Workshop 11

COMP90051 Machine Learning

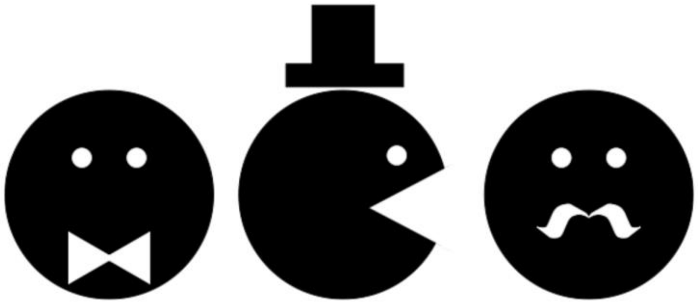
Semester 2, 2020

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
3. (extension) perform approximate inference on a PGM using PyStan

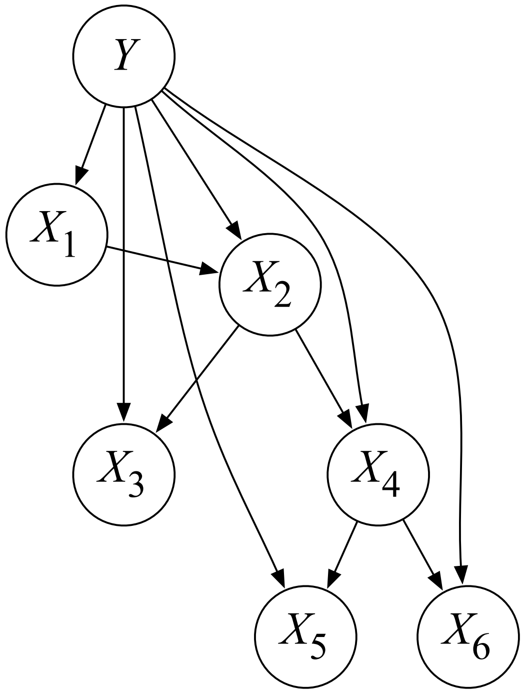
Context for Worksheet 11a



- Pacbaby's parents are trying to teach her to discriminate between Pacmen ($Y = 1$) and ghosts ($Y = -1$)
- She will use visual features such as presence of bowtie, hat, moustache etc., denoted by $X_1, X_2, \dots, X_6$
- The features are *not independent*, so Pacbaby's parents decide to use a tree-augmented Naïve Bayes (TANB) model

Q1a: TANB model

Assume all features $\mathbf{X} = (X_1, \ldots, X_6)$ 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

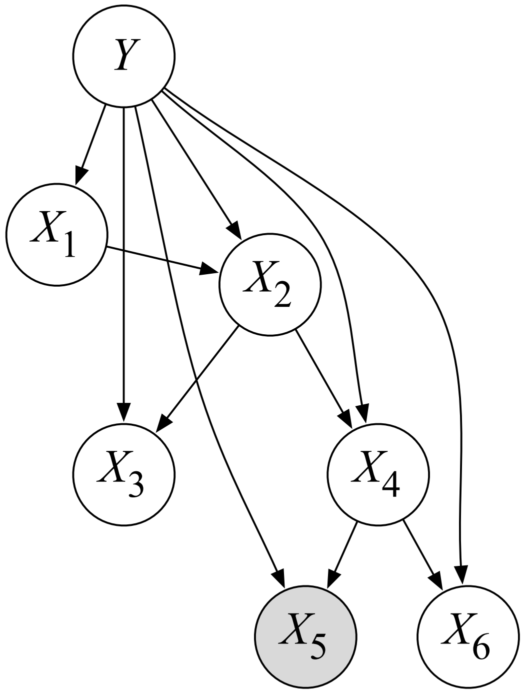
$$y^\star = \arg\max_y p(Y = y | \mathbf{X} = \mathbf{x})$$

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

$$p(Y = y | \mathbf{X} = \mathbf{x}) \propto p(y)p(x_1|y)p(x_2|x_1, y)$$
$$p(x_3|x_2, y)p(x_4|x_2, y)p(x_5|x_4, y)p(x_6|x_4, y)$$

Q1b: Efficient variable elimination

Specify an efficient elimination order for the query $p(Y|X_5 = x_5)$. 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 $X_6 \to X_3 \to X_4 \to X_2 \to X_1$

We eliminate these two since they are both 1 as non of their parents and children are depending the evidence node

$$p(Y|x_5) \propto \sum_{X_1,X_2,X_3,X_4,X_6} p(Y)p(X_1|Y)p(X_2|X_1,Y)p(X_3|X_2,Y)$$
$$p(X_4|X_2,Y)p(x_5|X_4,Y)p(X_6|X_4,Y)$$
$$= p(Y)\sum_{X_1} p(X_1|Y)\sum_{X_2} p(X_2|X_1,Y)\sum_{X_4} \underbrace{p(X_4|X_2,Y)p(x_5|X_4,Y)}_{\phi^1(X_2,X_4,Y)}$$
$$= p(Y)\sum_{X_1} p(X_1|Y)\sum_{X_2} \underbrace{p(X_2|X_1,Y)\phi^2(X_2,Y)}_{\phi^3(X_1,X_2,Y)}$$
$$= p(Y)\sum_{X_1} \underbrace{p(X_1|Y)\phi^4(X_1,Y)}_{\phi^5(X_1,Y)}$$
$$= \phi^6(Y)$$

Q1b: Efficient variable elimination

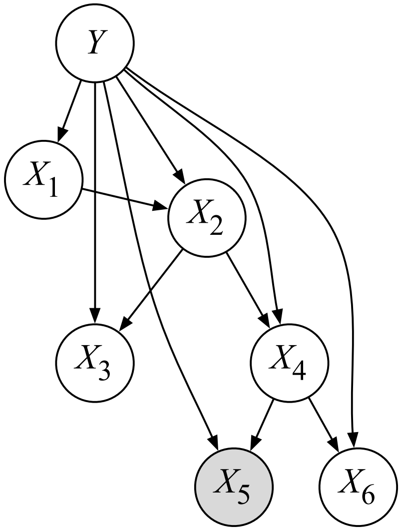
Try eliminating in the order $X_6 \rightarrow X_3 \rightarrow X_2 \rightarrow X_4 \rightarrow X_1$

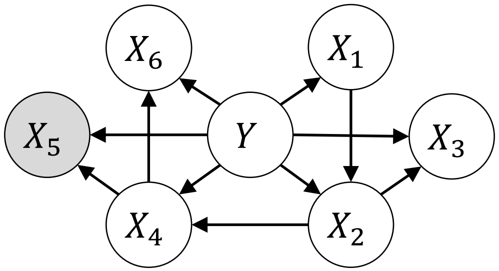
$$
\begin{aligned}
p(Y|x_5) \propto & \sum_{X_1,X_2,X_3,X_4,X_6} p(Y)p(X_1|Y)p(X_2|X_1,Y)p(X_3|X_2,Y) \\
& \qquad p(X_4|X_2,Y)p(x_5|X_4,Y)p(X_6|X_4,Y) \\
&= p(Y)\sum_{X_1} p(X_1|Y)\sum_{X_4} p(x_5|X_4,Y)\sum_{X_2} \underbrace{p(X_2|X_1,Y)p(X_4|X_2,Y)}_{\phi^1(X_1,X_2,X_4,Y)} \\
&= p(Y)\sum_{X_1} p(X_1|Y)\sum_{X_4} \underbrace{p(x_5|X_4,Y)\phi^2(X_1,X_4,Y)}_{\phi^3(X_1,X_4,Y)} \\
&= p(Y)\sum_{X_1} \underbrace{p(X_1|Y)\phi^4(X_1,Y)}_{\phi^5(X_1,Y)} \\
&= \phi^6(Y)
\end{aligned}
$$

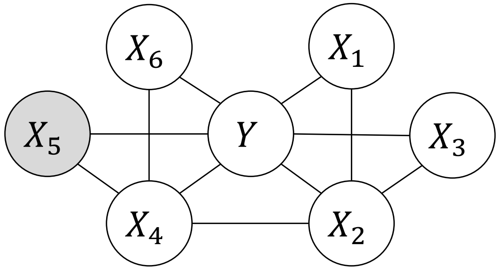
Tree width = 4 - 1 = 3

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

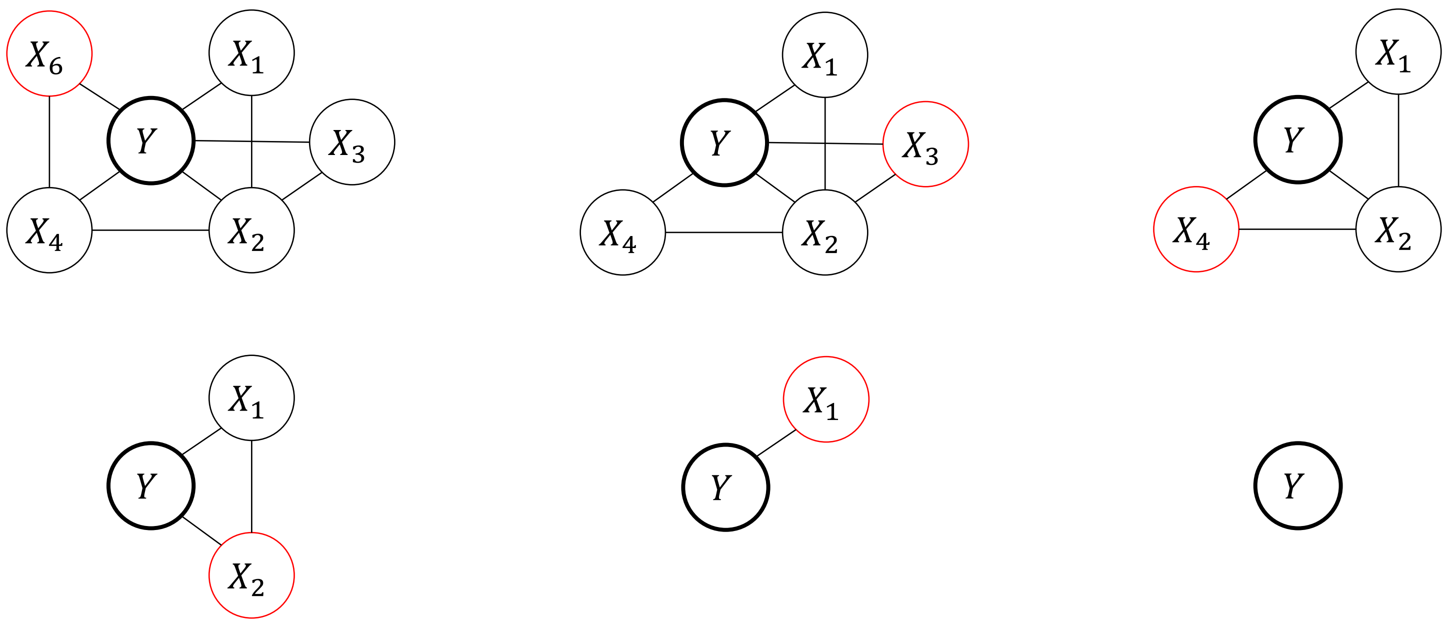






Q1b: Efficient variable elimination

Try eliminating in the order $X_6 \rightarrow X_3 \rightarrow X_4 \rightarrow X_2 \rightarrow X_1$

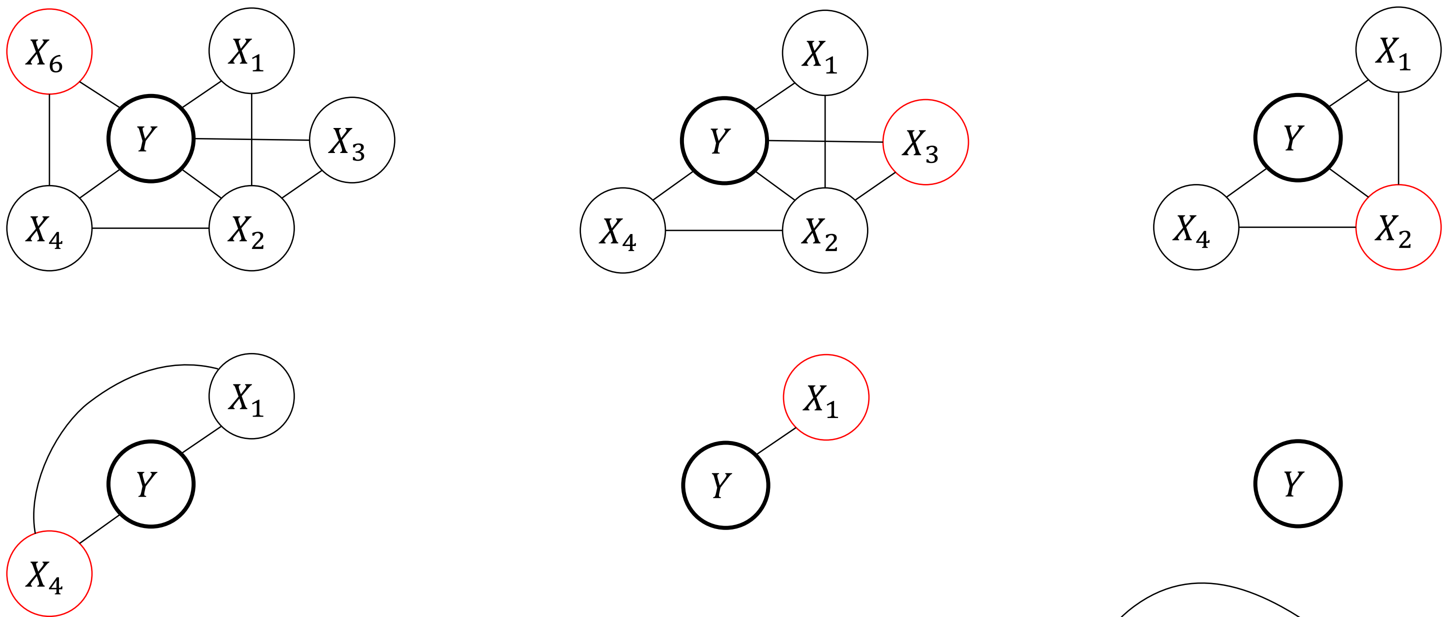


Induced graph is same as top left. Largest clique size is 3.

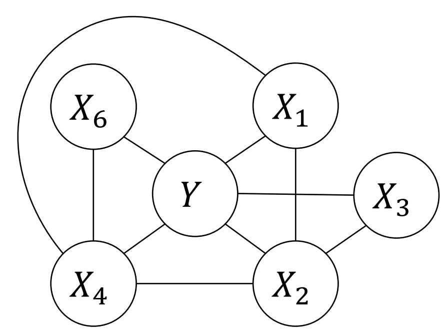
Q1b: Efficient variable elimination

Thus, we can conclude that the best strategy is not adding any extra edge to the original graph

Try eliminating in the order $X_6 \to X_3 \to X_2 \to X_4 \to X_1$

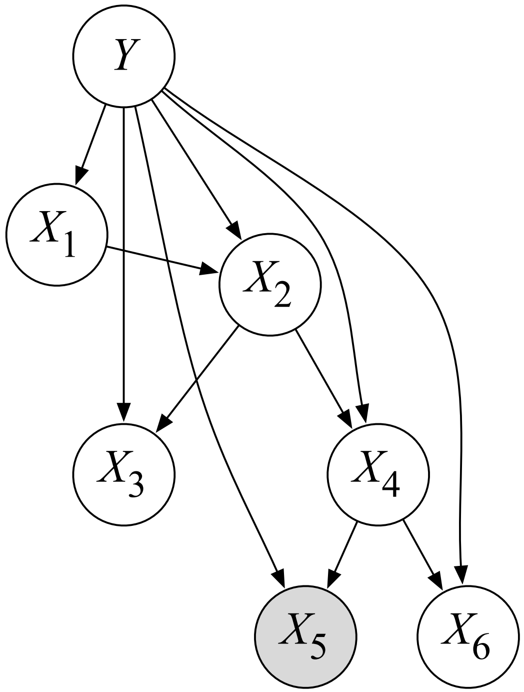


Induced graph has an additional edge between X_1 and X_4. Largest clique size is 4.



Q1c: Efficient variable elimination

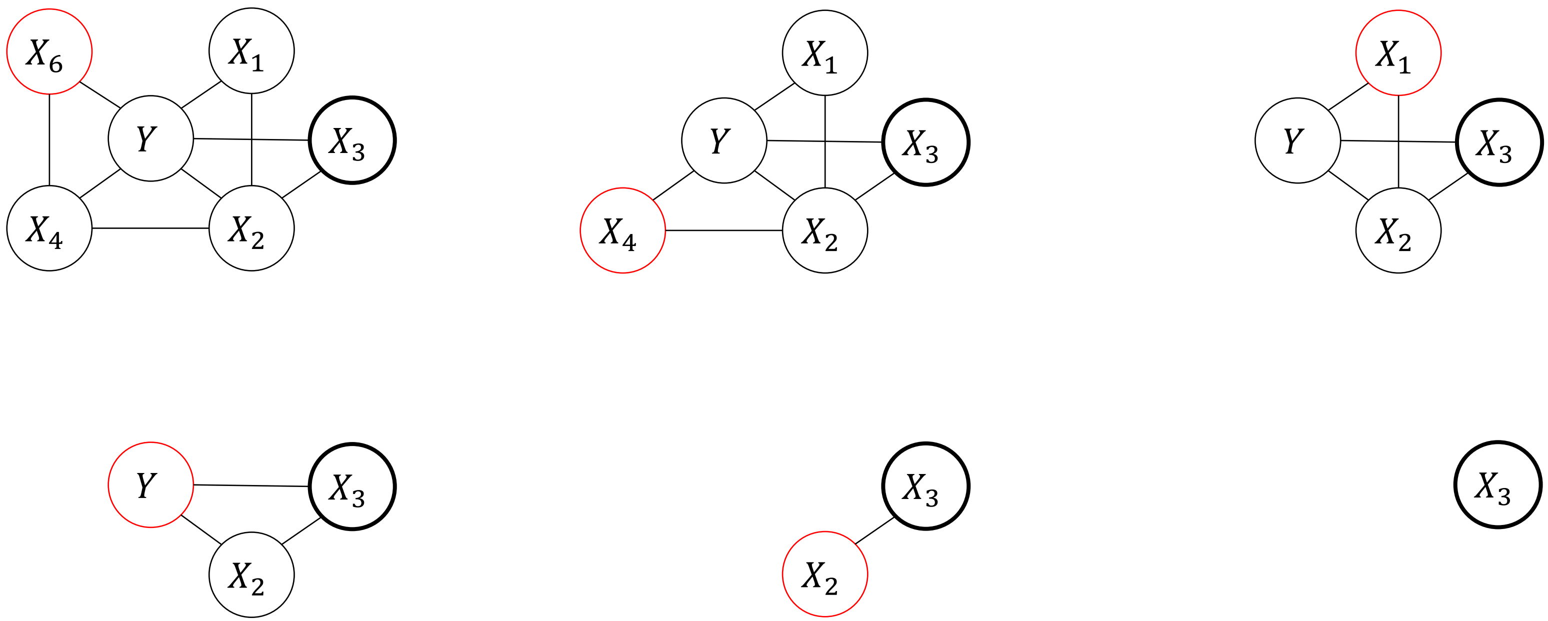
Specify an efficient elimination order for the query $p(X_3|X_5 = x_5)$. How many variables are in the biggest factor induced by variable elimination? Which variables are they?



We'll use the graphical approach.

Q1c: Efficient variable elimination

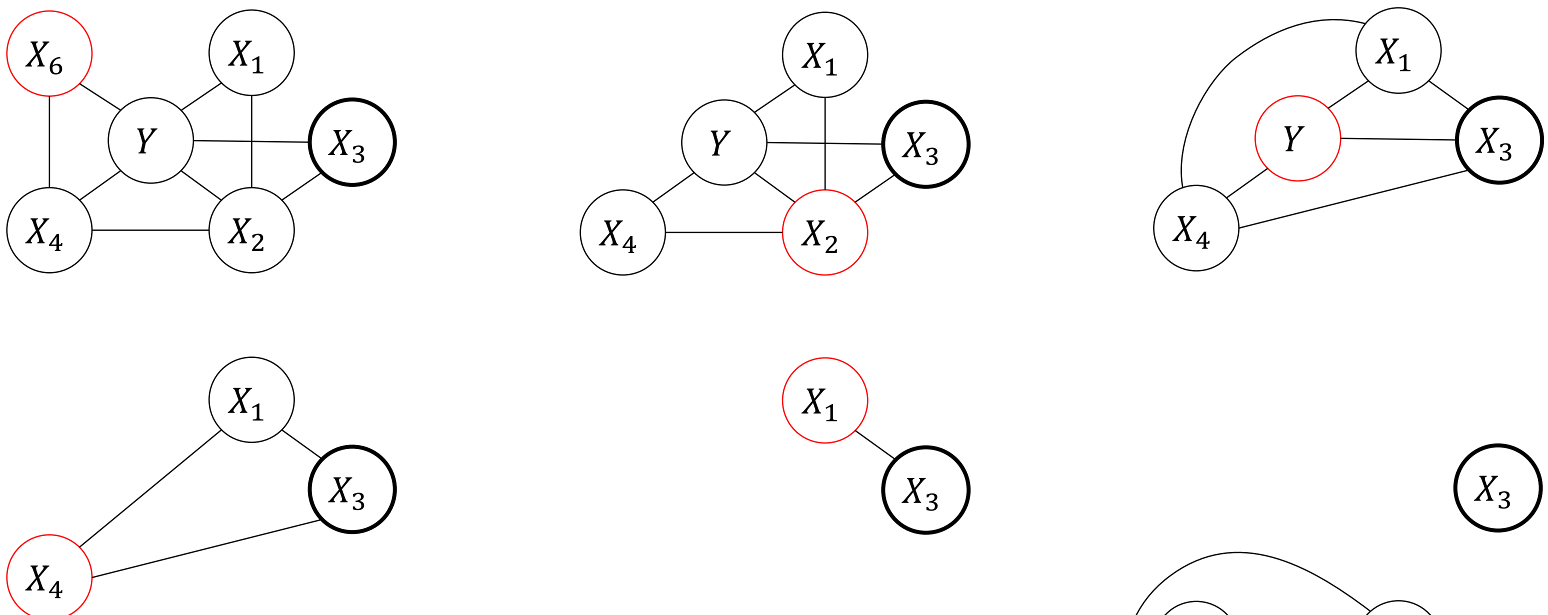
Try eliminating in the order $X_6 \rightarrow X_4 \rightarrow X_1 \rightarrow Y \rightarrow X_2$

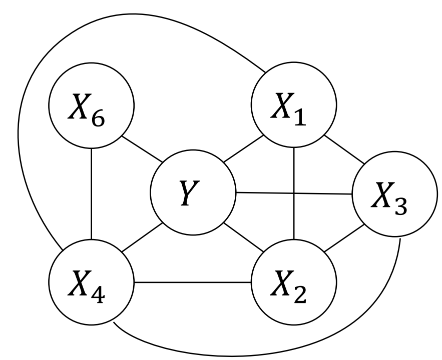


Induced graph is same as top left. Largest clique size is 3.

Q1c: Efficient variable elimination

Try eliminating in the order $X_6 \rightarrow X_2 \rightarrow Y \rightarrow X_4 \rightarrow X_1$



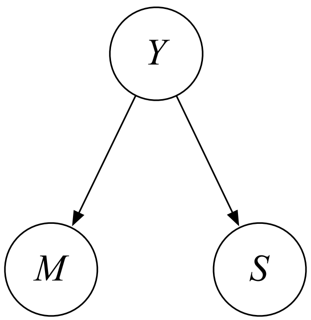


Induced graph has an additional edge between X_1 and X_4. Largest clique size is 5.

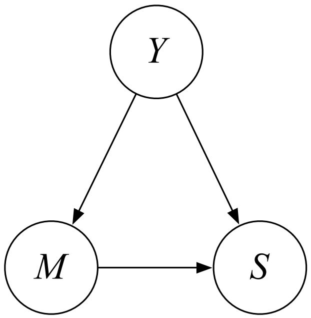
Q2a: CPTs

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

- Pacbaby observes $Y = 1$ or $Y = -1$ 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 $M = 1, -1$ and $S = 1, -1$)
- 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



(NB)



(TANB)

Q2a: CPTs

NB model

$P(Y = y)$	
$y = 1$	$y = -1$

y	$P(M = m \mid Y = y)$ $m = 1$	$m = -1$
1		
-1		

y	$P(S = s \mid Y = y)$ $s = 1$	$s = -1$
1		
-1		

TANB model

$P(Y = y)$	
$y = 1$	$y = -1$

y	$P(M = m \mid Y = y)$ $m = 1$	$m = -1$
1		
-1		

y	m	$P(S = s \mid Y = y, M = m)$ $s = 1$	$s = -1$
1	1		
-1	1		
1	-1		
-1	-1		

Q2a: CPTs

NB model

$P(Y=y)$	
$y=1$	$y=-1$
0.5	0.5

y	$P(M=m \mid Y=y)$ $m=1$	$m=-1$
1	0.5	0.5
-1	0.5	0.5

y	$P(S=s \mid Y=y)$ $s=1$	$s=-1$
1	0.5	0.5
-1	0.5	0.5

TANB model

$P(Y=y)$	
$y=1$	$y=-1$
0.5	0.5

y	$P(M=m \mid Y=y)$ $m=1$	$m=-1$
1	0.5	0.5
-1	0.5	0.5

y	m	$P(S=s \mid Y=y, M=m)$ $s=1$	$s=-1$
1	1	1	0
-1	1	0.5	0.5
1	-1	0	1
-1	-1	0.5	0.5

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

$$p(Y|M=1,S=1) \propto p(Y)p(M=1|Y)p(S=1|Y)$$

$$= \begin{cases} \left(\frac{1}{2}\right)^3, Y = 1 \\ \left(\frac{1}{2}\right)^3, Y = -1 \end{cases}$$

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

$$p(Y|M=1,S=1) \propto p(Y)p(M=1|Y)p(S=1|M=1,Y)$$

$$= \begin{cases} \left(\frac{1}{2}\right)^2, Y=1 \\ \left(\frac{1}{2}\right)^3, Y=-1 \end{cases}$$

Normalising we have $p(Y=1|M=1,S=1)=\frac{2}{3}$. So the model predicts that a Pacman was observed.

Worksheet 11b