PROBABILISTIC GRAPHICAL MODELS

- **INFERENCE** $\quad p(x_1, \dots, x_n) \rightarrow$ compute $p(x_i)$ marginals, $p(x_i \mid x_j)$ conditionals

$$p(\underbrace{x_1, \dots, x_n}_{\text{OBSERVED}}, z_1, \dots, z_m) \leftarrow$$

$$p(z_1, \dots, z_m \mid x_1, \dots, x_m) = \frac{p(x_1, \dots, x_n, z_1, \dots, z_m)}{\underbrace{p(x_1, \dots, x_m)}_{\underbrace{\int p(x_1, \dots, x_n, z_1, \dots, z_m)\, dz_1 \cdots dz_m}_{\text{HARD TO COMPUTE}}}}$$

$$\overbrace{p(x_1, \dots, x_n)}^{(a)} = \overbrace{p(x_2, \dots, x_n \mid x_1)\, p(x_1) = p(x_3, \dots, x_n \mid x_2, x_1) \cdot p(x_2 \mid x_1)\, p(x_1)}^{(b)}$$

$$= \underbrace{p(x_n \mid x_1 \dots x_{n-1})}_{p(x_n \mid x_{n-1})}\, \underbrace{p(x_{n-1} \mid x_1 \dots x_{n-2})}_{p(x_{n-1} \mid x_{n-2})} \cdots \underbrace{p(x_3 \mid x_1, x_2)}_{p(x_3 \mid x_2)}\, p(x_2 \mid x_1)\, p(x_1)$$

$$\underbrace{p(x_n \mid x_{n-1}) \cdot p(x_{n-1} \mid x_{n-2}) \cdots p(x_3 \mid x_2) \cdot p(x_2 \mid x_1)\, p(x_1)}_{(c)}$$

$x_1, \dots, x_n \in \{1, \dots, K\}$

(a) $O(K^n)$ MEMORY $[p(x_1, \dots, x_n)]$

STORING $p(x_1) =$

1	2	3	$\dots$	K
p_1	p_2	p_3	$\dots$	p_K

$K-1$ doubles $(p_1 + \dots + p_K = 1)$

$\boxed{O(K-1) = O(K)}$

(b) $\underbrace{p(x_n | x_1 \dots x_{n-1})}_{O(K^n)} + \dots \underbrace{p(x_2|x_1)}_{O(K^2)} + \underbrace{p(x_1)}_{O(K)}$

$O(K^n)$ MEMORY

$\underbrace{p(x_2|x_1)}$

$O(K(K-1))$

$O(K^2)$

x_1/x_2	1	$\dots$	K	
1	p_{11}	$\dots$	p_{K1}	$\sum p_{j1} = 1$
$\vdots$			$\vdots$	
K	p_{1K}	$\dots$	p_{KK}	$\sum p_{jK} = 1$

(c) $\underbrace{p(x_n | x_{n-1})}_{O(K^2)} \dots \underbrace{p(x_2|x_1)}_{O(K^2)} \underbrace{p(x_1)}_{O(K)}$

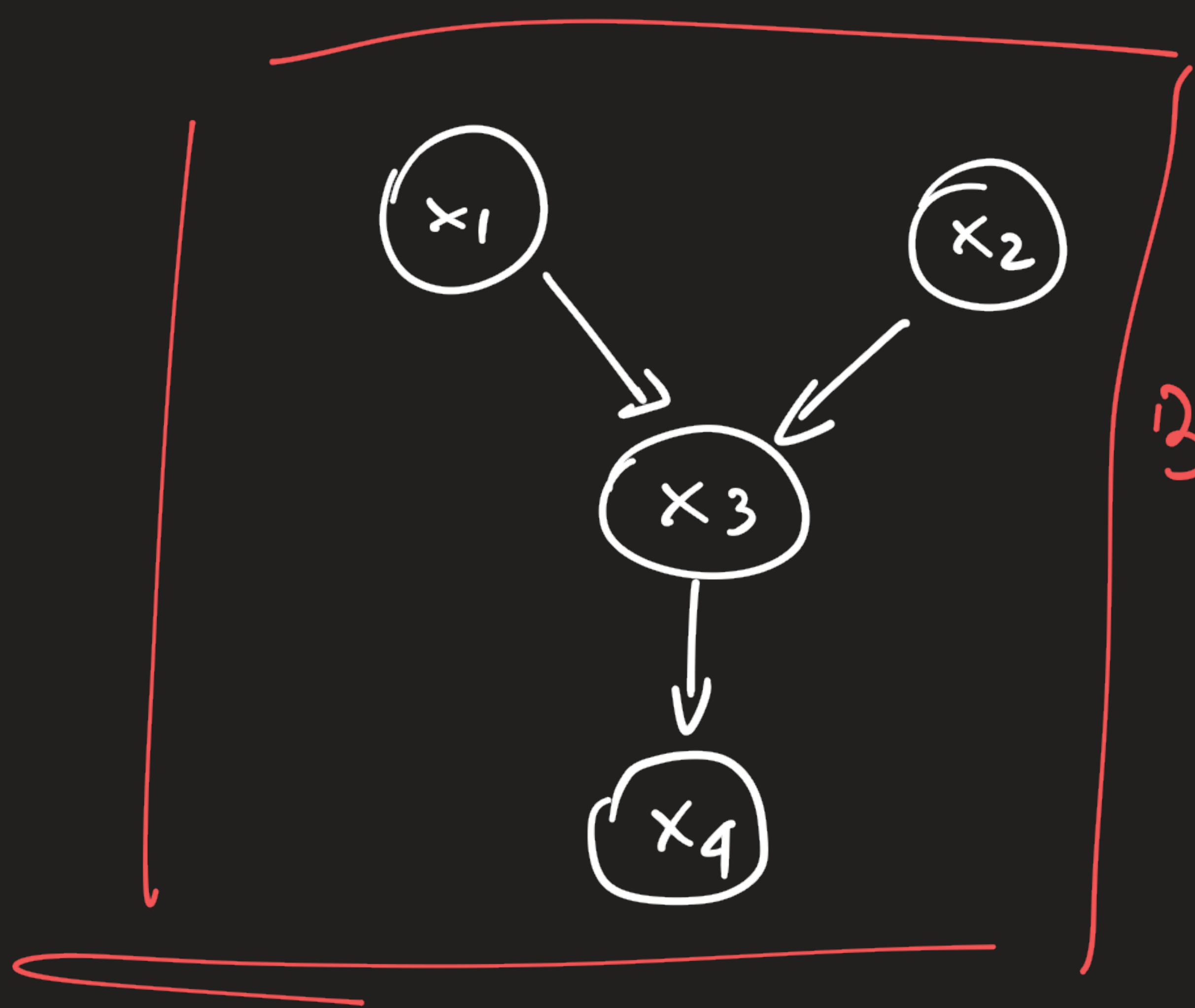
$O(nK^2)$

HOW TO REPRESENT FACTORIZATIONS IN A NICE WAY ?

- BAYESIAN NETWORKS (DIRECTED)
- MARKOV RANDOM FIELDS (UNDIRECTED)

(• FACTOR GRAPHS)

$$p(x_1, x_2, x_3, x_4) = p(x_4 | x_3)\, p(x_3 | x_1, x_2)\, p(x_2)\, p(x_1)$$



BAYESIAN NETWORK

BAYESIAN NETWORKS

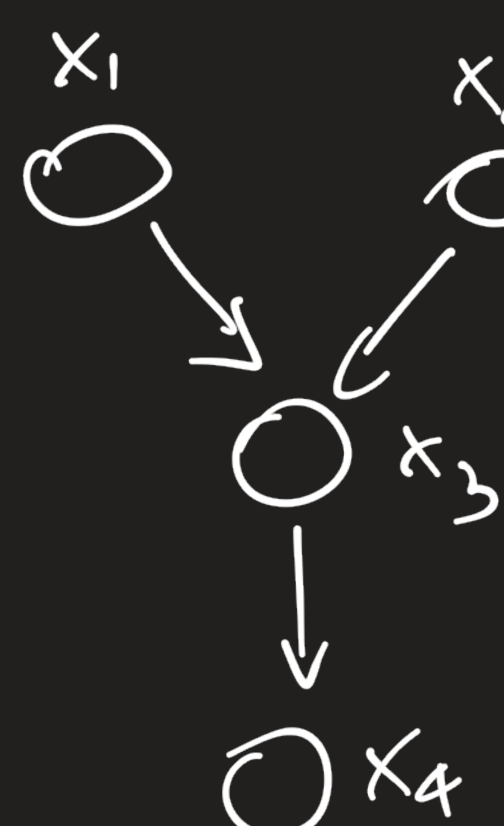
$$p(x_1,x_2,x_3,x_4) = p(x_4|x_3)\,p(x_3|x_1,x_2)\,p(x_2)\,p(x_1)$$

pa_k = parents of x_k $\quad pa_k \subseteq \{x_1, \dots, x_{k-1}\}$ are in the conditioning set of x_k

$\Downarrow$

$p(x_k | pa_k)$ is the factor for x_k

BAYESIA NETWORK is DAG, $\{x_1, \dots, x_n\}$ VERTICES



(x_i, x_j), $i<j$ and $x_i \in pa_j$

CONVENTION

OBSERVED NODES — SHADOWED

PLATED NODES

DETERMINISTIC QUANTITIES $\bullet^{\alpha}$ solid circles

$$\underbrace{p(x,z) = p(x|z)}_{\mathcal{N}(x \mid \mu(z), \sigma^2)} p(z)$$

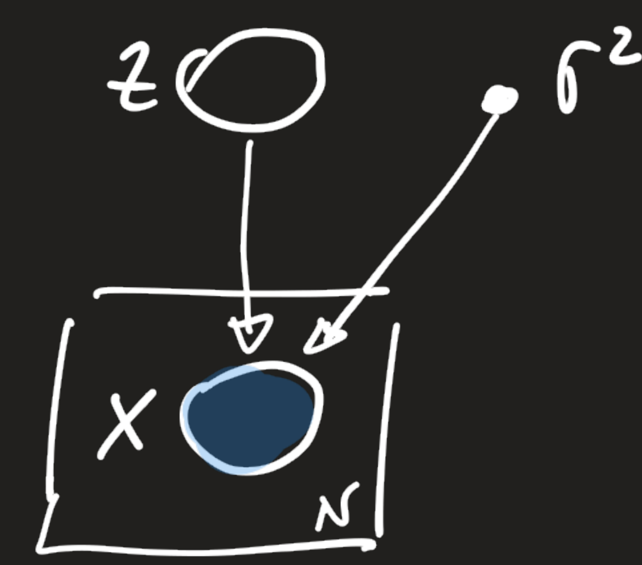
x: continuous $\mathcal{N}(\mu_z, \sigma^2)$

z: mixture component, discrete $\{1 \ldots K\}$

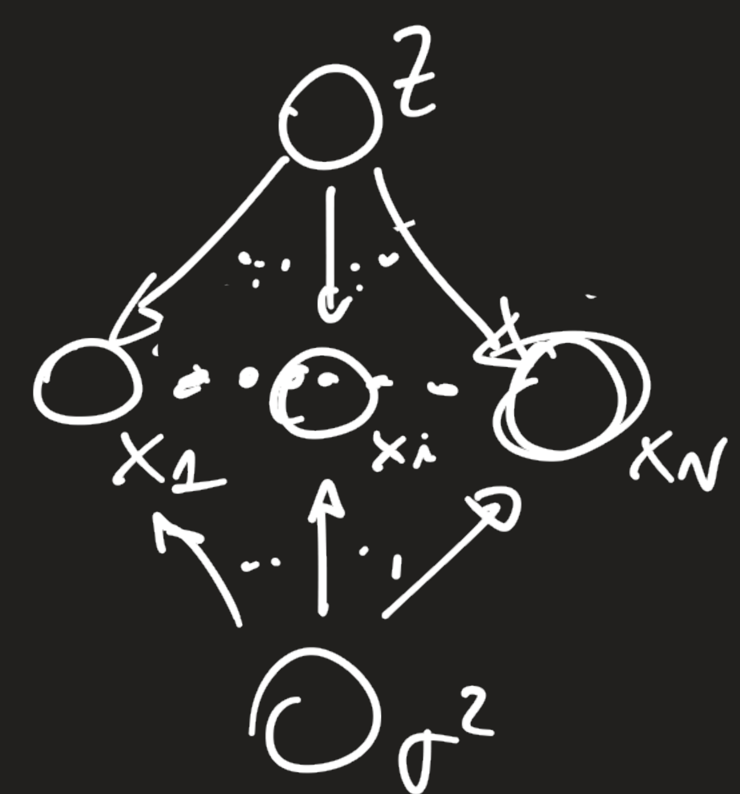
σ^2 constant

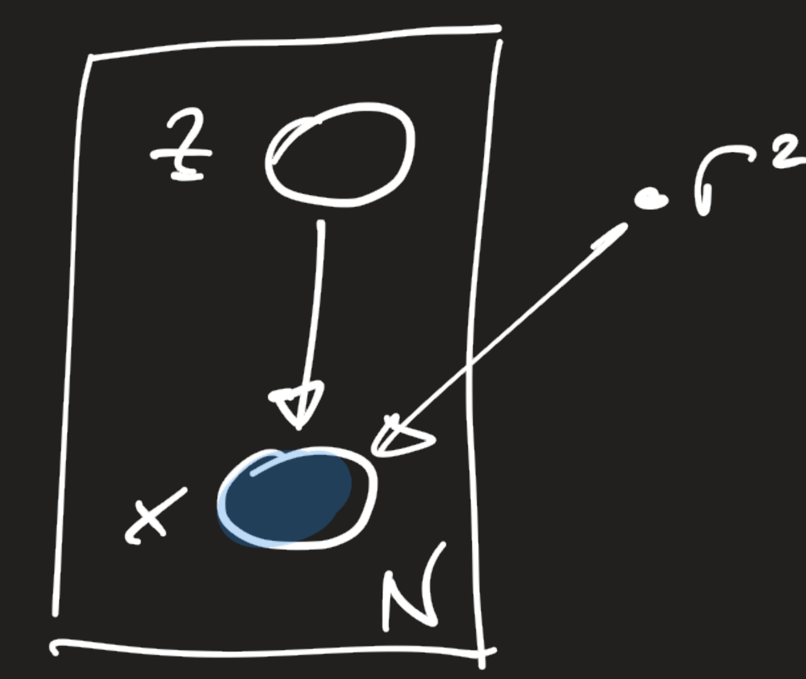
$x_1, \ldots, x_N = \underline{x}$

$$\underbrace{p(z \mid \underline{x})}_{p(z_1 \mid x_1)}$$



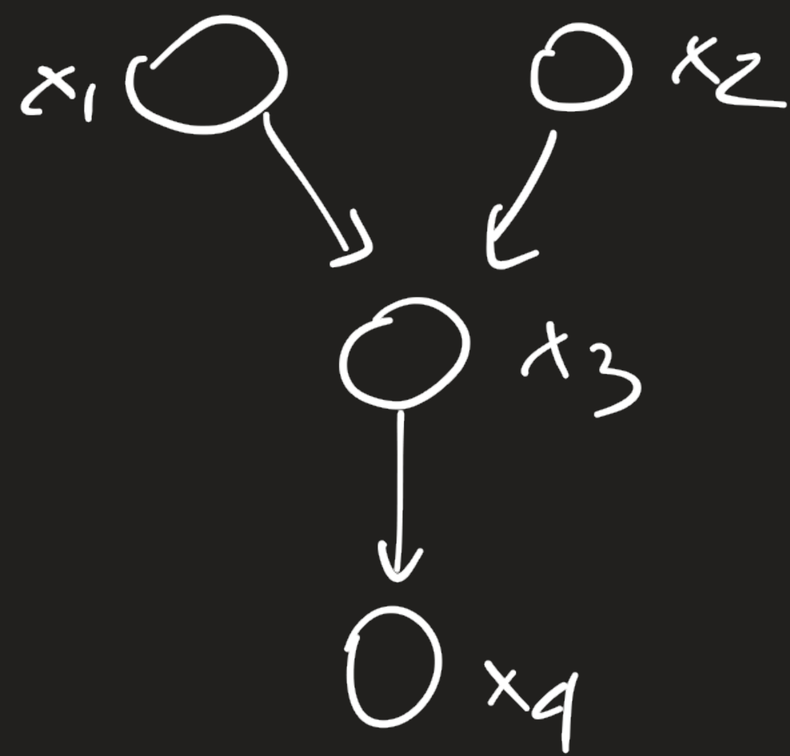
$\underline{x}$ come from the same z





ANCESTRAL SAMPLING

$$p(x_1, x_2, x_3, x_4) = p(x_4 \mid x_3)\, p(x_3 \mid x_1, x_2)\, p(x_2)\, p(x_1)$$

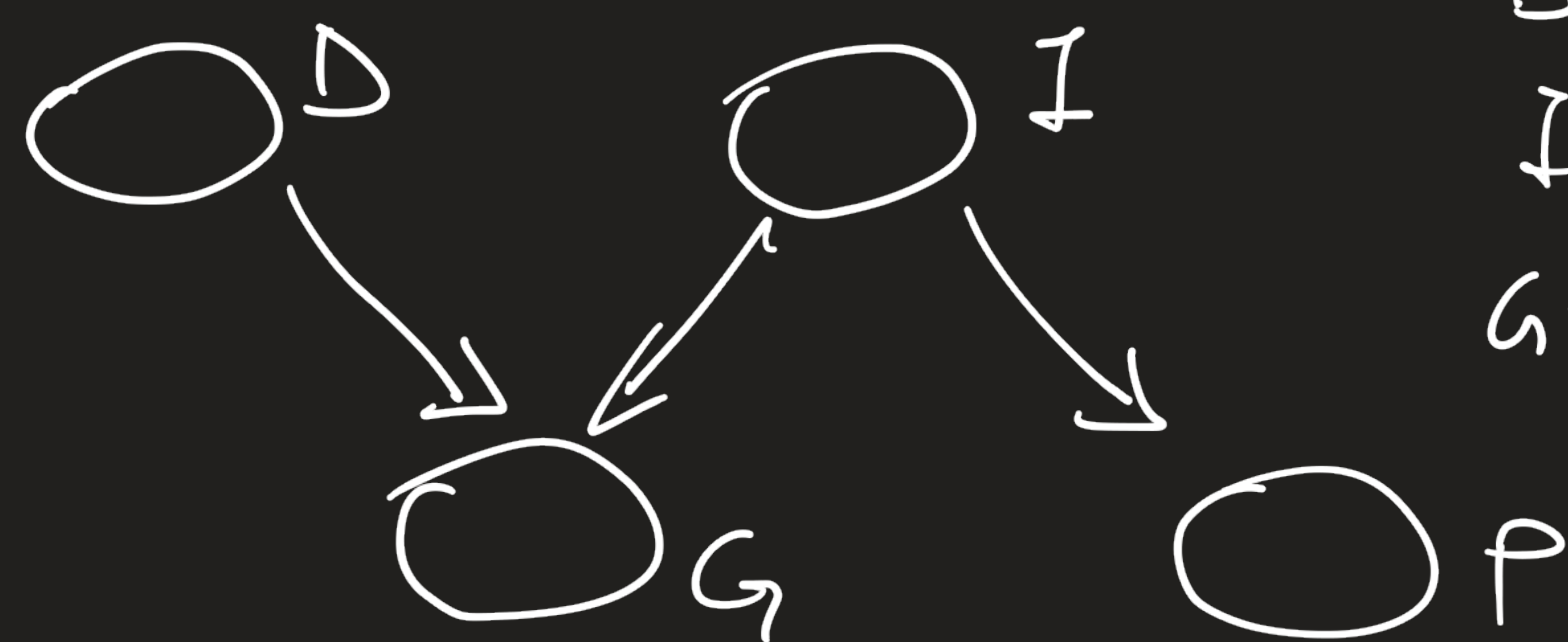


From top to bottom

from ancestrals to descendants

IF WE CAN SAMPLE FROM $\boxed{p(x \mid pa_x)}$

$P(D, I, P, G)$

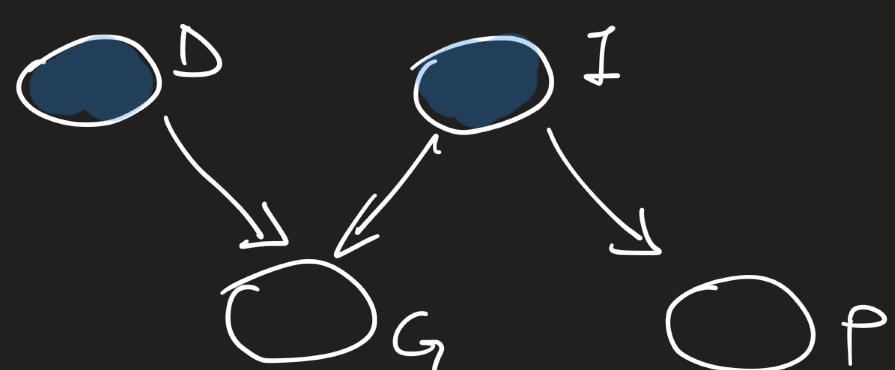


D = DIFFICULTY

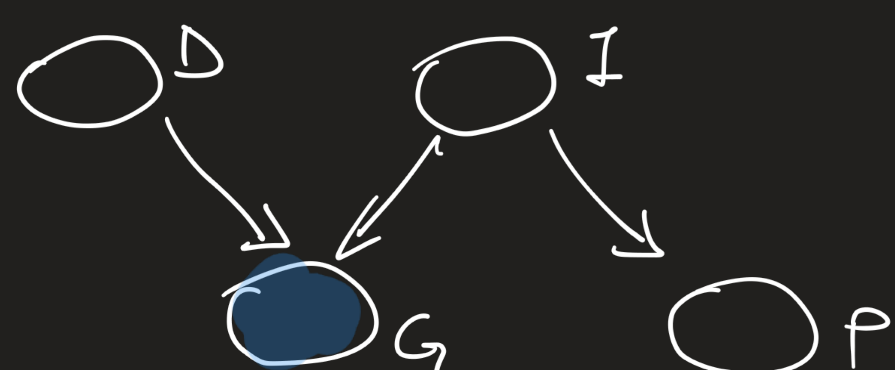
I = INTELLIGENCE

G = GRADE

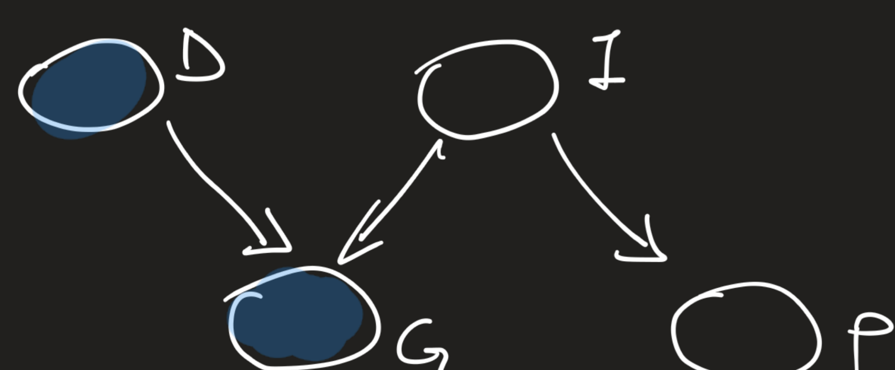
P = PASSED A HARD EXAM



CAUSAL REASONING



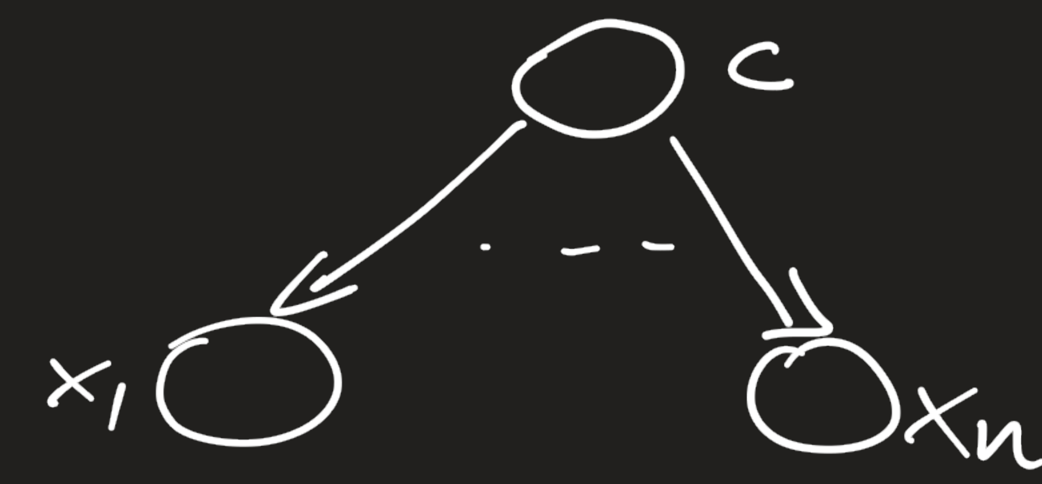
EVIDENTIAL REASONING



INTERCAUSAL REASONING

NAIVE BAYES

$p(x_1,\dots,x_n \mid c)$ ~ generative model class conditional



$p(c \mid \overbrace{x_1,\dots,x_n}^{\text{FEATURES}})$

$$p(x_1,\dots,x_n \mid c) = \prod_{i=1}^{n} p(x_i \mid c)$$

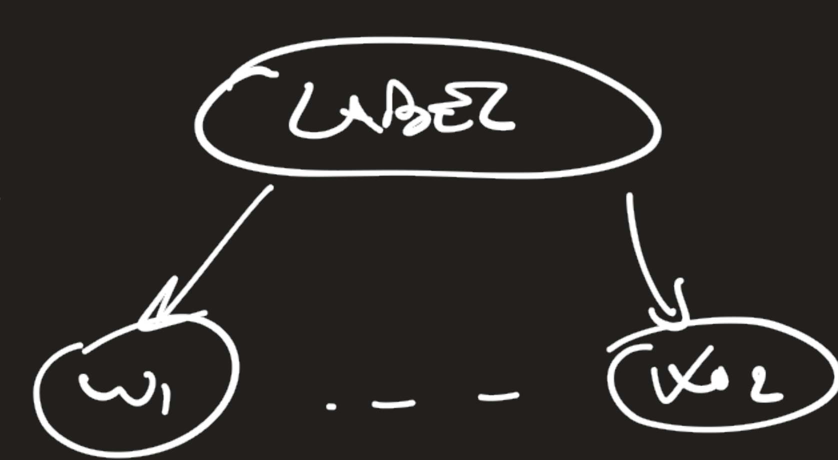
TRAIN N.B. given c, fix x_i, consider observation only of feature x_i for points of class c, and fit a model of $\underline{p(x_i \mid c; \theta)}$

$p(c)$ estimated

$$p(c \mid x_1 \dots x_n) \propto p(c) \cdot p(x_1 \dots x_n \mid c) = p(c) \prod_i p(x_i \mid c)$$

$$\frac{p(c=0 \mid x_1 \dots x_n)}{p(c=1 \mid x_1 \dots x_n)}$$ GOOD FOR CLASSIFICATION ONLY

$D_j : \begin{cases} 0 \\ 1 & \text{if word } j \text{ is in DOC} \end{cases}$



W_i = # of occurrences of word j in doc.

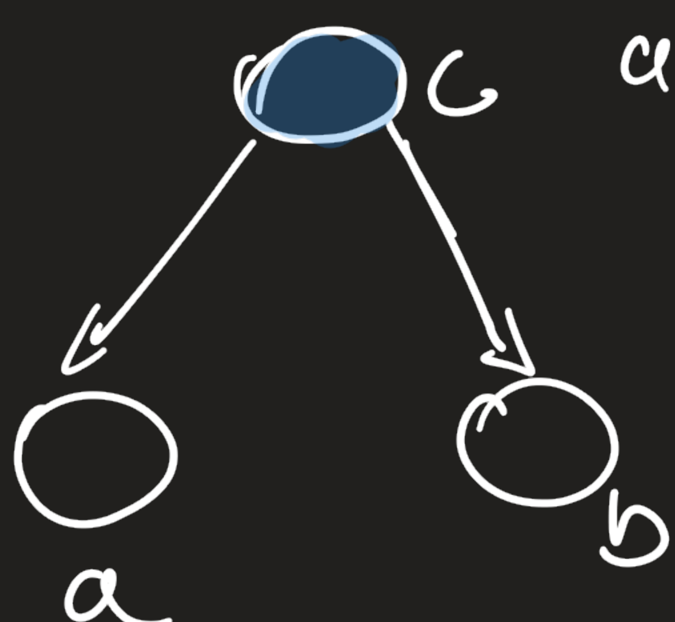
CONDITIONAL INDEPENDENCE IN BAYESIAN NETWORKS

def: a, b, c random variables.

a is COND. INDEP. of b given c $\quad a \perp\!\!\!\perp b \mid c$

iff

$$p(a \mid b, c) = p(a \mid c) \quad \text{OR} \quad p(a, b \mid c) = p(a \mid c)\, p(b \mid c)$$



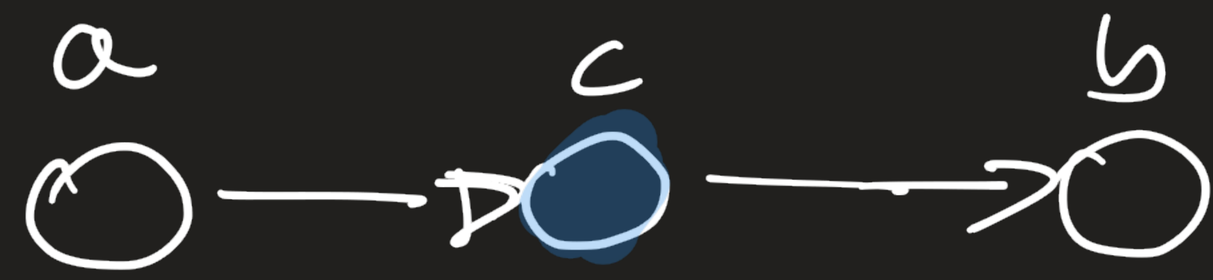
a, b TAIL TO TAIL with c

$$p(a,b) = \sum_c p(a \mid c)\, p(b \mid c)\, p(c) \neq p(a)\, p(b)$$

$$p(a,b \mid c) = \frac{p(a,b,c)}{p(c)} = \frac{p(a \mid c)\, p(b \mid c)\, \cancel{p(c)}}{\cancel{p(c)}} = p(a \mid c)\, p(b \mid c)$$

$a \not\perp\!\!\!\perp b$

$a \perp\!\!\!\perp b \mid c$



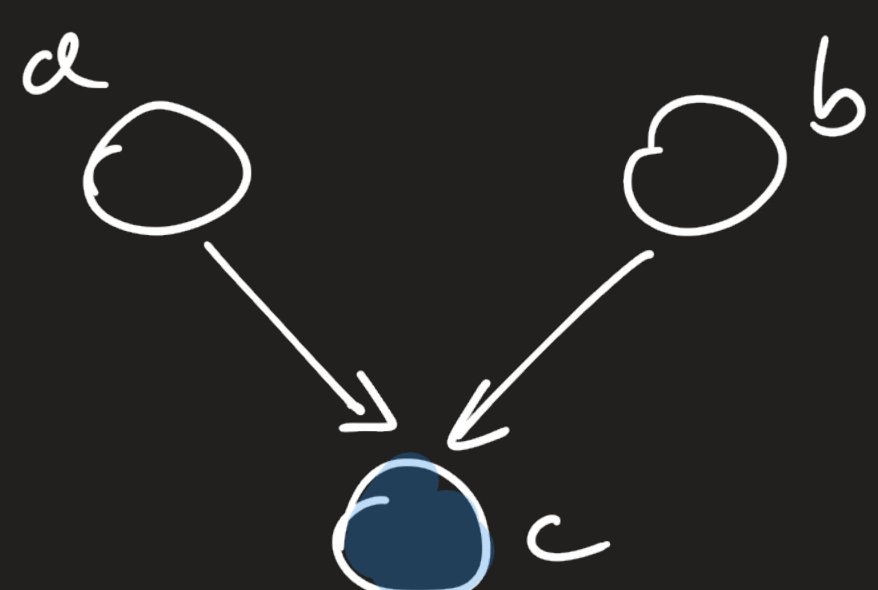
HEAD TO TAIL

$a \not\perp\!\!\!\perp b$

$$p(a,b) = \sum_c \underbrace{p(b|c)p(c|a)} p(a) = p(b|a)p(a)$$

$a \perp\!\!\!\perp b \,|\, c$

$$p(a,b|c) = p(b|c)\underbrace{\boxed{\frac{p(c|a)p(a)}{p(c)}}}_{p(a|c)} = p(b|c)p(a|c)$$



HEAD TO HEAD

$$p(a,b) = \sum_c p(a)p(b)p(c|a,b) = p(a)p(b)$$

$$p(a,b|c) = \frac{p(a)p(b)p(c|a,b)}{p(c)} \neq p(a|c)\,p(b|c)$$

$a \perp\!\!\!\perp b$

$a \not\perp\!\!\!\perp b \,|\, c$ $\rightarrow$ also $a \not\perp\!\!\!\perp b \,|\, d$ - for d descendant of c

a, b, c , a path from a to b is <u>blocked</u> by c iff

- c is observed and the path is head to tail or tail to tail in c
- c is not observed, nor any descendant of c, and the path is head to head in c.

let A, B, C subsets, if all paths from a node in A to a node in B are blocked by a node in C,

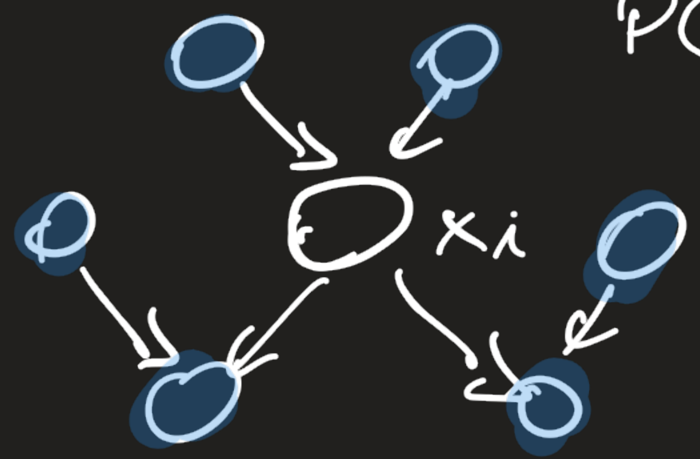
$$A \perp\!\!\!\perp B \mid C$$

MARKOV BLANKET

x_i, condition on $X_{\setminus i}$. Which nodes will remain in the conditioning set?

$$p(x_i \mid X_{\setminus i}) = \frac{p(x_1 \dots x_i \dots x_n)}{P(x_1 \dots x_{i-1} x_{i+1} \dots x_n)} = \frac{\prod_j p(x_j \mid pa_j)}{\sum_{x_i} \prod_j p(x_j \mid pa_j)}$$

only pa_i or those for which $x_i \in pa_j$



MARKOV BLANKET of x_i
- pa_i
- children
- co parents