Deadline: Friday January 22nd, 10:00 Upload a `.pdf` file via LearnWeb. (e.g. exported Jupyter notebook)

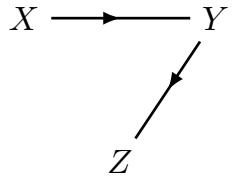
1 Conditional independence (8 points)

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?

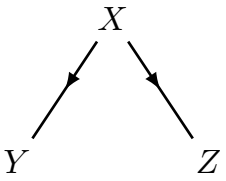
B [2p] Draw the Bayesian network associated with the joint pdf:

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

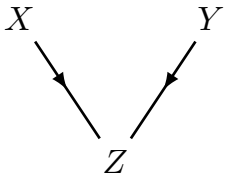
C [4p] Consider the following Bayesian Networks



(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)

A [5p] Given is an imbalanced data set for binary classification with 100,000 instances, only 10 labeled as positive, all remaining as negative. Our learned classifier is able to classify 99.99% of the instances correctly. What is the probability that an instance that is classified as positive is actually negative?

B [7p] Given is the following training data:

	Color	Type	Origin	Stolen
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	No
3	Blue	Sports	Domestic	Yes
4	Blue	Sports	Domestic	No
5	Blue	Sports	Imported	Yes
6	Blue	Grand tourer	Imported	No
7	Blue	Grand tourer	Imported	Yes
8	Blue	Grand tourer	Domestic	No
9	Red	Grand tourer	Imported	Yes
10	Red	Sports	Imported	Yes

[3 **a**] Calculate the probabilities

$$P(\text{Red}|\text{Yes}), \quad P(\text{Grand tourer}|\text{Yes}), \quad P(\text{Domestic}|\text{Yes}),$$
$$P(\text{Red}|\text{No}), \quad P(\text{Grand tourer}|\text{No}), \quad P(\text{Domestic}|\text{No})$$

[4 **b**]Predict the probability that a car with properties $X_1 = \text{Red}$, $X_2 = \text{Grand tourer}$, $X_3 = \text{Domestic}$ will be stolen.

[1] Provide a counter example.