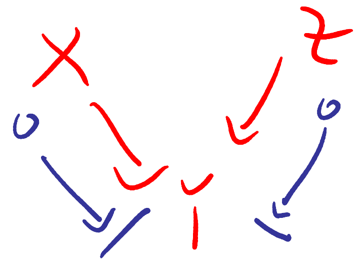
but are conditionally dependent

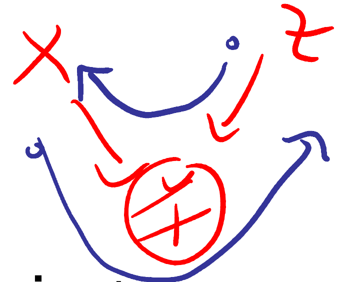
$$p(x, z|y) = \frac{p(x)p(z)p(y|x, z)}{p(y)} \neq f(x)g(z)$$

Explaining away

- Consider the v-structure

$X \perp Z$





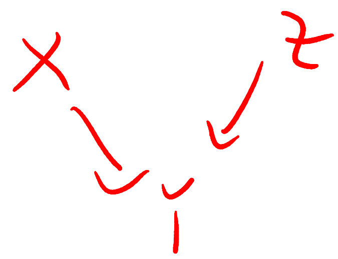
$X \not\perp Z \mid Y$

- Let X, Z $\in$ {0,1} be iid coin tosses.
- Let Y = X + Z.
- If we observe Y, X and Z are coupled.

X	Y	Z
0	0	0
0	1	1
1	0	1
1	1	2

Explaining away

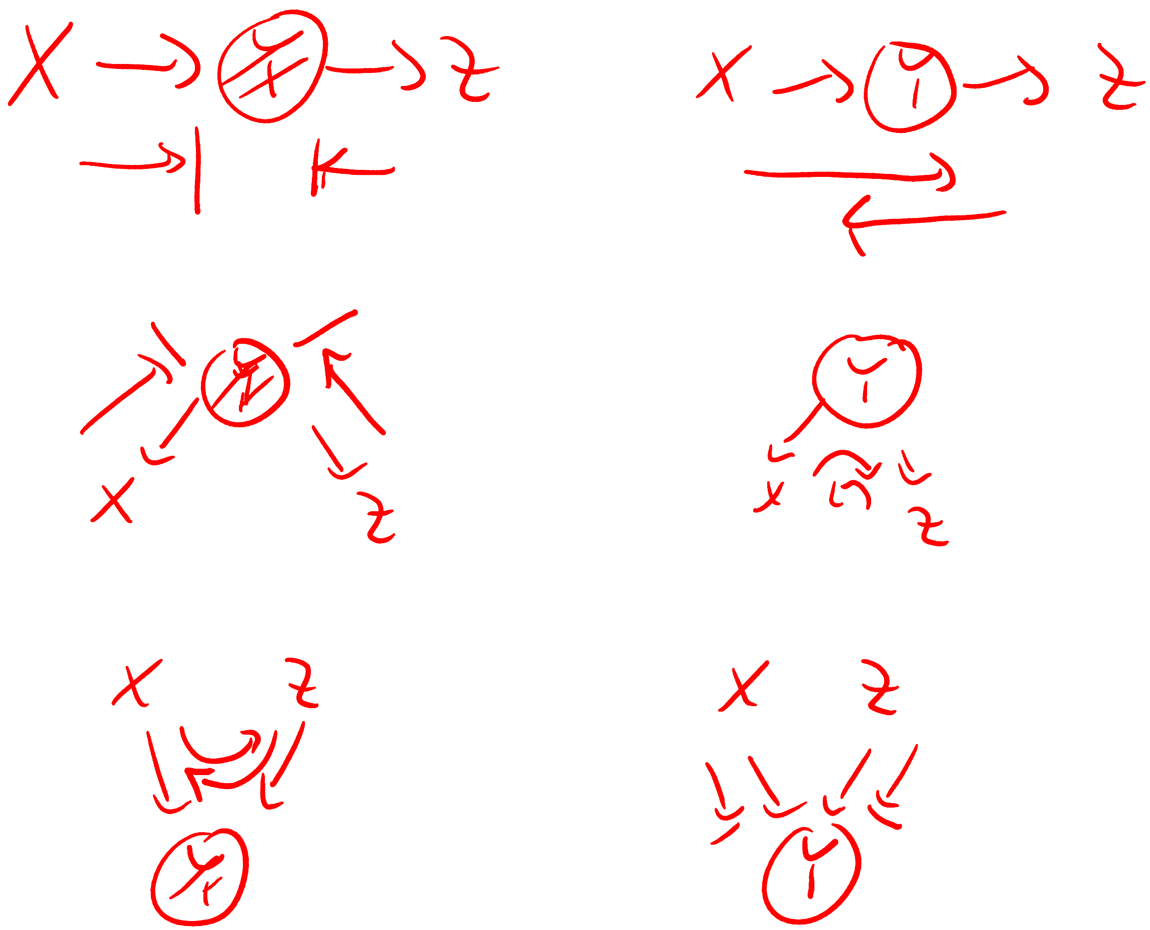
- Let Y = 1 iff burglar alarm goes off,
- X=1 iff burglar breaks in
- Z=1 iff earthquake occurred



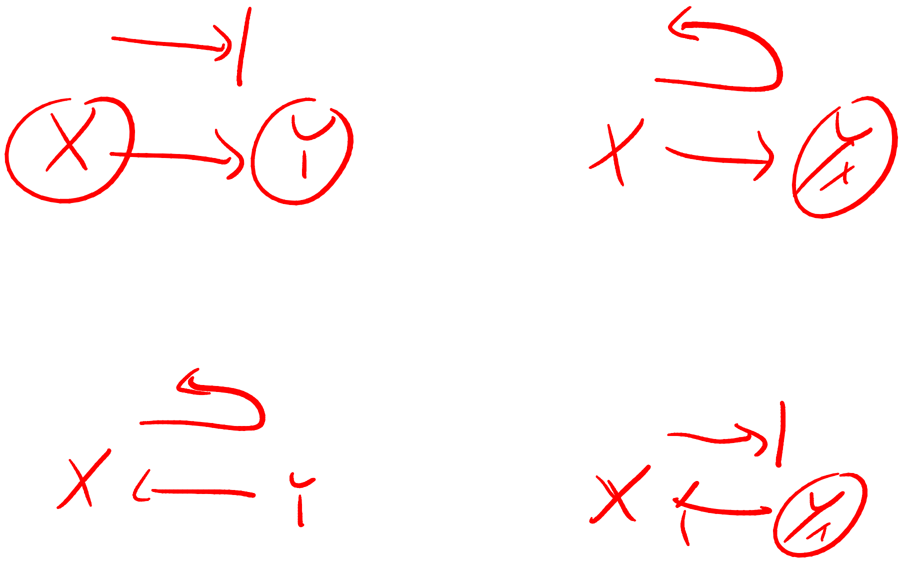
- X and Z compete to explain Y, and hence become dependent
- Intuitively, $p(X=1|Y=1) > p(X=1|Y=1,Z=1)$

Bayes Ball Algorithm

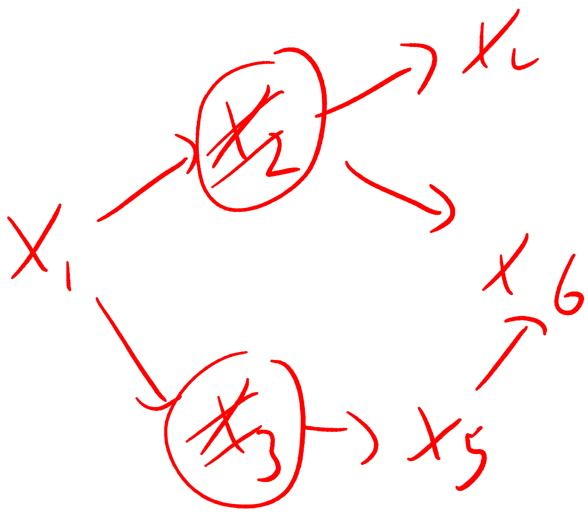
- $Z_A \perp Z_B \mid Z_C$ if every variable in A is d-separated from every variable in B when we shade the variables in C



Boundary conditions

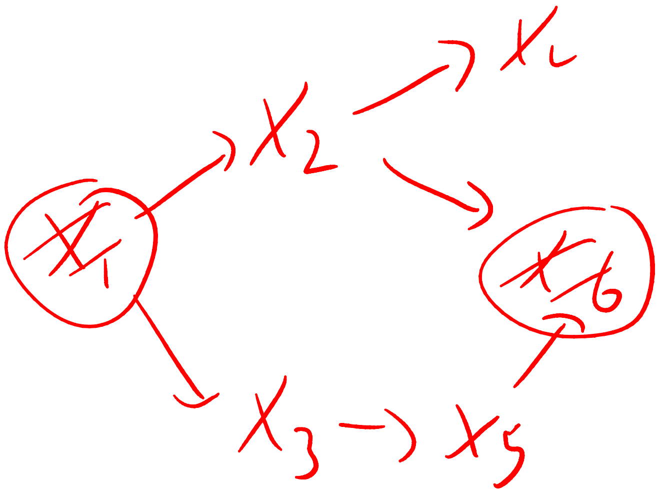


Example



$X_1 \perp X_6 \mid X_2, X_3$?

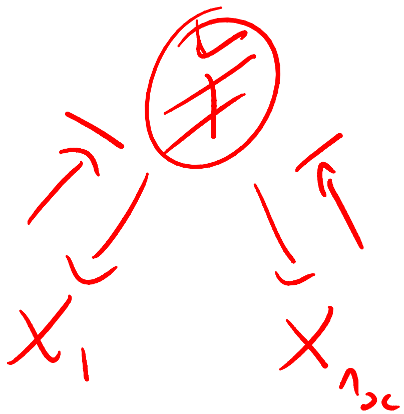
Example



$X_2 \perp X_3 \mid X_1, X_6$?

Naïve Bayes assumption

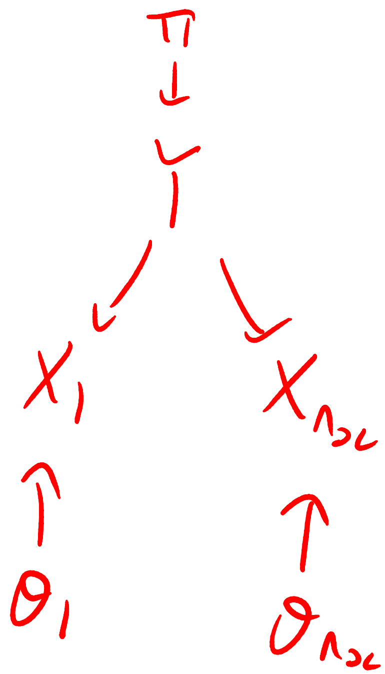
- Conditional on class, features are independent



$X_j \perp X_k \mid Y$

Parameters are rv's, too!

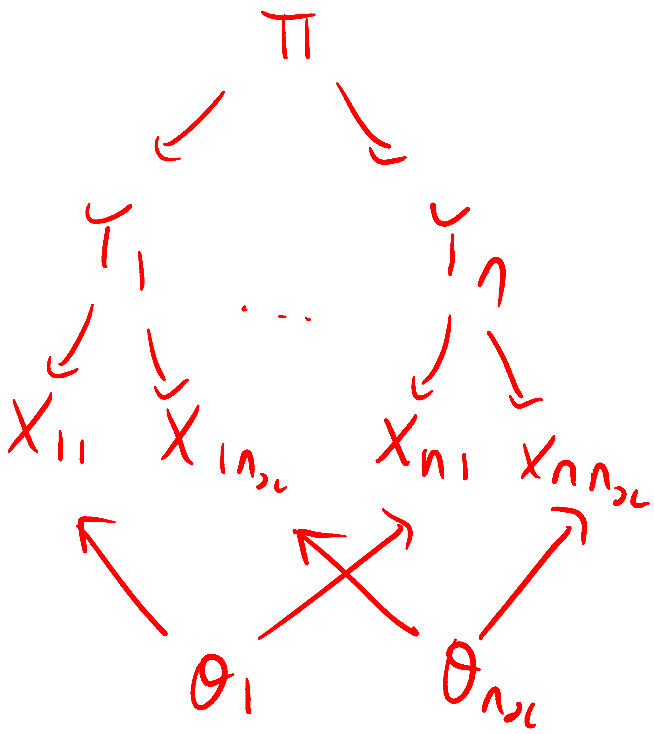
$$p(\mathbf{x}, y, \pi, \boldsymbol{\theta}) = p(\pi) p(y|\pi) \prod_{j=1}^{n_x} p(x_j|y, \theta_j) p(\theta_j)$$



This justifies our approach of estimating all the parameters independently

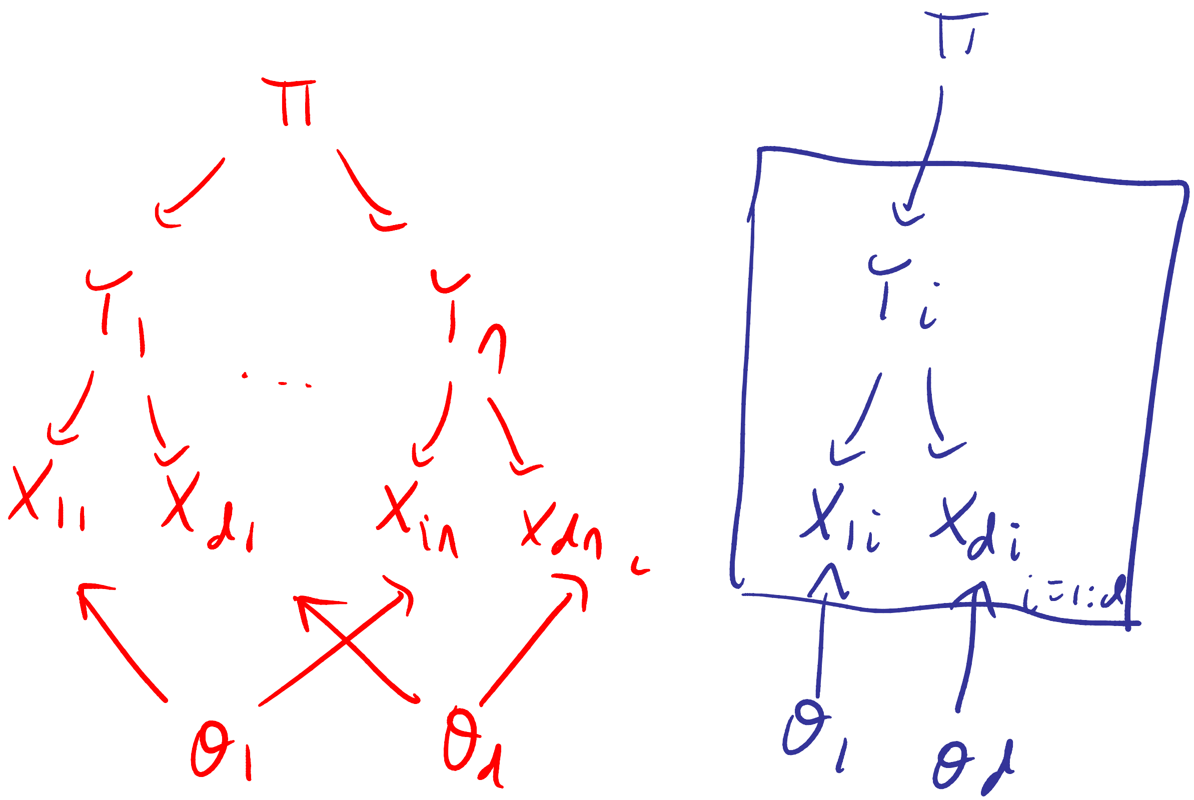
Repetitive structure

- When we have multiple samples, we replicate the variables, but the params are fixed



Plates

- We introduce a shorthand for repetitive structure



Plates

- We introduce a shorthand for repetitive structure