Summary Prev. Week

▷ **Bayesian Learning:** → Joint Probability Distribution

↳ Bayes Rule: $P(Y|X_1 \dots X_n) = \dfrac{P(X_1 \dots X_n|Y)\,P(Y)}{P(X_1 \dots X_n)}$

↳ Learning Problem: $f: X \to Y$ ⟹ $\text{Prob}(Y|X_1 \dots X_n)$

(Probabilistic)

▷ # Issues: Data Sparsity

↳ Smart Probability Estimation → MLE [$\text{Prob}(\mathcal{D}|\theta)$] → MAP [$\text{Prob}(\theta|\mathcal{D})$]

▷ **Bayes Classifier:**

↳ Discrete $Y$, Discrete $X_i$ ⟹ Naive Bayes Algorithm

[From: EXP. estimations in JPD]

Assume – Conditional Independence ($X_i$)

$$P(Y=y_k|X_1 \dots X_n) = \frac{\overbrace{P(Y=y_k)\cdot \prod_i P(X_i|Y=y_k)}^{\text{JD}}}{\sum_j P(Y=y_j)\cdot \prod_i P(X_i|Y=y_k)}$$

(Linear Est.)

Classification Problems

↳ Discrete $Y$, Continuous $X_i$ ⟹ Gaussian Naive Bayes

$$P(X_i=x|Y=y_k) = \frac{1}{\sqrt{2\pi\sigma_{ik}^2}}\, e^{-\frac{1}{2}\left(\frac{x-\mu_{ik}}{\sigma_{ik}}\right)^2}$$

[$\mu_{ik}$ Mean, $\sigma_{ik}^2$ Var]

• Today: Smart Representation of JPD ⟹ Bayes Net.

## Conditional Independence: $X \perp Y \mid Z$

$(\forall i, j, k) \rightarrow P(X = x_i \mid Y = y_j, Z = z_k) = P(X = x_i \mid Z = z_k)$

$\boxed{\text{Alt}}$ $P(X = x_i, Y = y_j \mid Z = z_k) = P(X = x_i \mid Z = z_k) \cdot P(Y = y_j \mid Z = z_k)$

## Marginal Independence: $X \perp Y$

$(\forall i, j) \quad P(X = x_i, Y = y_j) = P(X = x_i) \cdot P(Y = y_j)$

$= P(X = x_i \mid Y = y_j) = P(X = x_i)$

## Proof of Cond. Ind.:

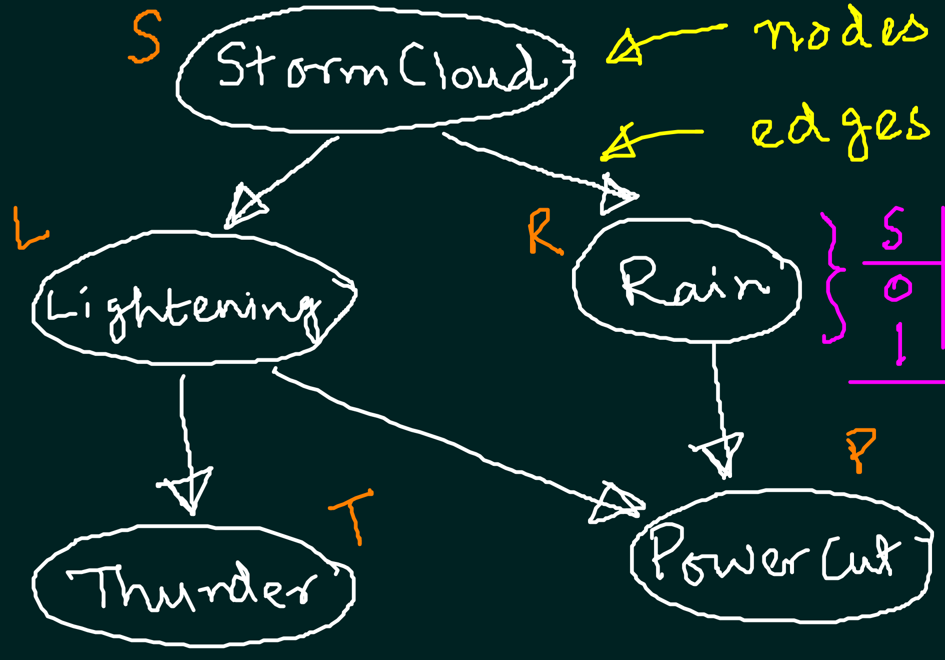
Assume, $P(X \mid YZ) = P(X \mid Z)$

Chain Rule

Now, $P(XY \mid Z) = \dfrac{P(XYZ)}{P(Z)} = \dfrac{\underbrace{P(X \mid YZ)}_{P(X \mid Z)} P(Y \mid Z) P(Z)}{P(Z)}$

$= \overbrace{P(X \mid Z)} P(Y \mid Z)$ — (Proved)

[Do the Reverse eqv. yourself !!]




# Bayesian Network

| S | R=1 | R=0 |
|---|---|---|
| 0 | | |
| 1 | | |

| S | L | R | T | P | → Prob |
|---|---|---|---|---|---|
| 1 | 0 | 1 | 0 | 1 | → 0.01 |
| . | | . - - - - | | | - - . |
| | - - - - | | | | . |

| Parent L | Parent R | P = 1 Prob | P = 0 Prob |
|---|---|---|---|
| 0 | 0 | 0.01 | 0.99 |
| 0 | 1 | - | - - |
| 1 | 0 | | - - - . |
| 1 | 1 | | |

JPD

DAG

(Graph)

[Causal dependencies]

→ $G = (V, E)$

$V = \{\text{Set of Attributes}\}$

$E = (n_1, n_2)$

$n_1 \longrightarrow n_2$

→ $Par(x_i)$ = Parent of $x_i$

→ $JPDT(x_i, Par(x_i))$

Bayesian Network

→ $P(x_1 \ldots x_n) = \prod_i P(x_i \mid Parent(x_i))$

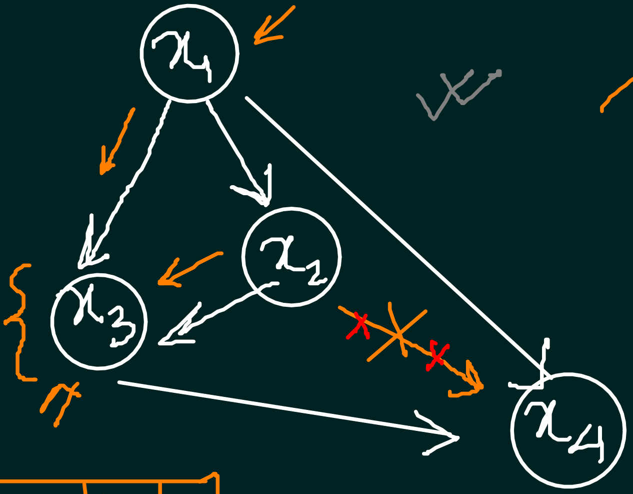
$P(S, L, \neg R, P, \neg T)$

$\Downarrow$

$P(S) * P(L|S) \cdot P(\neg R|S)$

$\boxed{P(P|L, \neg R)}$

$P(\neg T|L)$




| $x_1$ | $x_2$ | |
|---|---|---|
| - - | - - | 1 |
| - | — | + |

[GRAPHICAL MODELS]

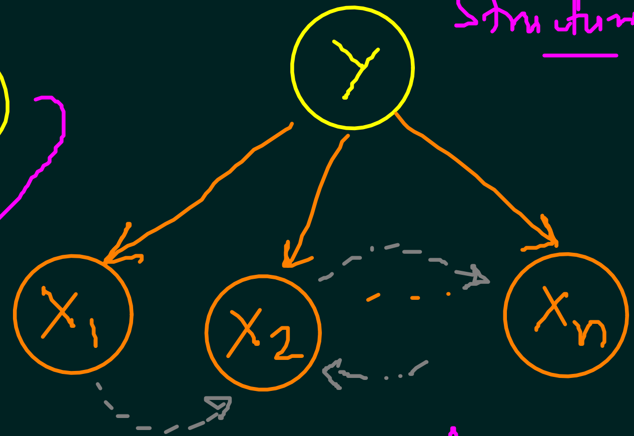
$$P(x_1, x_2, x_3, x_4)$$
$$= P(x_1) \cdot P(x_2 \mid x_1) \cdot P(x_3 \mid x_1, x_2) \cdot P(x_4 \mid x_1, \cancel{x_2}, x_3)$$

$$\left[ P(x \mid yz) = P(x \mid z) \text{ when } x \perp y \mid z \right]$$

▷ Bayes Net for Naive Bayes:

$$\boxed{P(X_1, X_2 \ldots X_n \mid Y)} = \frac{\prod_i P(X_i \mid Y)}{P(Y)}$$

Bayesian Network Structure



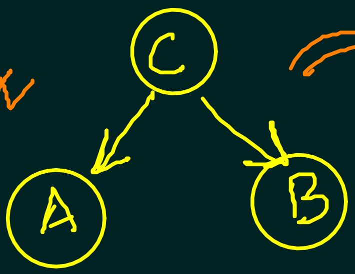

↳ Partially Cond. Depend.

Bayesian Network prob. Est.

$$P(X_1, X_2, \ldots, X_n, Y)$$

▷ 2 ement: Store Graph and Traverse } Algo.

$P(XY|Z) = P(X|Z) \cdot P(Y|Z) \iff X \perp Y \mid Z$




$$P(AB|C) = \frac{P(ABC)}{P(C)} = \frac{P(A|C) \cdot P(B|C) \cdot P(C)}{P(C)}$$

$$= P(A|C)\, P(B|C)$$

$A \perp B \mid C$

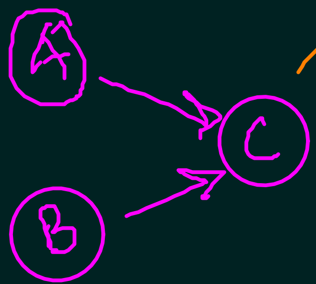
(Tail-to-tail)

$$P(AB|C) = \frac{P(ABC)}{P(C)} = \frac{P(C|A) \cdot P(B|C) \cdot P(A)}{P(C)}$$

$$= P(A|C)\; P(B|C)$$


$A \perp B \mid C$

(Head-to-tail)




(Head to Head)

$\boxed{A \perp B \mid C}$ ?

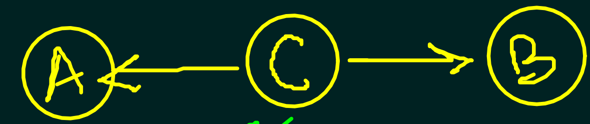
$\times$

$$P(AB|C) = \frac{P(ABC)}{P(C)} = \frac{P(C|A) \cdot P(C|B) \cdot P(A) \cdot P(B)}{P(C)}$$

$$\neq P(A|C) \cdot P(B|C)$$

✓

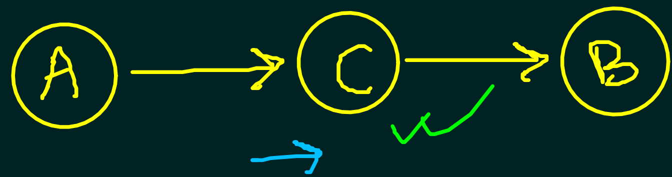
## D-Separation Algo. : (Conditional Indep.) $A \perp B \mid C$



1

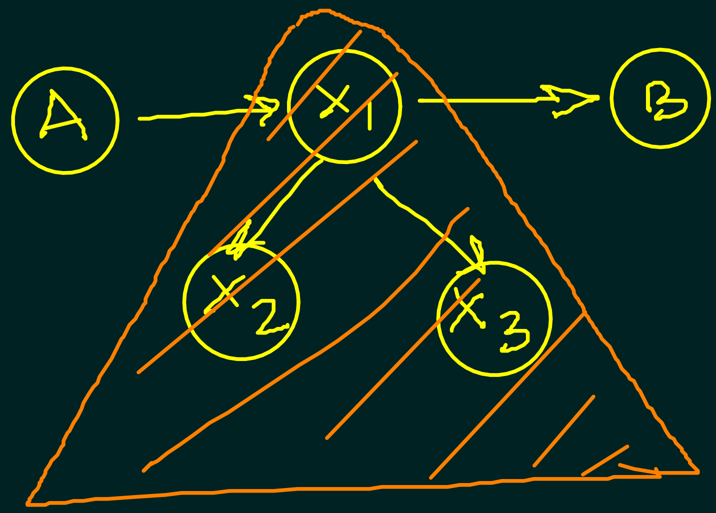
OR

2



OR

3



$C \notin \{x_1, x_2, \dots\}$

?

(Bishop) [Ch. 8]

Ex:

$X_1 \perp X_3 \mid X_2 \Rightarrow$ Yes

$X_1 \perp X_4 \mid X_2 \Rightarrow$ Yes ✓

$X_1 \perp X_4 \mid \{X_2, X_3\} \Rightarrow$ No ✓

$X_1 \perp X_4 \mid \{\phi\} \Rightarrow$ ??