Machine Learning 2 **Due: 13:00, April 16, 2018**

Homework 2

Instructors: Stephan Bongers, Stephan Alaniz Kupsch
Email: s.r.bongers@uva.nl, s.a.alanizkupsch@uva.nl

You are allowed to discuss with your colleagues but you should write the answers in *your own words*. If you discuss with others, write down the name of your collaborators on top of the first page. No points will be deducted for collaborations. If we find similarities in solutions beyond the listed collaborations we will consider it as cheating.
We will not accept any late submissions under any circumstances. The solutions to the previous homework will be handed out in the class at the beginning of the next homework session. After this point, late submissions will be automatically graded zero.

Problem 1. $(0.5 + 0.5 + 1 + 1 = 3$ pts$)$

1. Given three discrete random variables X, Y and Z. Give the definition of the mutual information $I(X;Y)$ and the conditional mutual information $I(X;Y|Z)$. Explain what the (conditional) mutual information measures.

Consider three variables x, y, $z \in \{0,1\}$ having the joint distribution $p(x,y,z)$ given in Table 1.

2. Evaluate the quantity $I(X;Y)$ and show that it is greater than zero. Hint: Compute the tables for $p(x,y)$, $p(x)$ and for $p(y)$. Moreover, remember that we use the convention that $0 \cdot ln(0) := 0$. Interpret this result, i.e. what does it mean that $I(X;Y) > 0$?

3. Evaluate $I(X;Y|Z)$ and show that it is equal to zero. Hint: Compute the tables for $p(x,y|z)$, $p(x|z)$ and for $p(y|z)$. Interpret this result, i.e. what does it mean that $I(X;Y|Z) = 0$?

4. Show that $p(x,y,z) = p(x)p(z|x)p(y|z)$, and draw the corresponding directed graph.

Table 1: The joint distribution over three binary variables.

x	y	z	$p(x,y,z)$
0	0	0	0.192
0	0	1	0.144
0	1	0	0.048
0	1	1	0.216
1	0	0	0.192
1	0	1	0.064
1	1	0	0.048
1	1	1	0.096

Problem 2. (2 pts)
Consider all the Bayesian networks consisting of three vertices X, Y and Z. Group them into clusters such that all the graphs in each cluster encode the same set of independence relations. Draw those clusters and write down the set of independence relations for each cluster.

Problem 3. (1 + 1 = 2 pts)

1. Given distributions p and q of a continuous random variable, Kullback-Leibler divergence of q from p is defined as
$$\mathcal{KL}(p||q) = -\int p(\mathbf{x}) \ln \left\{ \frac{q(\mathbf{x})}{p(\mathbf{x})} \right\} d\mathbf{x}$$
Evaluate the Kullback-Leibler divergence when $p(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $q(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\mathbf{m}, \mathbf{L})$

2. Entropy of a distribution p is given by
$$\mathcal{H}(p) = -\int p(\mathbf{x}) \ln p(\mathbf{x}) d\mathbf{x}$$
Derive the entropy of the multivariate Gaussian $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$