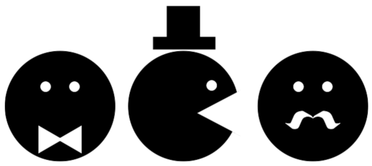
Worksheet 12: PGMs II*

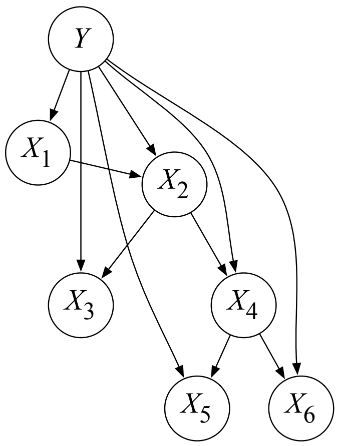
COMP90051 Statistical Machine Learning

Semester 2, 2024

Exercise 1. Mr. and Ms. Pacman have been searching for each other in the Pacman world (see http://ai.berkeley.edu/project_overview.html). Ms. Pacman has been pregnant with a baby, and this morning she has given birth to Pacbaby (congratulations, Pacmans!). To train Pacbaby to avoid encountering ghosts in the maze,[1] the Pacmans are trying to teach Pacbaby to distinguish Pacmen (pl.) from ghosts using discriminative visual features such as the presence of a bowtie, hat, mustache, etc.



Pacbaby has noticed that the features are not independent—nearly everyone who has a hat has a mustache, while those with bowties are always clean shaven. She decides to use a tree-augmented Naive Bayes model (TANB) to account for conditional dependencies. A TANB is an extension of a Naive Bayes model, where features are no longer assumed conditionally independent given the binary class $Y \in \{1, -1\}$ (Pacman or not-Pacman, respectively). Let $X_1, X_2, \ldots, X_6$ be the random variables corresponding to the features that Pacbaby observes. The TANB model arranges vertices in a tree-structured Bayes net with Y at the root:



(a) Assume all features $X_1, \ldots, X_6$ are observed in the TANB model. What is the classification rule? Your answer should be in terms of the prior and conditional probabilities.

*Based on Berkeley CS188 section

[1] Ghosts are nice enough not to eat Pacbaby, but they will take all her money.

When we perform a marginalisation operation—i.e. removing a variable from a joint distribution, we perform a sum over the product of all factors that include that random variable. For example, marginalising over X_4 in the joint distribution above involves a factor containing four random variables.

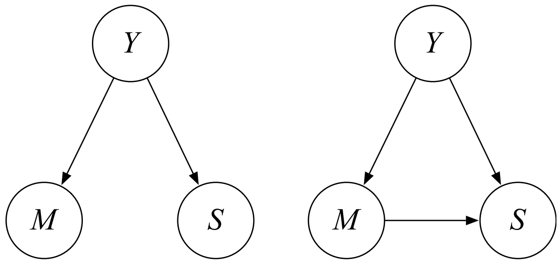
$$\sum_{X_4} \underbrace{p(X_4|X_2, Y)p(X_5|X_4, Y)p(X_6|X_4, Y)}_{\phi(X_2, X_4, X_5, X_6)}$$

This induces a dependency between all the random variables in the factor except the variable being marginalised—all subsequent operations will have to treat X_2, X_5, X_6 together (X_4 is summed out). Assuming there is no special algebraic structure in the summand that can be exploited, the complexity is exponential in the number of different random variables in the summand. Thus the overall complexity of the variable elimination algorithm is dominated by the number of variables in the largest elimination factor, $\phi(\ldots)$. Determining the optimal (lowest-complexity) elimination ordering is intractable, but a useful heuristic is to find an ordering that minimises the size of the largest factor generated.

(b) Specify an elimination order that is efficient for the query $p(Y|X_5 = x_5)$ in the TANB model above. How many variables are in the biggest factor induced by variable elimination with your ordering? Which variables are they?

(c) Specify an elimination order that is efficient for the query $p(X_3|X_5 = x_5)$ in the TANB model above. How many variables are in the biggest factor induced by variable elimination with your ordering? Which variables are they?

Exercise 2. Consider the Bayes nets below over the nodes Y (Pacbaby sees Pacman or not), M (Pacbaby sees a moustache), and S (Pacbaby sees sunglasses).



(Nb) (Tanb)

Empirically:

- Pacbaby observes $Y = 1$ or $Y = -1$ (Pacman or not) 50% of the time.
- Given $Y = 1$, Pacbaby observes $M = 1$ (moustache) 50% of the time and $S = 1$ (sunglasses) 50% of the time.
- When Pacbaby observes $Y = -1$, the frequency of observations are identical (equal probabilities of $M = 1, -1$, $S = 1, -1$).
- When Pacbaby observes $Y = 1$, anyone with a moustache wears sunglasses and anyone without a moustache does not wear sunglasses.
- If $Y = -1$ the presence/absence of a moustache has no influence on sunglasses.

(a) Based on the above information, fill in Pacbaby's conditional probability tables.

(b) Pacbaby sees someone with a moustache and wearing a pair of sunglasses. What prediction does the NB model make? What probability does it assign to its prediction? What prediction does Pacbaby's TANB model make? What probability does it assign to its prediction?

Workshop 12

COMP90051 Statistical Machine Learning

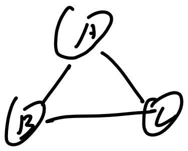
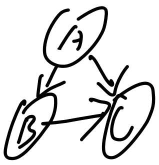
Semester 2, 2024

Learning Outcomes

By the end of this workshop you should be able to:

1. explain why variable elimination order affects the efficiency of inference on directed PGMs
2. specify a PGM based on a natural language description

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

no arrow

$P(y_1, y_2, \ldots, y_n) = \underbrace{P(y_1|y_2,\ldots,y_n)P(y_2|y_3,\ldots,y_n)}_{\text{所有的都给}}\underbrace{P(y_n)}_{\text{except } y_n}$

$= \prod_{i=1}^{n-1} P(y_i|y_{i+1},\ldots,y_n)P(y_n)$

NB: feature independence

Navie Bayes

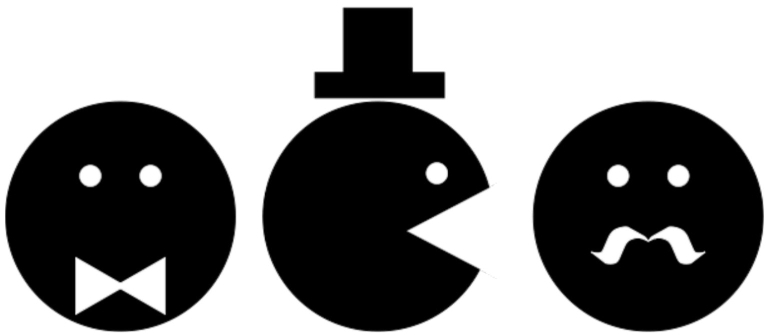
$P(X_1,\ldots,X_n,y) = P(X_1|X_2,\ldots,y)P(X_2|X_3,\ldots,y)$
$P(X_n|y)P(y)$

$= \underbrace{P(X_1|y)}_{\text{与其他X的条件独立}}P(X_2|y)\cdots P(X_n|y)P(y)$

$= \left[\prod_{i=1}^{n} P(X_i|y)\right]P(y)$

$\Sigma_A P(A,B) = P(B)$
(Marginalization)

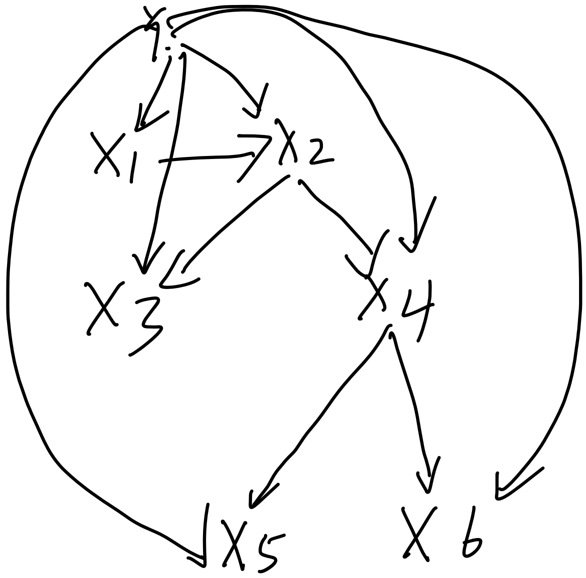
Context for Worksheet 12



- Pacbaby's parents are trying to teach her to discriminate between Pacmen () and ghosts ()

$y=1$ $y=-1$

- She will use visual features such as presence of bowtie, hat, moustache etc., denoted by $x_1 \cdots x_6$
- The features are *not independent*, so Pacbaby's parents decide to use a tree-augmented Naïve Bayes (TANB) model

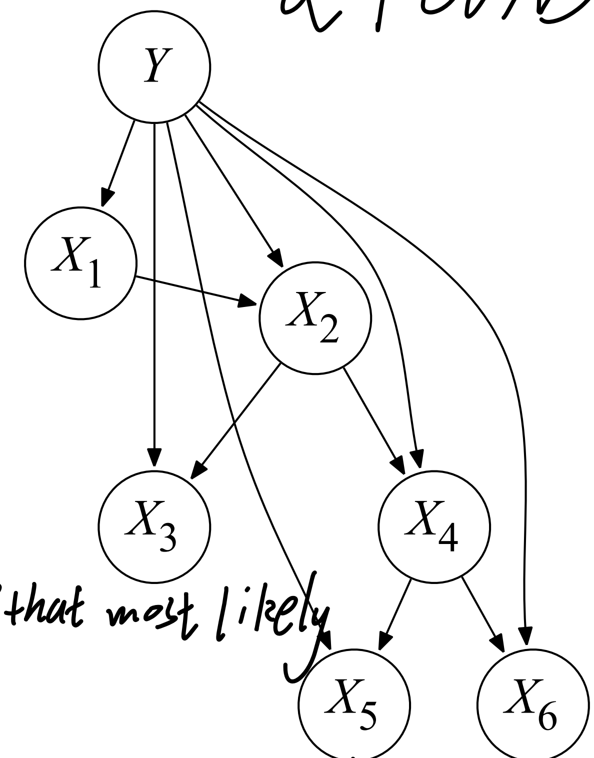

Y
X1
X2
X3
X4
X5
X6

Q1a: TANB model

$$\underbrace{P(V|V)}_{\Downarrow \text{ max}} = \frac{P(V,V)}{P(V)}$$

$$\propto P(V,V)$$

Assume all features are observed. What is the classification rule? Your answer should be in terms of the conditional distributions.



- Classification rule is the class that maximises the posterior probability

$$y^* = \arg\max_{y} P(y|x_1 \cdots x_6)$$ take the y that most likely

$\Downarrow$ $P(x_5|x_4, y)$

- Applying Bayes' rule and exploiting conditional dependence structure we have

$$P(y|x_1, \cdots, x_6) \propto P(y, x_1, \cdots, x_6)$$

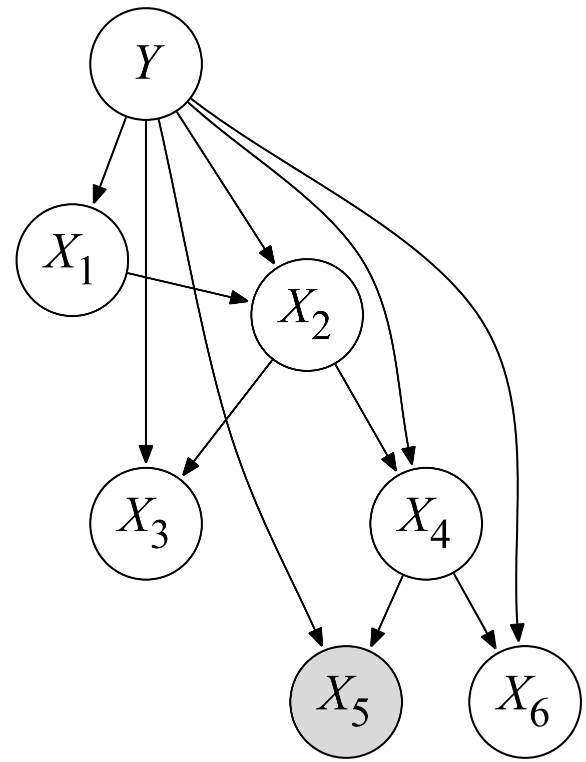
$$= P(x_1|y)P(x_2|y,x_1)P(y)$$
$$P(x_3|x_2,y)\, P(x_4|x_2,y)\, P(x_5|x_4,y)\, P(x_6|x_4,y)$$

但现在只看到了X_5 ⇒ 所以想计算的 ⇒ $P(Y|X_5)$ ⇒ By marginalisation.

Q1b: Efficient variable elimination

By

Specify an efficient elimination order for the query . How many variables are in the biggest factor induced by variable elimination? Which variables are they?



- Recall each step of elimination:
  - Removes a node
  - Connects node's remaining neighbours
- Time complexity is exponential in the largest clique of the induced graph
- Different elimination orderings produce different cliques

Q1b: Efficient variable elimination

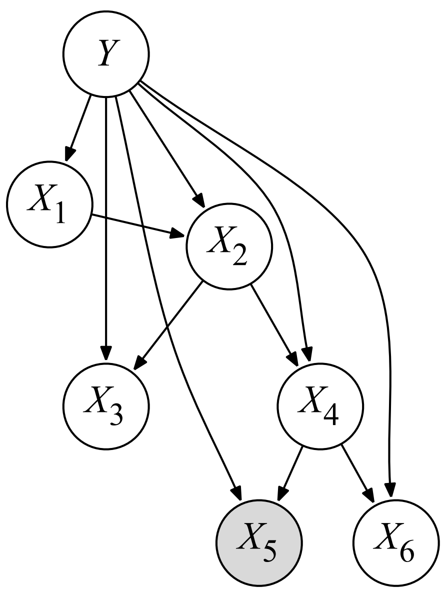
Try eliminating in the order $x_6 \to x_3 \to x_4 \to x_2 \to x_1$

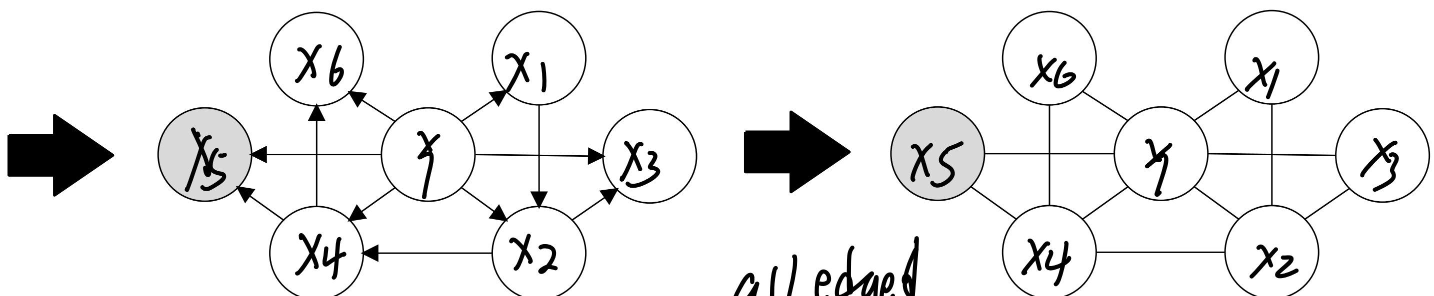
Q1b: Efficient variable elimination

Try eliminating in the order

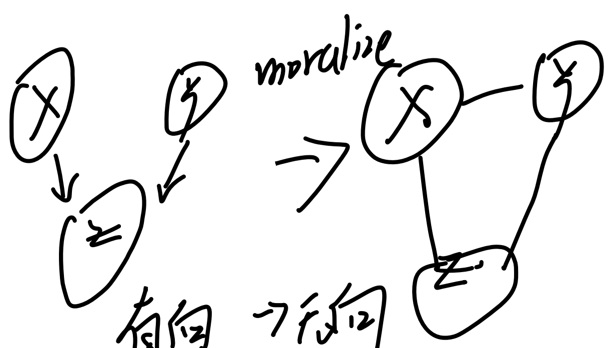
Q1b: Efficient variable elimination

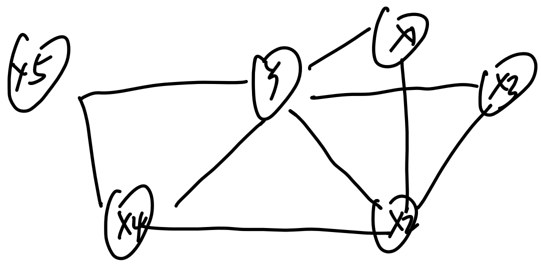
- Let's try a graphical approach.
- Re-arrange graph and moralise — add an edge between any nodes that share a child



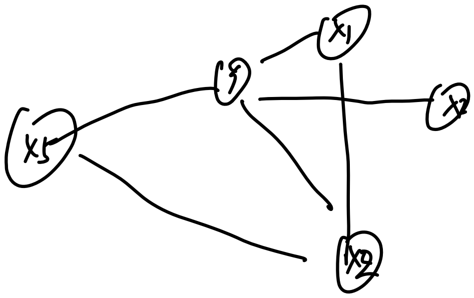


all edges are moralizeds





eliminate $x_4 \to x_2 - x_5$

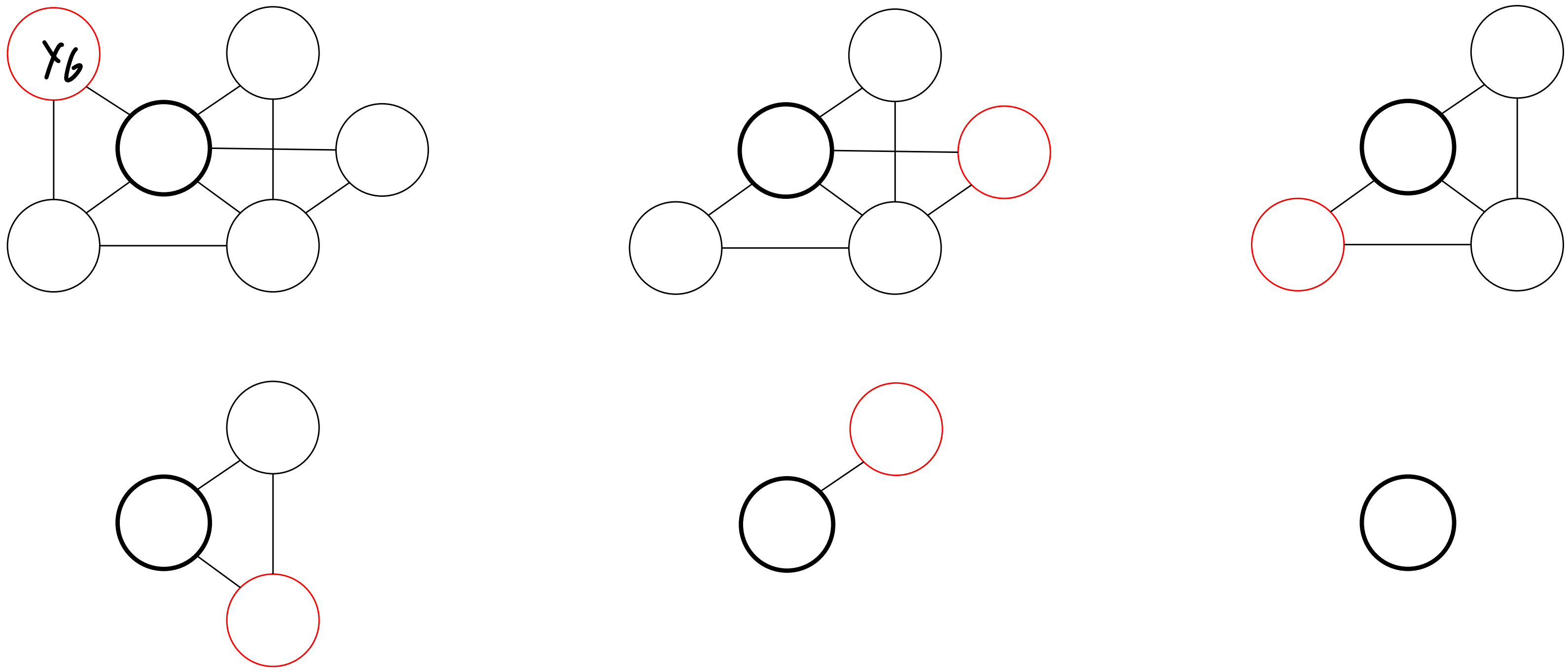


$p(X_3 \mid X_5 = x_5)$?

Q1b: Efficient variable elimination

why eliminate x_5?

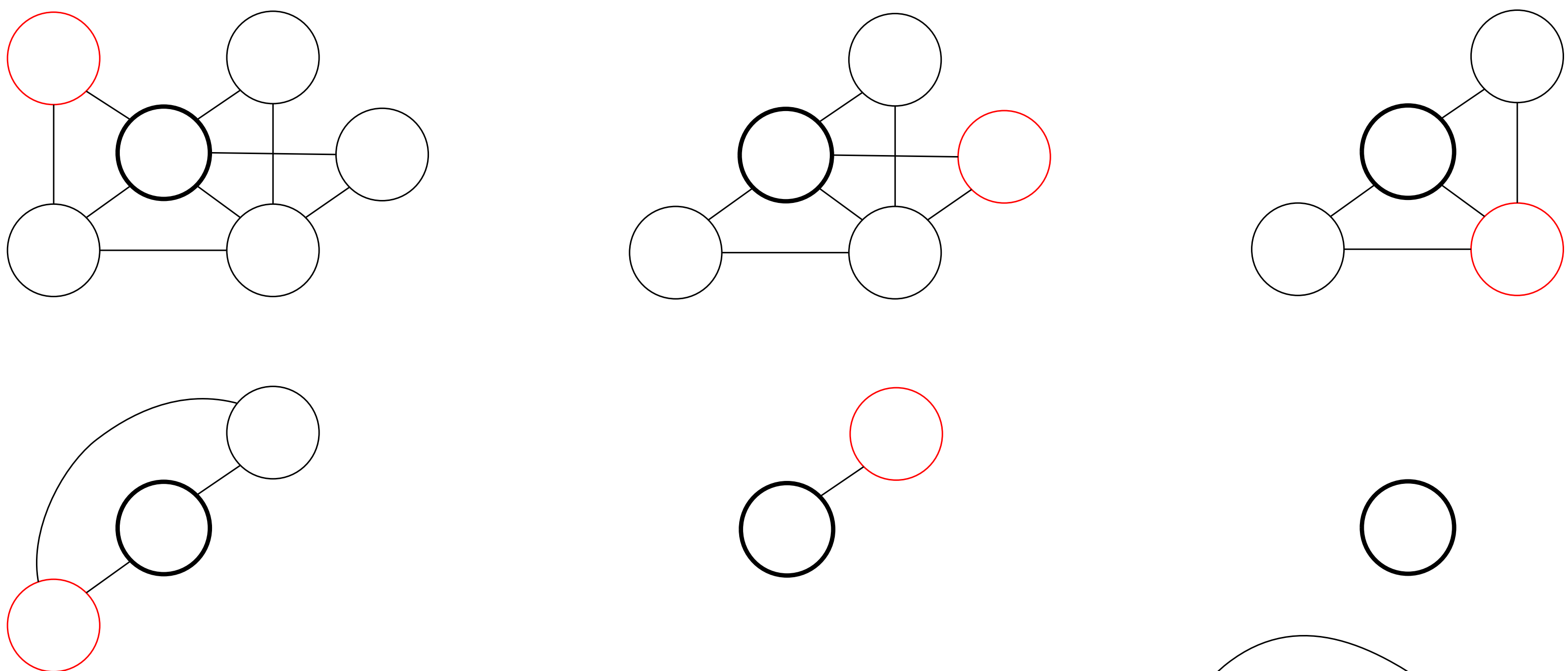
Try eliminating in the order



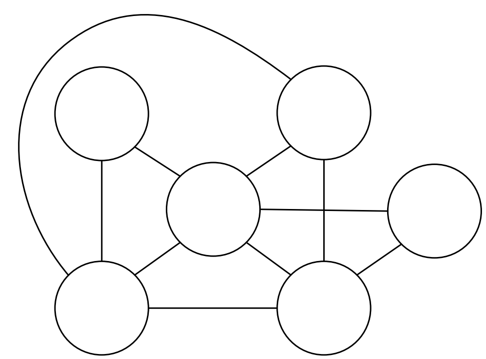
Induced graph is same as top left. Largest clique size is 3.

Q1b: Efficient variable elimination

Try eliminating in the order

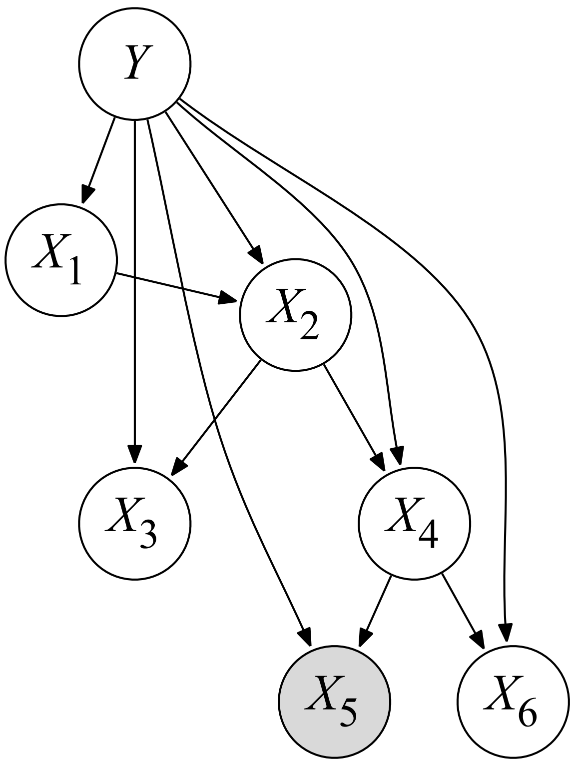


Induced graph has an additional edge between and . Largest clique size is 4.



Q1c: Efficient variable elimination

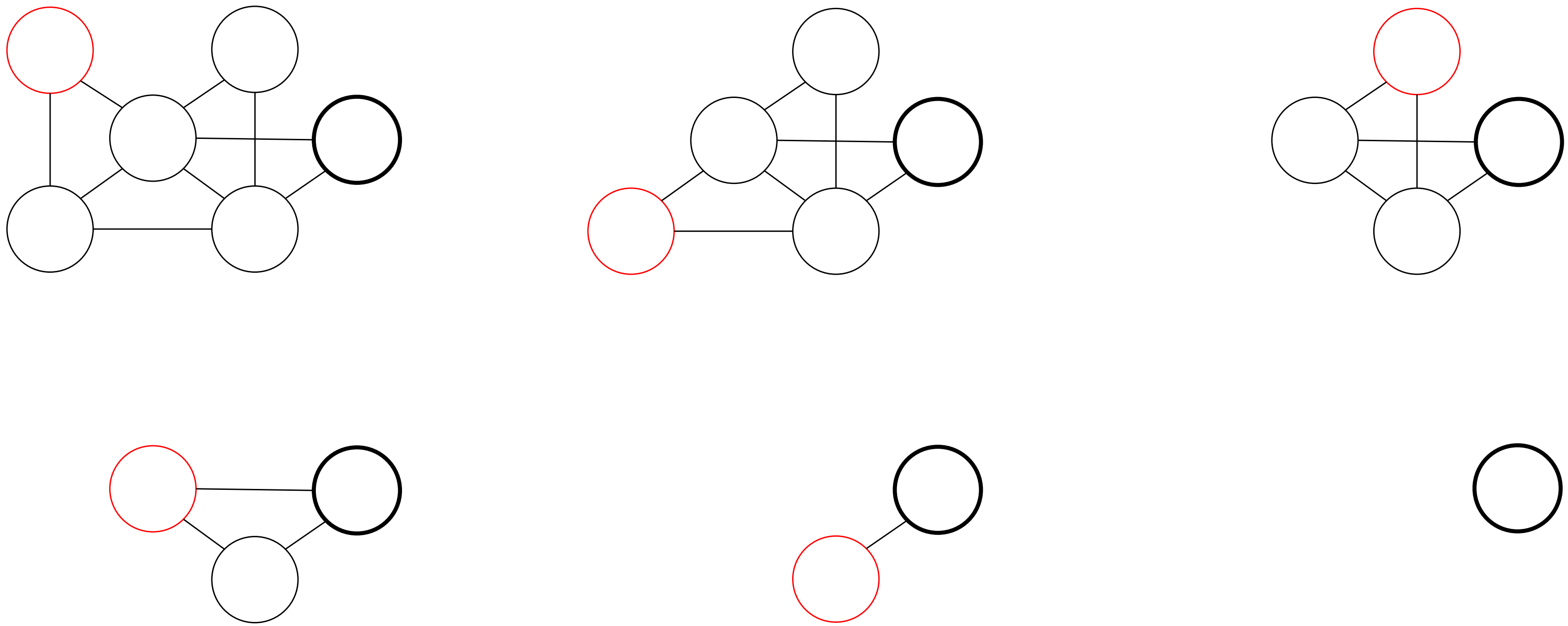
Specify an efficient elimination order for the query . How many variables are in the biggest factor induced by variable elimination? Which variables are they?



We'll use the graphical approach.

Q1c: Efficient variable elimination

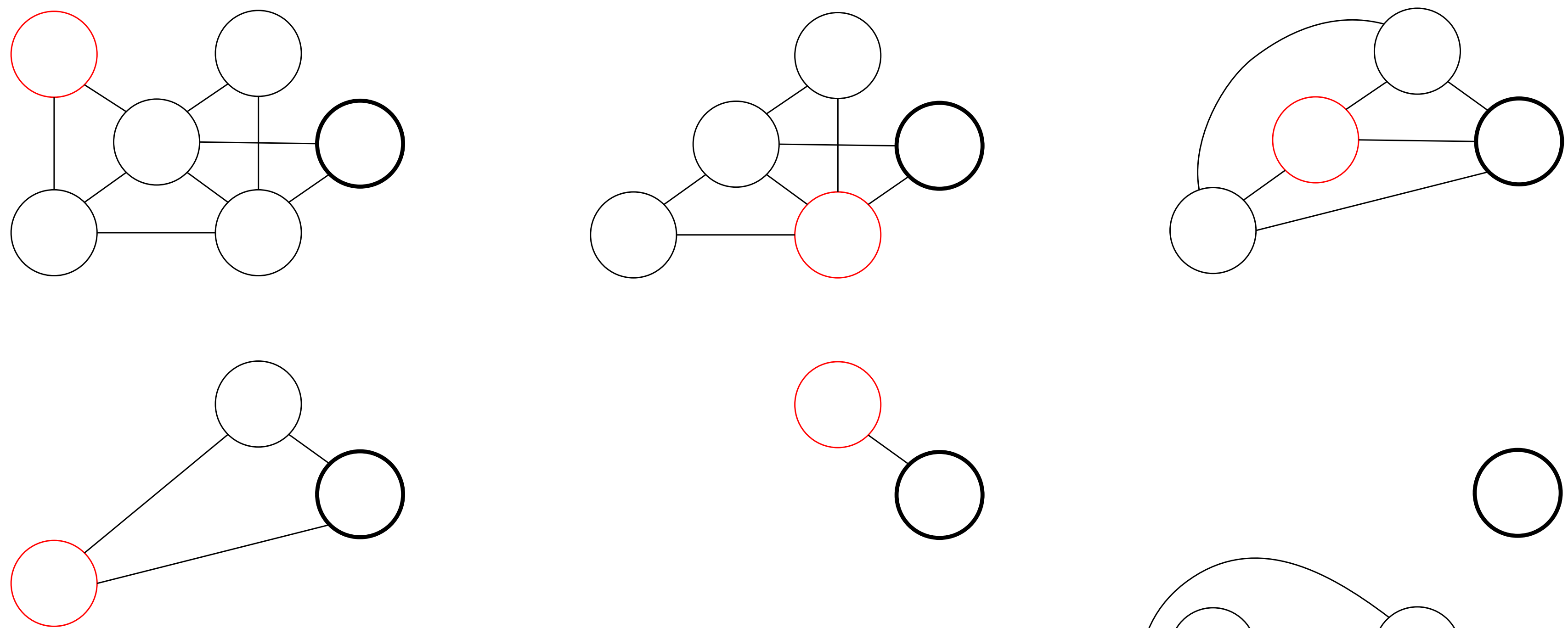
Try eliminating in the order

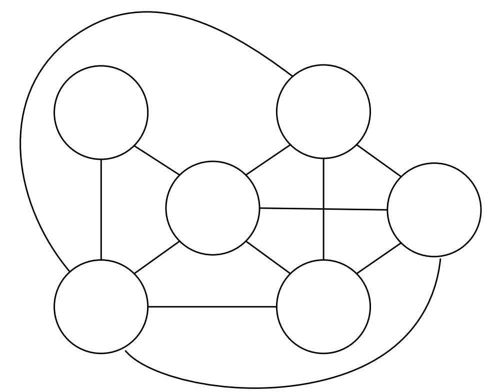


Induced graph is same as top left. Largest clique size is 3.

Q1c: Efficient variable elimination

Try eliminating in the order



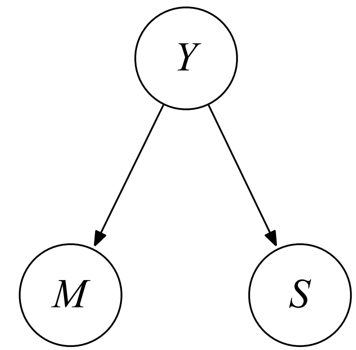


Induced graph has an additional edge between - , - and - . Largest clique size is 5.

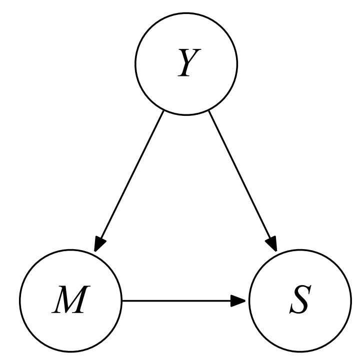
Q2a: CPTs

Use the following facts to fill out the conditional probability tables for the NB and TANB models:

- Pacbaby observes or 50% of the time
- Given , Pacbaby observes (moustache) 50% of the time and (sunglasses) 50% of the time
- When Pacbaby observes , the frequency of observations are identical (equal probabilities of and)
- When Pacbaby observes , anyone with a moustache wears sunglasses and anyone without a moustache does not wear sunglasses
- If the presence/absence of a moustache has no influence on sunglasses



(NB)



(TANB)

Q2a: CPTs

$p(y)p(M|y)p(S|y)$

NB model

$y=1$	0.5
$y=-1$	0.5

$p(M|y)$

	$y=1$	$y=-1$
$M=1$	0.5	0.5
$M=-1$	0.5	0.5

$p(S|y)$

	$y=1$	$y=-1$
$S=1$	0.5	0.5
$S=-1$	0.5	0.5

$p(Y)p(M|Y)p(S|M,Y)$

TANB model

$p(S|M,y)$

	$y=1$		$y=-1$	
	$M=1$	$M=-1$	$M=1$	$M=-1$
$S=1$	1	0	0.5	0.5
$S=-1$	0	1	0.5	0.5

Q2a: CPTs

NB model

$\arg\max_y P(y \mid M=1, S=1)$

$\propto P(M=1 \mid y) P(S \mid y) P(y)$

in NB: $\begin{cases} (0.5)^3 & y=1 \\ (0.5)^3 & y=-1 \end{cases} \Rightarrow \{y=1, y=-1\}$ 一样

0.5 0.5 normalize.

TANB model

$P(y \mid M=1, S=1)$

$\propto P(M=1 \mid y) P(S=1 \mid M=1, y) P(y)$

$\begin{cases} (0.5) \times 1 \times 0.5 & y=1 \\ (0.5 \times 0.5 \times 0.5) & y=-1 \end{cases}$

$\begin{cases} 0.25 \\ (0.5)^3 \end{cases}$

$\therefore$ $y=1$ is selected with $P = \frac{0.25}{(0.5)^3 + 0.25}$

Q2b: Query

Pacbaby sees someone with a moustache wearing a pair of sunglasses.
What prediction does the NB model make? What probability does it assign to its prediction?

Under the NB model

So there is a tie between the two classes.

Q2b: Query

Pacbaby sees someone with a moustache wearing a pair of sunglasses.
What prediction does the TANB model make? What probability does it assign to its prediction?

Under the TANB model

Normalising we have . So the model predicts that a Pacman was observed.