Chain Rule and CI

Bayesian Networks

Ordering of Variables and Causality

ORIE 4742

Mar 22, 2021

Spencer Peters

Reading off CI's — d-separation

BN as convenient notations
for complex models

Chain rule and conditional probabilities

$$P(X_1, X_2, X_3) = P(X_1)\,P(X_2 \mid X_1)\,P(X_3 \mid X_1, X_2)$$

$$\text{indep} \Rightarrow = P(X_1)\,P(X_2)\,P(X_3)$$

$$P(X_1, X_2, X_3) = P(X_1)\,P(X_2 \mid X_1)\,P(X_3 \mid X_1, X_2)$$

$$X_3 \perp X_1 \mid X_2 \Rightarrow = P(X_1)\,P(X_2 \mid X_1)\,P(X_3 \mid X_2)$$

Conditional independence

$$P(X_1, X_2, X_3) = P(X_1)\,P(X_2|X_1)\,P(X_3|X_1, X_2)$$

X_1	$P(X_1)$
0	0.1
1	0.9

X_1	X_2	$P(X_2\|X_1)$
0	0	0.4
0	1	0.3
1	0	0.6
1	0	0.7

X_1	X_2	X_3	$P(X_3\|X_1, X_2)$
0	0	0	0.1
0	0	1	0.9
0	1	0	0.2
0	1	1	0.8
1	0	0	0.7
1	0	1	0.3
1	1	0	0.1
1	1	1	0.9

$X_1 = 1 \quad X_2 = 1 \quad X_3 = 0$

$$P((1,1,0)) = 0.9 \cdot 0.7 \cdot 0.2$$

Bayesian Networks (BNs)

$$P(X_1, X_2, X_3) = P(X_1)\,P(X_2 \mid X_1)\,P(X_3 \mid X_1, X_2)$$

Bayes Net B is
a directed acyclic graph
(the nodes N are variables)
and conditional probabilities (CPT)

$$P(X_i \mid Pa(X_i))$$

for each $X_i \in N$.

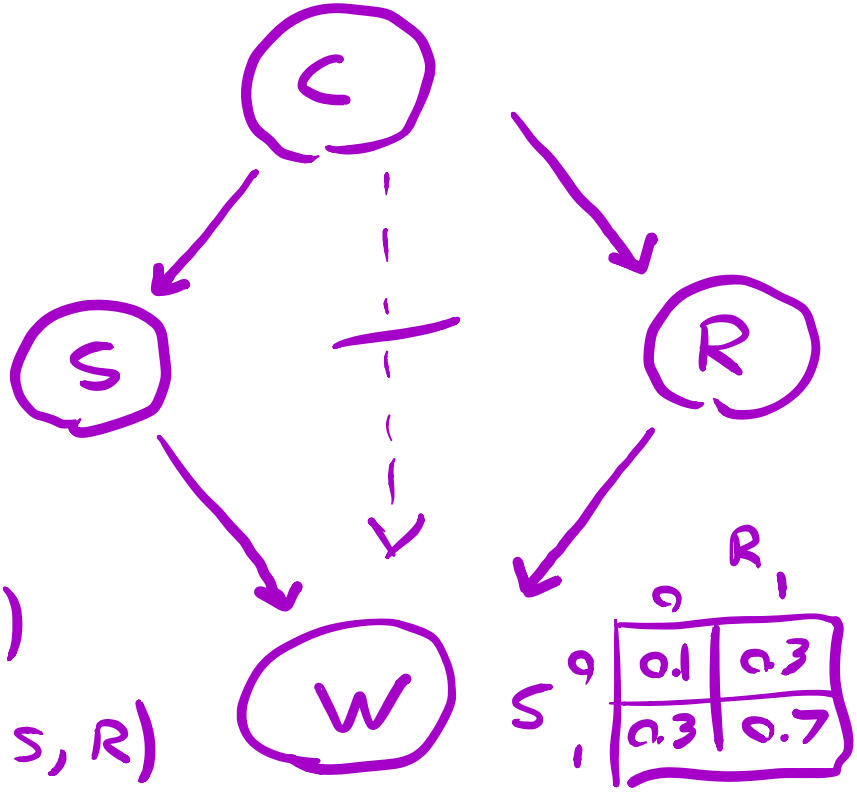
Rain or Sprinkler?

W: Grass wet

R: raining

S: sprinklers on

C: Cloudy

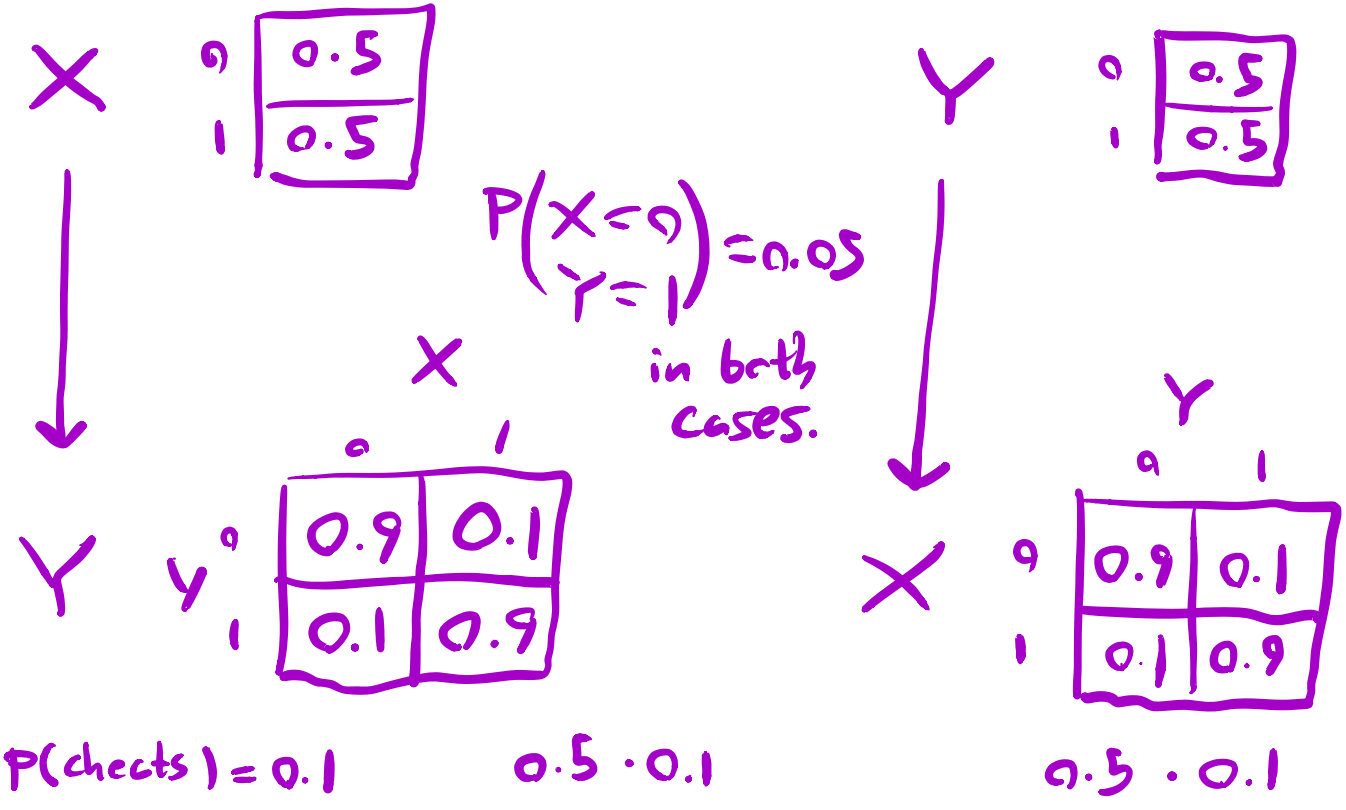


$P(W \mid S, C) > P(W \mid S)$

$P(W \mid S, C, R) = P(W \mid S, R)$

$W \perp C \mid S, R$

Ordering of variables in BNs and causality (1/2)



Ordering of variables in BNs and causality (1/2)

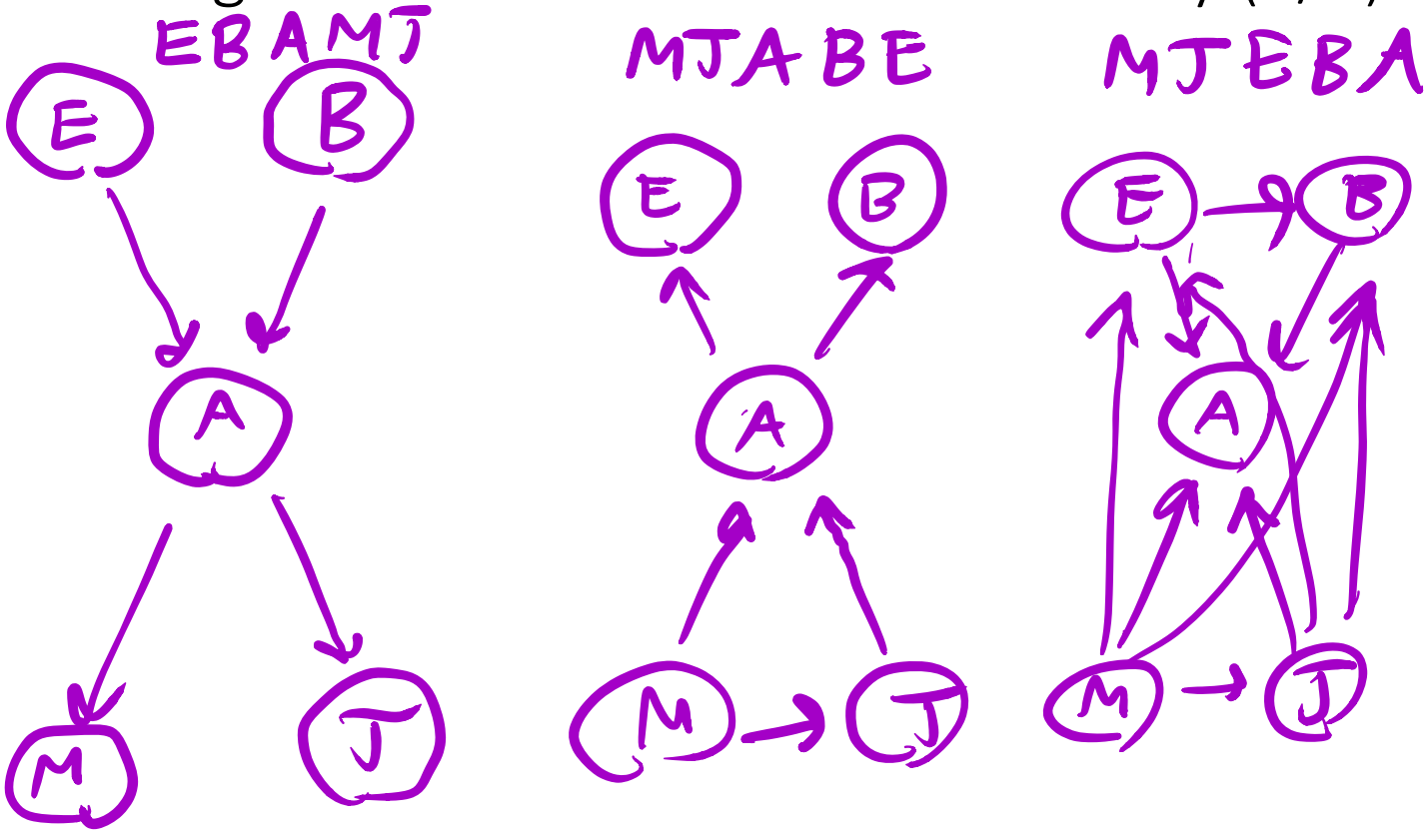
X = coin flip $\{0, 1\}$

Y = student's report $\{0, 1\}$

$$P(x, y) = P(x)P(y|x)$$

$$P(x, y) = P(y)P(x|y)$$

Ordering of variables in BNs and causality (2/2)

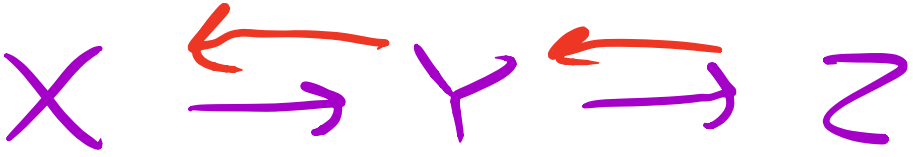


Reading off a BN's conditional independences (d-separation)

$$P(x,y) = P(x)P(y) \quad (= P(y)P(x))$$

$$X \perp Y \Rightarrow Y \perp X$$

Chains



$$P(Z \mid X, Y) = P(Z \mid Y) \Leftrightarrow Z \perp X \mid Y$$

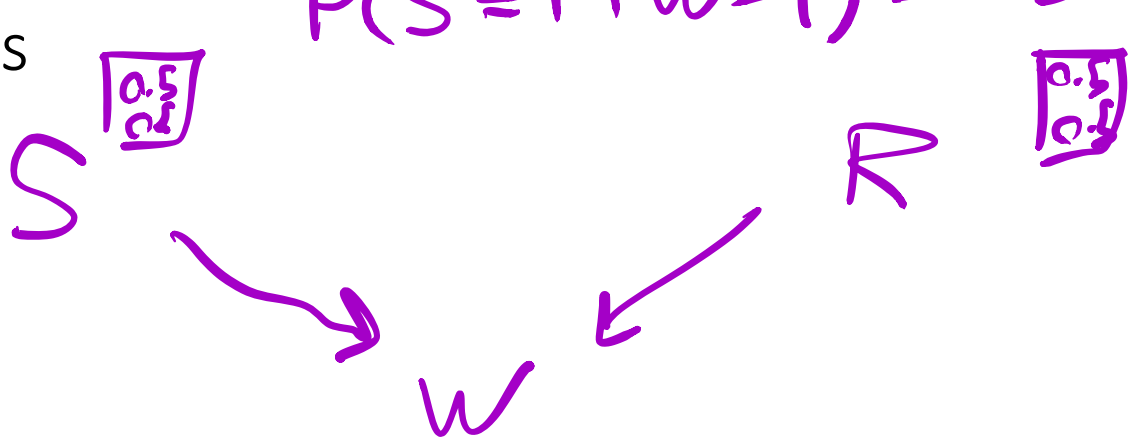
$$P(X \mid Y, Z) = P(X \mid Y) \Leftrightarrow X \perp Z \mid Y$$

Forks

$$P(X \mid Y, Z) = P(X \mid Y)$$
$$P(Z \mid Y, X) = P(Z \mid Y)$$

$$P(S=1 \mid W=1, R=0) = 1.$$
$$P(S=1 \mid W=1) = \frac{2}{3}$$

Colliders/Joins



$X \perp Z$

$X \perp Z \mid Y$ ✓✓

X

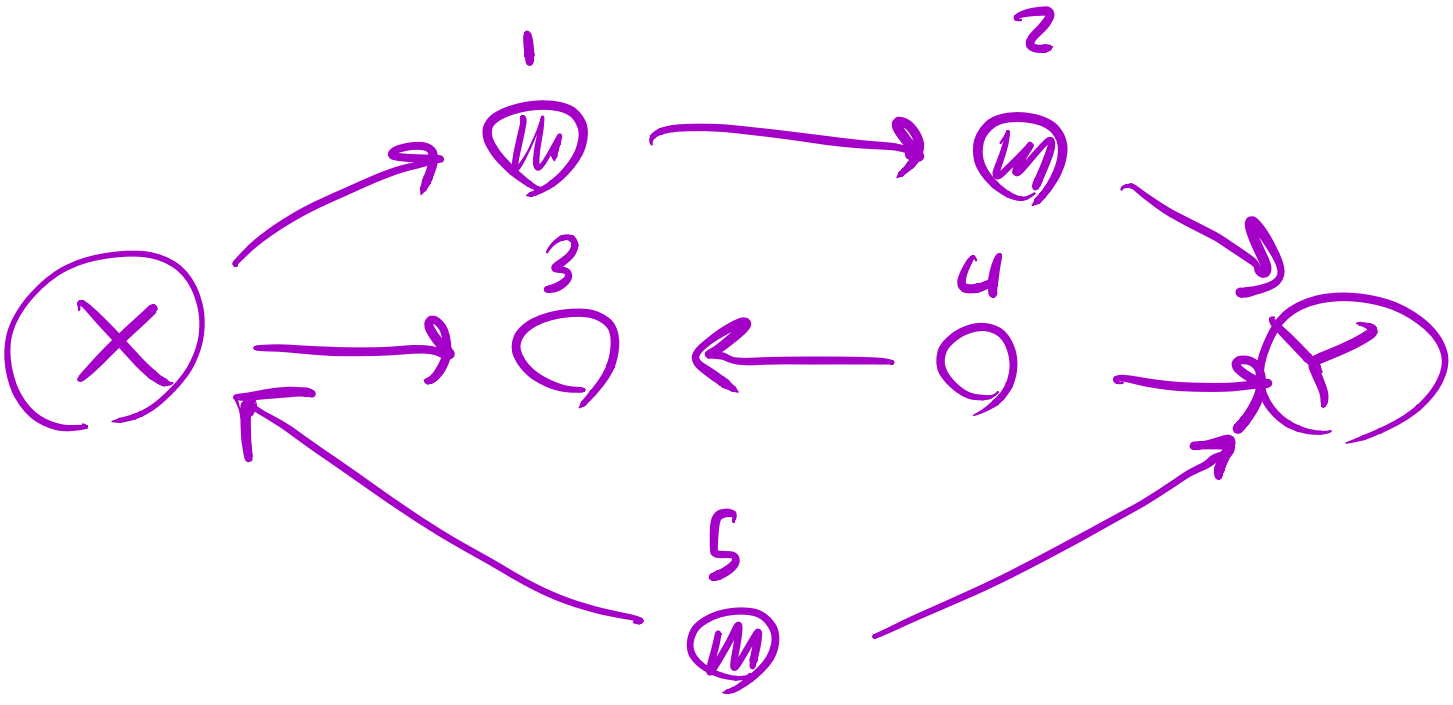
$P(W=1 \mid S=1 \text{ or } R=1)$

$=1$

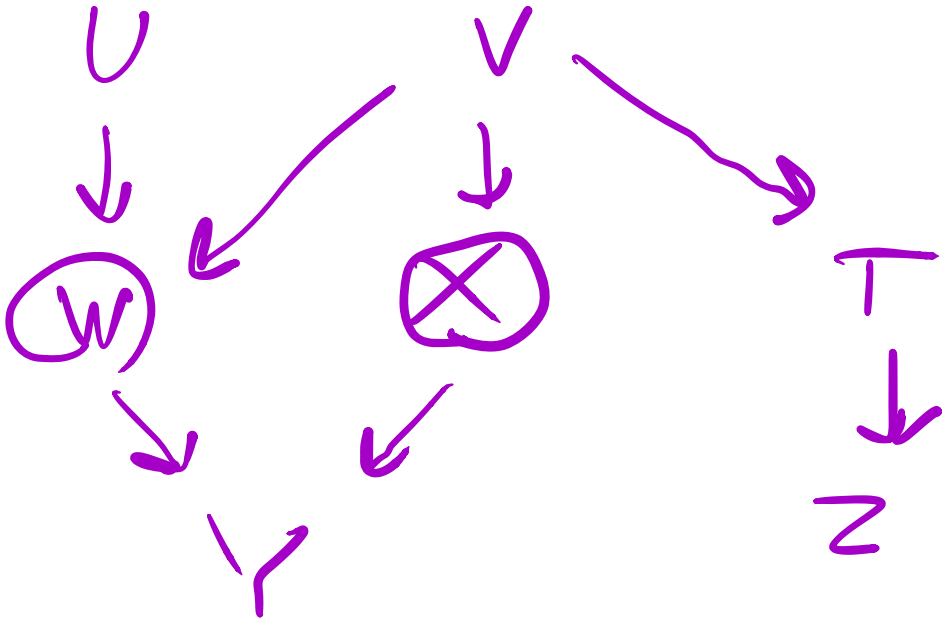
$P(W=1 \mid S=0 \text{ and } R=0)$

$=0$

d-Separation



d-Separation



Y ⊥ Z | X ?

False

'r ⊥ z | x, w

BNs as notation: Regression

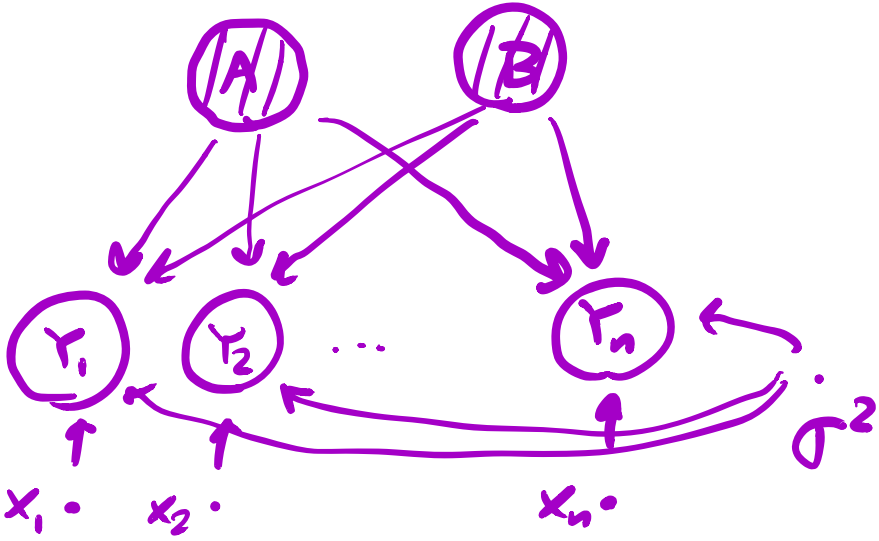
$(x_1, Y_1), (x_2, Y_2), \quad (x_n, Y_n)$

$Y_i = Ax_i + B + \varepsilon_i$

$\varepsilon_i = N(0, \sigma^2)$

$A = N(3, 1^2)$

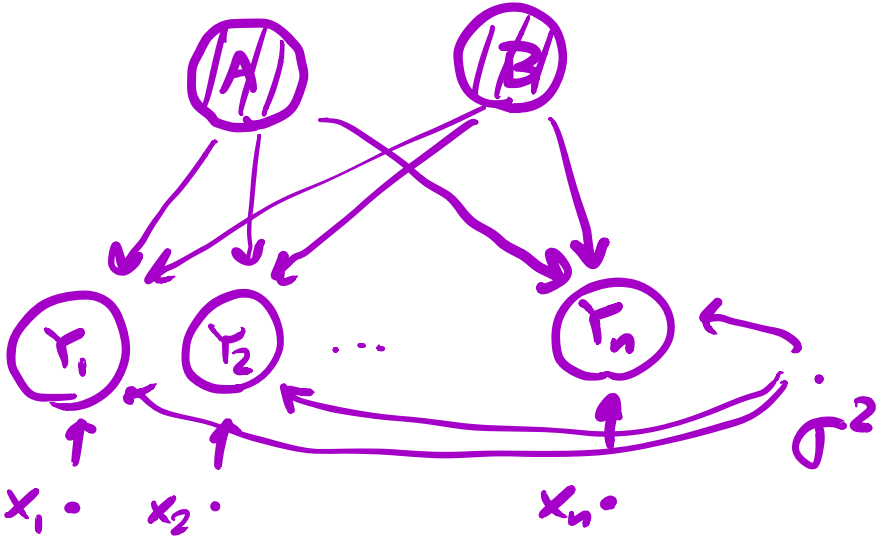
$B = N(0, 100^2)$

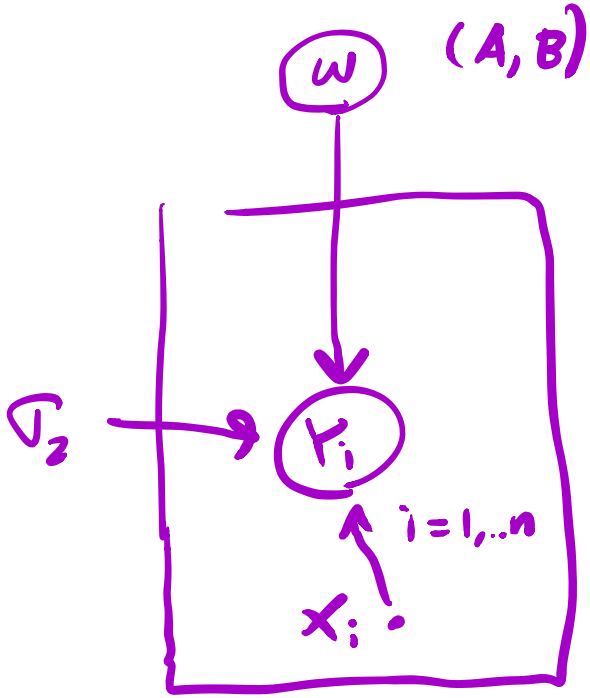


BNs as notation: Regression

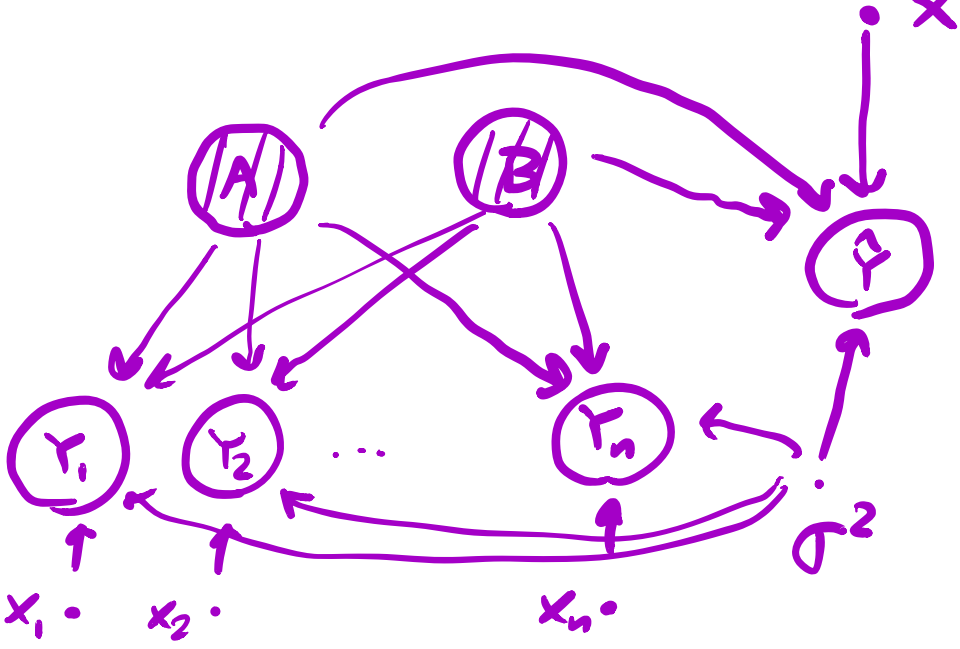
$(x_1, Y_1), (x_2, Y_2), \quad (x_n, Y_n)$

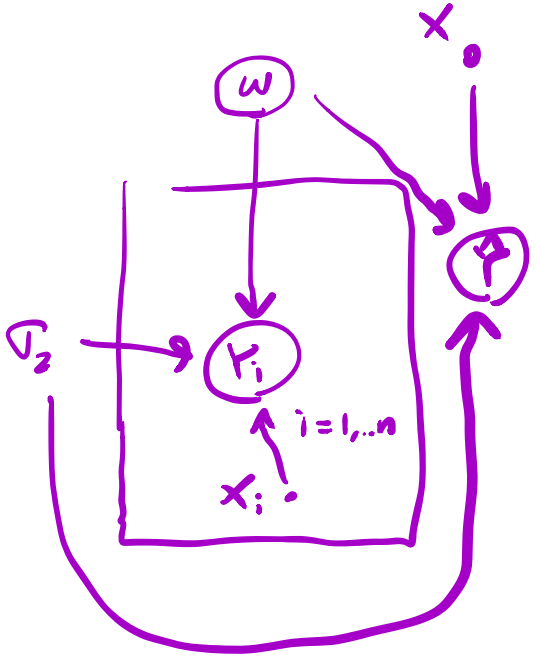
$Y_i = Ax_i + B + \varepsilon_i$





BNs as notation: Prediction





$$\hat{Y} = Ax + B + N(0, \sigma^2)$$

Bonus BNs as notation: Naïve Bayes

Categories $c_1, \ldots c_m$