Deadline: Th. January 16th , 14:00 Drop your printed or legible handwritten submissions into the boxes at Samelsonplatz. Alternatively upload a `.pdf` file via LearnWeb. (e.g. exported Jupyter notebook)

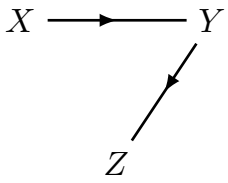
1. Conditional independence (8 points)

A. [2p] What does it mean in lay-mans terms if we say two events A and B are conditionally independent given that an event C occurred?

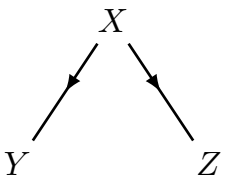
B. [2p] Draw the Bayesian network associated with the joint pdf:

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

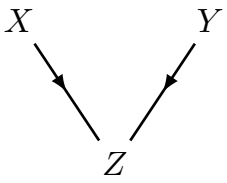
C. [4p] Consider the following Bayesian Networks



(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:

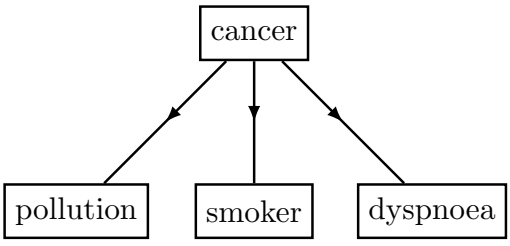
(a) X and Z are conditionally independent given Y

(b) Y and Z are conditionally independent given X

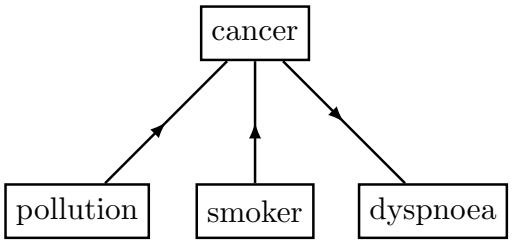
(c) X and Y are generally **not** conditionally independent given Z[1]

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), are a smoker and live in an area with high air pollution. We consider two different graphical models:



(a) "Naïve Bayesian classifier"



(b) "Tree Augmented Naïve Bayesian classifier"

A. [2p] For both models, write out the associated joint probability.

B. [8p] Train both models with the provided data for patient 1-12. Use $\alpha = 1$, i.e. add-one-smoothing for the Dirichlet prior. Provide the conditional probability tables (CPT) for each node.

[1] Provide a counter example.

C. [2p] 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