

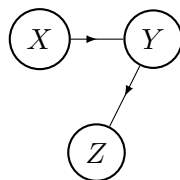
**Deadline: Th. Jan. 17, 10:00 am** Drop your printed or legible handwritten submissions into the boxes at Samelsonplatz, or upload them as `.pdf` or `.ipynb` files onto the LearnWeb.

**Exercise 1** (Conditional independence - 8 points).

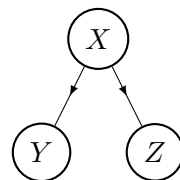
- (2) What does it mean in lay-mans term if we say two events  $A$  and  $B$  are conditionally independent given that an event  $C$  occurred?
- (2) Draw the Bayesian network associated with the joint pdf:

$$p(x_1, x_2, x_3, x_4, x_5) = p(x_5|x_1, x_3, x_4)p(x_4|x_2, x_3)p(x_3|x_1, x_2)p(x_2|x_1)p(x_1)$$

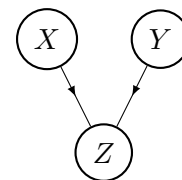
- (4) Consider the following Bayesian Networks



(a) "chain"



(b) "split" or "fork"



(c) "join" or "collider"

These graphs are associated with the joint probabilities:

(a)  $p(x, y, z) = p(z|y)p(y|x)p(x)$

(b)  $p(x, y, z) = p(y|x)p(z|x)p(x)$

(c)  $p(x, y, z) = p(z|x, y)p(y)p(x)$

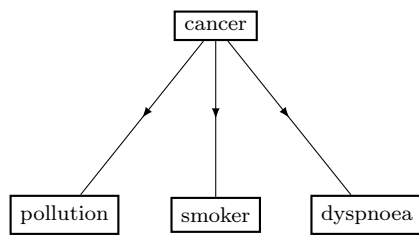
Show that, for the different cases respectively, holds:

- $X$  and  $Z$  are conditionally independent given  $Y$
- $Y$  and  $Z$  are conditionally independent given  $X$
- $X$  and  $Y$  are generally **not** conditionally independent given  $Z$ <sup>1</sup>

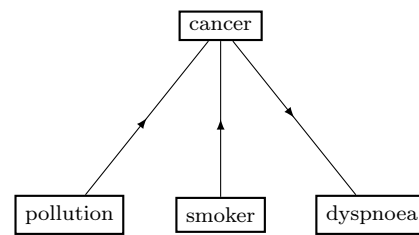
**Exercise 2** (Naïve Bayes - 12 points).

Given the data from table 1, we want to predict the probability that a patient has lung-cancer given that we know whether or not they show symptoms of dyspnoea (breathing problems), whether or not they are a smoker and whether or not they live in an area with high air pollution. We consider two different graphical models:

<sup>1</sup>Provide a counter example.



(a) "Naïve Bayesian classifier"



(b) "Tree Augmented Naïve Bayesian classifier"

1. (2) For both models, write out the joint probability.
2. (8) Train both models with the provided data for patient 1-12. Use  $\alpha = 1$ , aka add-one-smoothing as the prior and provide the CPT for each node.
3. (2) What do both models predict for the missing values for patient 13 and 14?

patient	air-pollution	smoker	dyspnoea	cancer
1	high	yes	yes	yes
2	high	yes	yes	yes
3	high	yes	yes	no
4	high	yes	no	yes
5	high	yes	no	yes
6	high	no	yes	yes
7	high	no	yes	no
8	low	yes	yes	yes
9	low	yes	yes	no
10	low	yes	no	no
11	low	no	yes	no
12	low	no	no	no
13	low	yes	yes	?
14	low	yes	?	yes

Table 1: Synthetic lung-cancer data-set