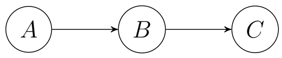
Machine Learning
Prof. S. Harmeling

Wednesday, 20. October 2021
DUE 23:55 Tuesday, 26. October 2021

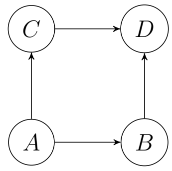
Exercise set #2

Please submit your solutions in teams of two using the sciebo file-drop folder. The link is available in ILIAS. For the formatting please stick to the `submission_guideline.pdf` that you can find on sciebo. In the case of multiple uploads we will consider the latest. Uploads after the deadline will be deleted without further notice.

1. **Bayes net calculations**
Consider the following directed acyclic graphs (DAGs):



$\mathcal{G}_1$



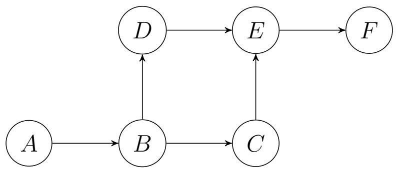
$\mathcal{G}_2$

   (a) Let A, B and C be binary random variables. Calculate $p(B)$ and $p(C)$ for the Bayes net defined by $\mathcal{G}_1$ and the (conditional) probabilities $p(A) = 0.3$, $p(B|A) = 0.2$, $p(B|\neg A) = 0.4$, $p(C|B) = 0.7$ and $p(C|\neg B) = 0.5$.

   (b) Let A, B, C and D be binary random variables. Calculate $p(B)$ and $p(D)$ for the Bayes net defined by $\mathcal{G}_2$ and the (conditional) probabilities $p(A) = 0.3$, $p(B|A) = 0.2$, $p(B|\neg A) = 0.4$, $p(C|A) = 0.7$, $p(C|\neg A) = 0.6$, $p(D|B, C) = 0.9$, $p(D|B, \neg C) = 0.5$, $p(D|\neg B, C) = 0.3$ and $p(D|\neg B, \neg C) = 0.3$.

25 points

2. **The d-separation criterion**
Consider the following DAG $\mathcal{G}$:



   (a) Write down the factorization of the joint distribution $p(A, B, C, D, E, F)$.

   (b) From your answer in (a) derive the marginal distribution $p(A, B, C, D, E)$.

   (c) Find a minimal set, that d-separates A and F (i.e. there is no other set with fewer elements, that also d-separates A and F). Prove that this is the case.

   (d) Prove that $C \perp\!\!\!\perp_{\mathcal{G}} D|B$ holds.

25 points

3. **Three node networks**
Prove all four cases from the lecture, i.e.:

(a) $A \to B \to C$	implies	$A \perp\!\!\!\perp C \mid B$
(b) $A \leftarrow B \leftarrow C$	implies	$A \perp\!\!\!\perp C \mid B$
(c) $A \leftarrow B \to C$	implies	$A \perp\!\!\!\perp C \mid B$
(d) $A \to B \leftarrow C$	implies	$A \perp\!\!\!\perp C$

For this write out the factorization of $p(A, B, C)$ that is implied by the graph and show the wanted conditional independence.

20 points

4. **Properties of expected value and variance**
Let X and Y be two discrete, not necessarily independent, random variables and $a \in \mathbb{R}$ a real number.

(a) Show that the expectated value is an linear operator, i.e.,

$$\mathrm{E}(aX + Y) = a\,\mathrm{E}(X) + \mathrm{E}(Y)$$

(b) Show that

$$\mathrm{Var}(aX) = a^2\,\mathrm{Var}(X)$$

10 points

5. **Mode of the Gaussian distribution**
The mode of a continuous probability distribution is the point at which the probability density function attains its maximum value. Prove that the mean μ is also the mode of the Gaussian normal distribution

$$\mathcal{N}(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right).$$

Hint: Try maximizing $\log \mathcal{N}(x|\mu, \sigma^2)$ and argue why it has the same maximum.

10 points

6. **Plotting distributions (programming task)**
The goal of this task is to make yourself familiar with some continuous distributions and plotting pdfs and histograms. First, add the following import statements:

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import beta, expon, gamma, laplace, norm
```

For each of the following distributions, plot its pdf in the given interval and a histogram based on 1000 samples. Use `.pdf(x, *args)` to compute the probability densities for the given array of x values, e.g., `norm.pdf(x, a, b)`. Use `.rvs(*args, n)` to sample `n` values, e.g., `norm.rvs(a, b, n)`. For plotting you may either use `plotly` or `matblotlib`.

(a) Normal distribution: $\mathcal{N}(5, 1.5)$. Plot interval: $[0, 10]$
(b) Laplace distribution: $\mathrm{Laplace}(5, 1.5)$. Plot interval: $[-5, 15]$
(c) Exponential distribution: $\mathrm{Exp}(0, 1/1.5)$. Plot interval: $[-1, 8]$
(d) Gamma distribution: $\mathrm{Gamma}(5, 0, 1)$. Plot interval: $[0, 15]$
(e) Beta distribution: $\mathrm{Beta}(2, 1.5, 0, 1)$. Plot interval: $[0, 1]$

If you are interested, play around with the parameters of the distributions and see what influence they have on the probability densities.

Important: Make sure you also submit the output produced by your Python code.

10 points